

МЕТОДЫ И СРЕДСТВА ПОСТРОЕНИЯ ОНТОЛОГИЙ ДЛЯ ИНТЕЛЛЕКТУАЛИЗАЦИИ СЕТИ ИНТЕРНЕТ

© 2002 С.В. Батищев¹, Т.В. Искварина¹, П.О. Скобелев²

¹ Самарский филиал Физического института им. П.Н. Лебедева РАН

² Институт проблем управления сложными системами РАН, г. Самара

В настоящей работе рассматривается проблема интеллектуализации сети Интернет и создания так называемого Semantic Web за счет реализации нового поколения программных Интернет-приложений, использующих понятие "онтология" для понимания смысла информационных ресурсов и кооперативного взаимодействия между собой. Сравниваются различные определения понятия "онтология", описываются критерии построения онтологий и онтологический базис первичных понятий и отношений для построения онтологий как "фрагментов мира". Рассматриваются методы формализации и интеграции онтологий, а также языковые средства для представления и работы с онтологиями. Формулируются требования к программным инструментам для построения онтологий и даются примеры успешной реализации онтологий в сети Интернет.

Введение

В настоящее время Интернет предоставляет пользователю огромное количество информации, а также разнообразных услуг, от web-магазинов до бюро путешествий. И с каждым днем интенсивность информационных потоков все увеличивается, что вызывает целый ряд новых проблем.

Первая проблема – желание пользователя получать только ту информацию или услугу, которая его интересует, и не получать огромное количество не нужных ему ссылок. Для решения этой задачи должны существовать более интеллектуальные инструменты, более точно понимающие желания пользователя. В частности, поисковые машины должны стремиться к действительному "пониманию" запросов, сделанных на естественном языке, накапливать знания о предметной области или пользователе и уметь делать из них определенные выводы.

Вторая проблема – совместимость различных приложений. В настоящее время ситуация сложилась довольно драматичная – огромное количество существующих приложений никак не в состоянии "понять" друг друга и, тем более, "разговаривать" друг с другом и о чем-то договориться, чтобы вместе предоставить потребителю определенную услугу (как, например, должны договариваться

ся программа продажи некоторого товара в магазине и программа доставки его со склада потребителю). Вместе с тем, взаимодействие программ могло бы положить конец постоянному дублированию разработок и унификации сервисов в сети.

К сожалению, современные системы, представленные в Web, не располагают средствами для решения этих проблем, и эта ситуация усугубляется разнородностью различных форматов представления данных, неполнотой, недоопределенностью и противоречивостью информации из различных источников, отсутствием средств приобретения, систематизации и интеграции знаний и т.д.

Для решения указанных проблем W3C Consortium разрабатывает идею Семантического Веба (Semantic Web). "Семантический Веб (Semantic Web) – это не отдельная сеть, а расширение современной сети, в которой информации присваивается четко определенный смысл, позволяющий людям и компьютерам лучше работать в сотрудничестве" [1]. Суть этой идеи заключается в том, чтобы научить, наконец, машины понимать содержание Интернет-страниц, т.е. распознавать и интерпретировать смысл (значение) той информации, которая находится на странице.

Для семантизации информации в Интернет в настоящее время активно развивается

онтологический подход. В рамках этого подхода каждый информационный квант сети получает описание в терминах специально-го формализованного фрагмента представления мира – так называемой "онтологии". Для системы бронирования билетов такая модель мира может включать взаимосвязанные представления понятий "самолет", "билет", "рейс", "дата вылета" и т.д. Для Интернет-магазина – "покупатель", "товар", "цена", "продажа", "поставка" и др. При заказе конкретного билета или покупке выбранного товара некоторым покупателем в системе (центре продажи билетов или магазине) сложится определенная сцена, описываемая комбинацией условий и действий, которая может быть интерпретирована программами, в результате чего пользователю может быть предложен другой рейс или дополнительный товар с большой скидкой [2].

При этом особый интерес представляют онтологии для интеллектуальных агентов – программных объектов, способных воспринимать ситуацию, принимать решения и коммуницировать с себе подобными [3]. Загружая ту или иную онтологию, агенты могут мгновенно настраиваться на новую предметную область, а также обмениваться между собой не только данными, но и знаниями, что создает предпосылки для создания самоорганизующихся и саморазвивающихся распределенных систем знания.

В настоящей работе мы рассмотрим основные подходы к созданию онтологий, которые могут применяться в разрабатываемых мультиагентных системах.

Определение понятия онтология

В словаре Брокгауза и Ефона [4] говорится, что "онтология ("ontologia") - вообще учение о сущем. Здесь рассматриваются понятия нечто и ничто, возможное и невозможное, определенное (или действительное) и неопределенное, количество и мера, качество, порядок истина (определяемая формально, как единство многообразного или согласие различного), свойства и условия действительного бытия - пространство, время, дви-

жение, форма, происхождение из другого и переход в другое...". Иммануил Кант писал: "Онтология – это теория обо всех моих концептах, которые могут быть поняты мной только априори" [5].

Как следует из этих определений, в философии онтология – это системный раздел о бытии, т.е. о том, что нас непосредственно окружает (для домашнего кабинета – это стол, компьютер, шкаф и т.д.). Для компьютерных систем "существует" только то, что может быть представлено и, в простейшем понимании, онтология – это способ представления знаний о фрагменте окружающего мира, или знания "как они есть" ("knowledge as it is").

По более формальному определению Тома Грубера [6], впервые применившего это понятие в области информационных технологий, онтология – это "спецификация концептуализации". Концептуализация, связанная с введением абстрактных объектов и других понятий для описания мира – это абстрактный, упрощенный взгляд на мир, который мы хотим представить для некоторой цели.

Под формальной моделью онтологии О часто понимают упорядоченную тройку вида $O = \langle C, R, F \rangle$, где C – конечное множество концептов (понятий) предметной области, которую определяет онтология O ; R – конечное множество отношений между концептами (понятиями) предметной области; F – конечное множество функций интерпретации (аксиоматизация), заданных на концептах и/или отношениях онтологии O [7]. Естественными ограничениями, накладываемыми на множество C являются конечность и не пустота. Что касается множеств R и F , то они могут быть пустыми, что соответствует частным видам онтологии, когда она вырождается в простой словарь, таксономию понятий и т.д.

Таким образом, онтологии на базовом уровне должны прежде всего обеспечивать словарь понятий (терминов) для представления и обмена знаниями о предметной области и множество связей (отношений), установленных между понятиями в этом словаре.

Критерии построения онтологий

Цель создания онтологий – обеспечить поддержку деятельности по накоплению, разделению и повторному использованию знаний. Исходя из этой цели, в [8] введен ряд критериев, которым должна отвечать онтология:

– Прозрачность – онтология должна эффективно передавать подразумеваемое значение определенного термина.

– Связность – онтология должна быть связной, т.е. она должна позволять делать выводы, которые согласуются с исходными определениями понятий. По крайней мере, определяемые аксиомы должны быть логически согласованы между собой.

– Расширяемость - онтология должна быть разработана с возможностью использования разделяемого и пополняемого словаря.

– Независимость от синтаксиса - концептуализация должна быть специфицирована на уровне знания максимально независимо от представления понятий на уровне символов.

– Минимальный базис при высокой выразительности - онтология должна вводить минимальный базовый набор понятий, но их должно быть достаточно, чтобы моделировать мир в требуемых целях и создавать сложные сцены.

Базис для создания онтологий

Для формализации знаний в онтологии необходимо выбрать базис, в котором будут описываться концепты.

В качестве примера одного из таких базисов в [9] предложен следующий набор компонент:

- классы (classes) – обычно организованы в таксономии;

- отношения (relations) – представляют тип связей между концептами предметной области. Формально они определяются как подмножество n множеств таких, что: R: C₁ x C₂ x ... x C_n. Примеры простых бинарных отношений: "быть частью" ("part-of"), "подкласс-класс" ("subclass-of") или "связанный-с" ("connected-to");

- функции (functions) – специальный слу-

чай отношений, в котором n-й элемент отношения определяется по значениям (n-1) предшествующих элементов: F: C₁ x C₂ x ... x C_{n-1} ⇒ C_n. Примеры: функция "Price-of-a-used-car", которая вычисляет цену подержанной машины в зависимости от ее модели, даты выпуска и числа километров;

- аксиомы (axioms) – моделируют предложения, которые всегда истинны. Пример аксиомы: "если что-то сделано из дерева, оно может гореть";

- экземпляры (instances) – представляют элементы. Например, моя конкретная маленькая белая мышь является экземпляром класса "Мышь".

Заметим, что данный базис – отнюдь не единственный, например, в [10] рассматривается вариант "объекты" - "отношения" - "роли" - "атрибуты" при построении онтологий организаций. В работе [11] была предложена и реализована другая модель онтологического базиса: "объекты" - "свойства" - "процессы" - "отношения" - "атрибуты", восходящая к работам Аристотеля, первым обнаружившим, еще задолго до появления первых языков объектного программирования, что "объекты – суть свойства".

Еще одним примером онтологического базиса может служить формальная онтология свойств, представленная в [12], в которой свойство – центральная сущность. Здесь рассматривается в первую очередь проблема формирования правильной таксономической структуры онтологии. Свойство соответствуют узлу таксономии. Формальная онтология основывается на наборе мета-свойств, построенных вокруг философских понятий идентичности (identity), единства (unity), сущности (entity) и зависимости (dependency). Комбинации мета-свойств определяют разновидности свойств такие, как категория, тип, роль, атрибут и прочие. Таким образом, мета-свойства налагают некоторые ограничения на категоризацию и помогают "очищать" таксономии, т.е. прояснить неправильно выделенные концепты (misconcepts), делая их более понятными, облегчая их сравнение и интеграцию. На практике, ответив на простые вопросы о каждом свойстве в иерархии ("Присуще ли

это свойство всем его экземплярам?", "Зависит ли от какого-нибудь другого свойства?" и т.п.), можно определить наборы мета-свойств каждого узла и, соответственно, отнести его к определенной разновидности свойств. Таким образом, строится основная линия наследования – *backbone taxonому*, в которую входят категории, типы и квази-типы. Остальные свойства (различные атрибуты, роли и т.п.) становятся элементами боковых линий. Результатом такой "очистки" в примере, приведенном в [13], стало удаление многих случаев множественного наследования. Это не является обязательной целью, однако, показывается, что множественное наследование часто используется как средство представить нечто большее, чем просто категоризацию. Например, свойство "Страна" было потомком свойств "Место" ("Location") и "Социальная сущность". После первого этапа это понятие было признано не строгим, т.к. страна, например Пруссия, может уже не существовать, в то время как все еще можно поехать на ее географическое местоположение. Но в то же время страна включает в себя два понимания: географический регион (строгое свойство) и политическая сущность (строгое свойство). В итоге же свойство "страна" стала потомком только "свойства" "Социальная сущность", но появилось свойство "Географический регион" - потомок свойства "Место".

Важно отметить, что выбор онтологического базиса приобретает ключевое значение особенно в случае, когда необходимо приобретение новых знаний, например, когда в сети встречаются два агента, имеющие разные онтологии. При наличии единого базиса эти агенты имеют шанс "договориться", поскольку их системы знания базируются на едином основании. В частности, в модели [3] агент в ответ на запрос с указанием на неизвестный объект сможет спросить "что это?" и получить ответ "объект", после чего запросит определения этого объекта и "подцепит" (встроит) его в свою онтологию, получив возможность рассуждать о соответствующих классах и экземплярах объектов.

Формализованное представление знаний

Для формализованного представления знаний наибольшее распространение получили подходы на основе фреймов, семантических сетей и продукционных правил [2]. Отметим, что в уже описанной формальной онтологии свойств используется одна из разновидностей *фреймовой модели представления онтологии*.

Семантические сети являются на сегодня наиболее сложными и наиболее удобными формами концептуализации знаний. Семантические сети состоят из узлов и упорядоченных отношений (связей), соединяющих эти узлы. Узлы выражают понятия или предположения, а связи описывают взаимоотношения между этими узлами.

Для представления знаний обычно используют неоднородные семантические сети [14]. В самом простом случае неоднородную сеть можно представить как совокупность следующих объектов: множества *S* имен предметов и процессов реального мира и семейства *R* отношений на множестве *S*. Различают экстенсиональные и интенсиональные семантические сети. Экстенсиональные сети задаются посредством перечисления всех экземпляров. Интенсиональные сети – это сети, задаваемые посредством обобщающих концептов. Например, вместо отношений "Венди - дочь миссис Дарлинг" и "Майкл - сын миссис Дарлинг" будет введено отношение *быть ребенком* между *детьми и родителями*.

В результате развития подхода к представлению знаний в виде семантических сетей, был разработан ряд методик по их построению и использованию, в числе которых выделим отдельно как наиболее интересную *Динамические семантические сети* (Dynamical Semantic Network – DSN).

DSN, как определяет их Смоленцев С.В. (см. <http://inftech.webservis.ru/it/conference/scm/2000/session8/smolen.htm>), – это метод представления знаний, позволяющий строить сетевые интеллектуальные системы (или сети

интеллектуальных систем). В основу концепции динамической семантической сети положены следующие принципы: интеграция процедурных и декларативных знаний; параллельное функционирование всех компонентов сети; эволюция сети в реальном времени. Динамическая семантическая сеть – это семантическая сеть, у которой каждый узел является выполняемым вычислительным процессом, обладающим: алгоритмами обработки информации и выработки решений; средствами общения с другими узлами сети; определенным поведением, что, как легко заметить, в целом очень близко понятию программного агента.

На основе DSN может быть построена иерархическая DSN, которая представляет собой граф, узлы которого – выполняемые процессы, обладающие набором атрибутов и присоединенных функций. Структура графа отражает текущую модель предметной области. Однако множество дуг графа не ограничивается дугами, отражающими отношения "класс-подкласс", а включает в себя дуги, отражающие отношения "система-подсистема" и ассоциативные связи между узлами. Каждый узел содержит в себе кроме уникальных данных информацию о своей окрестности и множество процедур. Такая технология представляется перспективной для описания современных распределенных и эволюционирующих программных систем.

Нельзя также не рассмотреть и ряд других моделей описания онтологий. Например, наличие развитых алгоритмов и множества стандартных библиотек делает *продукционные модели* также удобными для описания предметных областей. Однако более чем десятилетний опыт разработки экспертных систем показывает, что сложные системы правил продукции плохо поддаются формализации и структурированию.

В последнее время помимо рассмотренного формализованного "явного" представления знаний, все большее распространение получают методы их "неявного" представления, примером которого являются *нейронные сети* [15]. Для нейронных сетей в настоящее время также существует много развитых ал-

горитмов. Нейронные сети и самообучающиеся автоматы являются красивой и удобной моделью для построения самообучающихся агентов, но требуют весьма рутинной процедуры обучения по примерам, что в ряде случаев сделать не представляется возможным; для эффективного обучения требуется большая выборка; даже на уровне отдельных "особей" трудно выделить семантику принимаемых решений.

Поэтому в настоящее время часто применяются комбинированные схемы, соединяющие преимущества явного и косвенного представления знаний. Примером такого подхода является, например, инструмент визуального конструирования знаний CAKE [16], агентная система Agent Builder компании Reticular Systems, которые используют фреймо-продукционную модель. В некоторых случаях применяется также сочетание фреймов и нейронных сетей для представления онтологий (см., например, [17]).

Интеграция знаний от различных экспертов

При интеграции знаний от различных экспертов в рамках единой онтологии возникает новая проблема, связанная с индивидуальной природой знаний у экспертов и неравномерностью распределения знания между группой географически удаленных экспертов. В [18] для решения этой проблемы предлагается использовать технику репертуарных решеток Дж. Келли в рамках теории персональных конструктов. В рамках этой теории считается, что в основе нашего знания о мире лежат базовые категории, которые всегда помогают нам упрощать его восприятие, укладывать знания в определенную, понятную нам систему. Отдельные категории, как прутья, складываются в индивидуальные, групповые и профессиональные структуры (которые могут иметь разные названия – иерархии категорий и ценностей, когнитивные модели, и т.п.), всегда имеющие утилитарные значения, которые и позволяют их называть решетками. Метафора решетки очень удобна, т.к. она позволяет говорить об отдельных элементах, из которых состоит решетка. Описать

модель – значит выявить ту решетку, на которой "приготавливается продукт". Использование техники персональных конструктов позволяет расширить границы когнитивного пространства, в рамках которого живут и эксперт, производящий знания, и рядовой пользователь этого знания.

Интеграция знаний призвана помочь в организации и координации работы в современных "распределенных" компаниях. В таких случаях при создании онтологий ориентируются в первую очередь не на структуру объектов предметной области, а на структуру производственных и бизнес-процессов.

Для описания процессов и координации выполнения производственных операций все чаще используется методология *WorkFlow* (в буквальном переводе – поток работ) [19]. В основе методологии *WorkFlow* лежат следующие понятия: *объект* – информационный, материальный или финансовый объект, используемый в бизнес-процессе; *событие* – внешнее действие, произошедшее с объектом; *операция* – элементарное действие, выполняемое в рамках рассматриваемого бизнес-процесса; *исполнитель*. В рамках этой модели процесс состоит из операций и других процессов. Операция адресуется исполнителям, которые, в свою очередь, отвечают за выполнение одной или нескольких операций бизнес-процесса. Объекты участвуют в выполнении операции. События могут влиять на выполнение операций, например, изменяя результат операций или последовательность их выполнения. Операции обрабатывают события, являясь реакцией системы на внешние действия. Жизненный цикл объекта связан с внешними событиями и операциями, выполняемыми в составе процесса.

Заметим, что для описания такого рода онтологий могут быть использованы и классические *сети Петри* и *алгебры процессов* [20]. Сети Петри – это класс формализмов, основанных на представлении системы состояниями с вершинами двух видов: позициями для представления локальных состояний системы и переходами, соответствующими атомарным действиям. Алгебры процессов – это также класс формализмов для описания и

анализа параллельных систем, при этом процесс описывается с помощью алгебраического выражения над атомарными действиями.

Однако, несмотря на многочисленные теоретические исследования в области формализованного представления знаний, продукции, семантических сетей и фреймов, до сих пор на практике не сложился унифицированный стандарт и готовые для применения библиотеки онтологий, которые бы могли значительно помочь в создании прикладных систем примерно так же как стандартные библиотеки математических функций.

Языки представления онтологий

Язык описания онтологий должен удовлетворять следующим требованиям [21]: быть достаточным для описания концептов предметной области; быть легко используемым на существующих платформах; давать возможность неспециалистам легко строить концептуальные модели.

Опишем кратко структуру некоторых языков, наиболее часто применяемых в этой области, опустив детали синтаксиса.

Knowledge Interchange Format (KIF) – язык обмена знаниями (подробную спецификацию можно найти по адресу: <http://inftech.webservis.ru/it/conference/scm/2000/session8/smolen.htm>). При разработке этого языка пожертвовали наглядностью: наиболее важной была лишь возможность любое представление знаний перевести в KIF, а затем однозначно восстановить или преобразовать с минимальными искажениями в другой формат. В результате KIF позволяет представлять как декларативные, так и процедурные знания, а его синтаксис очень напоминает LISP.

KIF является логическим функциональным языком. Сразу обратимся к примерам. Предложение (material table1 wood) (shape table1 round) означает, что "table1 сделан из дерева и table1 имеет круглую форму". Нетрудно описать, используя эту функциональную нотацию, и более сложные явления. Например, (\Rightarrow (material ?x wood)(can burn ?x)), т.е. "если что-то сделано из дерева, оно может гореть". А фраза: (believe Bob \wedge (shape, table1, round)) значит, что агент Bob верит, что table1

– круглый. Отметим, что знаки ‘^’ и ‘,’ означают, что предложение (shape table1 round) нужно воспринимать не как список, а как единый объект, единое выражение. Тогда конструкция (\Rightarrow (believe Bob ?x)(believe myself?x)) будет значить "я верю во все, во что верит Bob". Таким образом, язык KIF способен описывать и метазнания (знания о знаниях).

Язык KIF обладает широким набором математических функций, логических операций и описателей для формирования метазнаний. Неудивительно, что уже разработана, например, система алгебраических выражений (мир алгебры) в терминах KIF, которая позволяет производить обмен знаниями в области алгебры.

Другой вариацией в области концептуального моделирования является язык Unified Modelling Language (UML). Так, например, в [22] исследуется потенциал объектно-ориентированных стандартов для онтологического моделирования, и в частности описывается язык представления онтологий, основанный на подмножестве UML (универсальный язык моделирования) вместе со связанным с ним языком объектных ограничений OCL (Object Constraint Language).

Большинство же разработок по организации хранения знаний в Интернет в настоящее время основывается на языке XML (eXtended Markup Language). Популярность XML вызвана тем, что этот формат предназначен для описания структур любой степени вложенности. Такие структуры могут быть легко встроены в Web-страницу, т.к. XML очень легко интегрируется с HTML. В настоящее время разработано несколько стандартов расширения этого языка, например:

- XML DTD и XML Schema предназначены для определения структуры документов. С помощью этого языка можно задавать грамматики языков, производных от XML. Одним из таких языков является RDF (Resource Description Framework) – язык для описания ресурсов и связей между ними, рекомендованный W3C для обмена описаниями содержания страниц (см. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>);

- XPath и XPointer – расширения для орга-

низации ссылок между различными объектами и частями документов.

В настоящее время на основе XML и RDF разработано несколько языков, претендующих на то, чтобы быть стандартными языками онтологий:

- SHOE (Simple HTML Ontology Extentions);
- OIL (Ontology Inference Layer);
- DAML (DARPA Agent Markup Language).

Все они предназначены для описания контента HTML-страничек и других объектов Web через установление смысла их содержания и имеют схожие архитектуры.

Рассмотрим структуру таких языков на примере языка OIL (описание языка составлено по материалам <http://www.ontoknowledge.org/oil/downl/dialects.pdf>), созданного исторически после SHOE и использованного как база для определения DAML-Ont (а затем и языка DAML + OIL).

OIL состоит из Standard OIL и Instance OIL, реализует фреймовую модель и содержит конструкции, содержащиеся во многих широко распространенных языках представления знаний.

Standard OIL описывает только декларативную часть онтологии в виде определений классов (class-def) и слотов (slot-def), а так же аксиом.

Определение класса связывает описания класса с его именем (поле name). Класс задается типом (type) – defined или primitive, – последовательностью классов, подклассом которых он является (subclass-of) и набором ограничений на слоты (slot-constraint).

Тип может быть defined или primitive. Первый указывает на то, что определение (т.е. комбинация последовательности предков и ограничений на слоты класса) является критерием для того, чтобы можно было сказать, что объект принадлежит классу. Например, определен defined класс "плотоядное", который является подклассом класса "животное" со слотом "еда". На этот слот наложено ограничение: "еда плотоядного – мясо". Тогда любое животное, у которого значением слота "еда" является "мясо", – "плотоядное". Тип primitive класса определяет ограничения только для этого класса. Т.е. если слон – живот-

ное с серой кожей, это не значит, что все животные, у которых кожа серая, являются слонами.

Очень интересным свойством языка является развитая методика описания таксономии. *Subclass-of* задается в виде списка классовых выражений (*class-expression*). *Class-expression* может быть именем класса, перечислимым классом (*one-of...*), слотовым ограничением или булевой комбинацией перечисленных выше вариантов с использованием операторов *and*, *or* и *not*.

Определения слота связывает имя слота с его описанием. В описании могут быть указаны: список слотов, подслотом которых является определяемый слот; классы, которым принадлежит слот; диапазон значений; слот, обратный к описываемому (слот "есть" является обратным к "быть-съеденным"); свойства (*properties*) – транзитивность, симметричность.

Аксиомы определяют некоторые дополнительные ограничения для классов онтологии. Например: *плотоядные* и *травоядные* являются несовместными (*disjoint*) классами, т.е. нет ни одного экземпляра, который был бы экземпляром сразу этих двух классов. Разрешены следующие аксиомы: *disjoint*, *covered*, *disjoint-covered*, *equivalent*.

Instance OIL добавляет языку Standard OIL возможность определять экземпляры классов и ролей, причем экземпляр может быть экземпляром одного или более классов. Означивание слотов происходит через слотовые отношения (*slot relations*), которые соотносят означаемый экземпляр с другими экземплярами и/или значениями данных.

Программные средства конструирования онтологий

Инструменты поддержки конструирования онтологий в последнее время становятся также все более популярными. При их создании и использовании необходимо помнить о следующих проблемах [21]:

- Поддержка совместных и распределенных работ, поддержка совместимости при работе на разных платформах.
- Масштабируемость, расширяемость,

верификация, безопасность.

- Поддержка анализа, который будет фокусировать внимание пользователя на областях онтологии, где, возможно, необходимы модификации для того, чтобы сделать их более корректными.

- Жизненный цикл онтологии, т.е. наличие инструментов поддержки развития онтологий (например, инструменты слияния, система контроля кода и т.п.).

- Простота использования. В том числе, необходима поддержка простого, но разнообразного пользовательского интерфейса как для неискусшенных, так и продвинутых пользователей, поддержка разных стилей представления информации (текстовый, графический или др.).

В инженерии знаний существует понятие поля знаний, которое позволяет инженеру по знаниям описывать основные объекты предметной области и связи между ними в виде графа, диаграммы или любого другого наглядного представления. Поэтому инструменты визуального проектирования онтологий являются одной из наиболее важных частей инструментов поддержки. Преимущество визуального проектирования состоит в том, что разработчик может не заботиться о языке представления онтологии, а оперировать в терминах понятий, связей и т.п. Как следствие этого удобное графическое представление позволяет легко строить концептуальные модели сложных предметных областей даже неспециалистам.

Кроме того, сам процесс визуального проектирования онтологии является "достаточно эффективным гносеологическим инструментом" [23, 24]), помогая разработчику или исследователю осознавать структуру предметной области, ее основные и производные концепты и отношения (связи) между ними.

И особенно много вопросов при создании инструментов вызывает проблема поддержки *интеграции онтологий*.

В онтологических работах слово "интеграция" используется в различных значениях. Так в [25] обсуждаются различные значения "интеграции" и предлагаются 3 слова для различия этих значений: интеграция

(Integration), слияние (Merge) и использование (Use/Application). Другой взгляд на проблему – исследование по слиянию (merging) и "выравниванию" (alignment) онтологий [26]. Здесь, при слиянии онтологий создается одна согласованная онтология. При выравнивании онтологий создаются связи между онтологиями, позволяя повторно использовать информацию из одной онтологии в другой. Т.е. в отличие от слияния, результат выравнивания – две онтологии с постоянными связями, установленными между ними.

К проблемам интеграции относиться также задача трансляции (передачи) знаний между онтологиями. В [27] предлагается производить трансляции знаний на основе кластеризации ресурсов. В этой архитектуре ресурсы кластеризуются на базе похожести концептуальных моделей их областей (не в одной большой, а во многих, но более маленьких разделяемых онтологиях). Такой подход объясняется предположением, что ресурсы с похожими концептуализациями могут вести более "глубокие" переговоры, чем те, которые меньше разделяют свои концептуальные модели. Эта архитектура удобна для выполнения поддержки и масштабируемости, а также близка к человеческой модели концептуализации.

Онтологические кластеры организуются в иерархическую структуру. В кластере верхнего уровня лежат общие концепты, разделяемые всеми ресурсами. В кластерах более низкого уровня концепты, определенные в кластерах-предках, уточняются и расширяются, чтобы представить локальную концепцию для одного или более ресурсов. Если при построении онтологии какой-то агент "знает" некий концепт, который не разделяют другие агенты, то создается новый кластер. Концепты выражаются в терминах *унаследованных и отличающихся* (*distinguishing*) атрибутов.

Для трансляции концептов из онтологии одного кластера в онтологию другого часто не достаточно просто подняться по иерархии до первого общего кластера и затем спуститься ко второй онтологии, т.к. в процессе обобщения могут потеряться важные детали (например, может быть утеряны отличающие

атрибуты), и, соответственно, смысл понятий исказится.

Для того чтобы сохранить смысл, нужно добавить некоторые ограничения. Трансляцию между двумя локальными онтологиями предлагается получать с помощью последовательных обобщений и специализаций, пытаясь найти первый более общий концепт (*гипернум (hypernum)*) такой, что существует трансляция между ним и концептом в разделяемой онтологии и создавая непосредственно поверх его онтологии концепты, выраженные в терминах разделяемой онтологии, вместе с их отличающими (*distinguishing*) атрибутами. Такая трансляция эффективна только, если концепты исходной и целевой онтологии имеют общего предка, который расположен не слишком высоко от них в иерархии, иначе шаги обобщения могут привести к слишком общему предку. В этом случае информация потеряется, и полученная трансляция может быть пустой. Атрибуты и отношения, связывающие концепты, играют важную роль в процессе обобщения, позволяя избежать потери информации. Фактически они не только дают идентификацию гипернуму концепта (в той же или в разделяемой онтологии), но также позволяют "привязать" некоторую характеристическую информацию к каждому концепту, показывая, таким образом, различие между концептом и его предком.

Специализированные методы и средства построения сложных онтологий, проведения онтологического анализа для интеграции знаний и реализации систем компьютерного моделирования разработаны в работе [28].

Применение онтологического подхода для построения онтологии инженерных знаний на примере машиностроения рассмотрено в работе [29].

Примеры успешных Интернет-применений

В настоящее время существует уже довольно много Интернет-проектов, в основу которых положены онтологии. Рассмотрим некоторые из этих систем.

Ontolingua Server (<http://ontolingua.stanford.edu/>) [30] – это инструмент для созда-

ния, поддержки и использования онтологий, хранимых на сервере. Здесь онтология - это набор определений (на формальном языке) фрагмента декларативных знаний, ориентированный на совместное многократное использование различными пользователями в своих приложениях. В онтологии вводятся термины, типы и соотношения (аксиомы), описывающие фрагмент знания. В системе используется собственный язык *Ontolingua*, который использует принципы объектно-ориентированного подхода и является расширением языка KIF для обмена знаниями и взаимодействия между программами.

Инструмент рассчитан на web-приложения. Это проявляется, в частности, в том, что он имеет HTML-интерфейс для визуального создания и поддержки онтологий.

Онтология по запросу приложения может быть переведена на языки Loom, IDL, CLIPS и некоторые другие. Определены алгоритмы объединения онтологий.

В [31] приведен пример создания крупномасштабной базы знаний в виде онтологии. Взята The CIA's World Fact Book (коллекция географических, экономических, социальных и др. фактов), описанная на SGML. Специально написанный парсер извлекает знания и записывает их в виде KIF-аксиом. В процессе всего этого получают словарь терминов, факты и исключения (куски текста, которые не удалось разобрать). На следующем этапе производится "очистка" знаний и построение таксономии терминов и строятся очевидные функциональные связи терминов. Затем строится онтология (построение классов и отношений между терминами, создание новых KIF-аксиом в этой онтологии на основе аксиом 1-го этапа). Созданная онтология подцеплена к нескольким существующим проектам представления онтологий (DARPA's High Performance Knowledge Base Upper-Level ontology, Ontolingua Units and Dimensions Ontology).

Программа SMART [32], разработанная как часть системы поддержки моделирования знаний *Protégé*, базируется на алгоритме полув автоматического слияния и "выравнивания" онтологий. SMART помогает разработ-

чику онтологий автоматически решать однозначные задачи и "подсказывать" в задачах, в которых возникает неоднозначность. SMART определяет неконсистентность в онтологических утверждениях, возможно из-за действий пользователя, предлагает пути решения. Определено множество базовых операций, применяемых при слиянии и "выравнивании", их действие. SMART опирается на некоторую "всеобщую" модель знаний и потому может быть применен на различных платформах. Алгоритм применен при "выравнивании" онтологий, разработанной как часть проекта DARPA's High Performance Knowledge Base.

Для "желтых страниц" и каталогов продуктов, структурированное онтологическое представление содержания вместе с лингвистической онтологией может повысить и скорость отклика, и точность основанной на содержании возвращаемой информации. Информационно-поисковая система *OntoSeek* [33] разработана для семантически ориентированного поиска информации, комбинируя управляемый онтологией механизм установления соответствия смыслов и системы моделирования. Система *OntoSeek* использует "язык ограниченной выразительности" для представления контента и использует большую онтологию основанную на *WordNet* для матчинга содержания.

Применения в агентных системах

В последнее время онтологии находят все большее применение в различных агентных системах, позволяя отделить знания предметной области от программного кода агента и сделать их повторно используемыми. Здесь мы приведем только два примера, показывающих как использование онтологий агентами, так и использование агентов для построения онтологий.

Система Bee-gent (<http://www2.toshiba.co.jp/beegent/index.htm>), разработанная в Toshiba Corporation, Systems and Software Research Laboratories, является любопытным примером агентной системы с использованием онтологий для организации коммуникации. Bee-gent полностью "агентифицирует"

("Agentifies") коммуникацию, которая происходит между неагентными программными приложениями. Приложения запускают агентов, которые с этого момента поддерживают все сообщения между приложениями. Таким образом, Bee-gent позволяет разработчикам строить гибкие открытые распределенные системы, которые оптимально используют существующие приложения.

В Bee-gent есть 2-х типа агентов. "Агенты-Обертки" ("Agent Wrappers") используются для агентификации существующих приложений, т.е. они управляют состоянием "своих" приложений и вызывают их обработчики при необходимости. "Агенты-посредники" ("Mediation Agents") поддерживают координацию коммуникации между приложениями. Агенты используют онтологии при коммуникации. Когда в переговорах агентов-посредников и wrapper-агентов встречаются не интерпретируемые слова, агент-посредник может запросить перевод слова, т.к. в системе уже подготовлены сервисы онтологий. В Bee-gent сервисы онтологий доступны посредникам, если агенты-wrapper-ы являются клиентами OKBC (Open Knowledge Base Connectivity)

Интересное применение агентных технологий для построения онтологий описано в [34]. В этой системе создана и поддерживается метабаза данных онлайновых ресурсов по экономике (например, электронные публикации) – *RePEc*. Для решения проблемы информационной ориентации пользователей в усиливающемся потоке сведений о результатах исследований для каждого исследователя создается активный информационный робот – сетевой агент. Основной целью активности такого "и-робота" является поиск максимально благоприятных мест в научной системе разделения труда для своего "хозяина" и продвижение результатов "хозяина" среди исследователей, занимающихся близкой тематикой. Одновременная активность множества и-роботов в достижении подобной цели для всех членов некого научного сообщества (например, исследователей-экономистов), порождает процесс формирования онтологии ("тематической структуры предмет-

ной области"), происходит привязка к отдельным элементам онтологии имен исследователей. Отдельный исследователь занимает тем более высокое место в онтологической иерархии, чем большее количество работ, выполненных в данном сообществе, используют (цитируют) его результаты.

Заключение

В настоящей работе рассмотрена проблема описания семантики информационных ресурсов в сети Интернет на основе онтологического подхода. Онтологии позволяют разработчикам программных систем создавать семантические описания ресурсов, строить модели мира и обеспечивать возможность коммуникации для программных агентов в мультиагентных системах, а правильно спроектированные онтологии позволяют повторно использовать представленные в них знания и сокращать сроки проектирования систем.

Однако на пути внедрения онтологического подхода существует много проблем. Одна из них – отсутствие общего стандарта на описание и представление онтологий, с чем связано и отсутствие универсальных инструментов для создания и поддержки онтологий, и библиотек готовых решений. Отсутствие стандарта влечет и проблемы интеграции гетерогенных онтологий. Кроме того, существующие инструменты не всегда доступны эксперту и требуют привлечения программистов для создания полноценного продукта.

В этой связи в настоящее время важно разработка нового поколения инструментов и методик представления онтологий, которые бы обеспечивали:

- полноценный онтологический базис для представления понятий и отношений любой предметной области;
- конструирование онтологий с помощью удобных визуальных средств, которые были бы доступны экспертам и пользователям системы;
- представление гетерогенных знаний в виде смешанных моделей (комбинации семантических сетей, продукции, сценариев);

- динамическое обновление и структуризацию онтологии при одновременной работе многих экспертов;
- интеграцию разрабатываемых инструментальных средств с наиболее распространенными информационными технологиями (сервера Web, сервера приложений, сервера реляционных баз данных);
- интеграцию с другими системами на базе формирующихся в настоящее время de-facto стандартов XML.

СПИСОК ЛИТЕРАТУРЫ

1. *Berners-Lee T., Hendler J., Lassila O.* The Semantic Web // Scientific American, May 2001.
2. *Huhns M. N., Singh M. P.* Ontologies for Agents // IEEE Internet Computing, November–December 1997.
3. *Batishev S.V., Ivkushkin C.V., Minakov I.A., Rzevski G.A., Skobelev P.O.* MagentA Multi-Agent Systems: Engines, Ontologies and Applications // Proc. of the 3rd Intern. Workshop on Computer Science and Information Technologies CSIT'2001, Ufa, Russia, 21-26 September, 2001. Ufa State Aviation Technical University. Institute JurInfoR-MSU. Vol.1: Regular Papers. 2001.
4. Брокгауз Ф.А., Ефрон И. Энциклопедический словарь в 86 т. СПб., 1890-1907 гг.
5. Kant I. Lectures on metaphysics - Part III. Metaphysik L2 (1790-1791) // Translated and edited by Karl Ameriks and Steve Naragon. Cambridge University Press, 1997.
6. Gruber T. The role of common ontology in achieving sharable, reusable knowledge bases // Proceedings of the Second International Conference, 1991.
7. Гавrilова Т.А., Хорошевский В.Ф. Базы знаний интеллектуальных систем. СПб.: Питер, 2000.
8. Gruber T. Toward Principles for the Design of Ontologies // International Workshop on Formal Ontology, March. 1993. Padova. Italy.
9. Gomez-Perez A., Benjamins V.R. Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods // Proc. of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm. Sweden. August 2. 1999.
10. Виттих B.A. Онтологии сложных систем – организаций // Препринт ИПУСС РАН. Самара. 1998. № 3.
11. Андреев В.В., Ивкушкин К.В., Минаков И.А., Ржевский Г.А., Скобелев П.О. Конструктор онтологий для разработки мультиагентных систем // Труды 3-ей Международной конференции по проблемам управления и моделирования сложных систем, Самара, 4-9 сентября 2001. Самара: СНЦ РАН, 2001.
12. Guarino N., Welty Ch. A Formal Ontology of Properties // Proc. of 12th Int. Conf. on Knowledge Engineering and Knowledge Management, Lecture Notes on Computer Science. © Springer Verlag. 2000.
13. Guarino N., Welty Ch. Ontological Analysis of Taxonomic Relationships // Proceedings of ER-2000: The International Conference on Conceptual Modeling. October, 2000.
14. Осипов Г.С. Приобретение знаний интеллектуальными системами: Основы теории и технологии // Проблемы искусственного интеллекта. М.: Наука. Физматлит, 1997.
15. Grand S., Cliff D. Creatures: Entertainment Software Agents with Artificial Life // Autonomous Agents and Multi-Agent Systems 1(1). 1998.
16. Воинов А.В. Моделирование интуитивных рассуждений эксперта методами психосемантики и вывода с неопределенностью // Новости искусственного интеллекта. М., 1998. №2.
17. Леонтьев А.А., Леонтьев Л.А. Способы представления знаний и их компьютерное моделирование // Обработка текста и когнитивные технологии / Под ред. В.Д. Соловьева. Пущино: 1999. Вып.3.
18. Травина Л.Л. Извлечение знаний у эксперта, основанное на технике репертуарных решеток Дж. Келли, в пространстве Всемирной паутины // Обработка текста и когнитивные технологии / Под ред. В.Д. Соловьева. Пущино: 1999. Вып.3.
19. Уринцев А.И. Об использовании бизнес-объектов в рамках создания ЭИС // Корпоративные системы / М.: Моск. госуд. ун-т экономики, статистики и информатики, 2000.
20. Зазовская А.А., Ломазова И.А. О сравни-

- тельной выразительности вложенных рекурсивных сетей Петри и алгебры процессов // Труды 7-ой Национальной конференции ИИ-2000, 24-27 октября 2000, Переславль-Залесский, Россия. М., 2000. Т.1.
21. McGuinness D.L. Ontologies Come of Age // The Semantic Web: Why, What, and How, MIT Press, 2001.
22. Cranefield S., Purvis V. UML as an Ontology Modelling Language // Proceedings of the IJCAI-99 Workshop on Intelligent Information Integration
23. Gavrilova T., Voinov A., Vasilyeva E. Visual Knowledge Engineering as a Cognitive Tool // Proc of International Work Conference on Artificial and Neural Networks IWANN'99. Spain. 1999.
24. Гаврилова Т. Визуальное проектирование баз знаний как гносеологический инструмент // Труды 7-ой Национальной конференции ИИ-2000, 24-27 октября 2000, Переславль-Залесский, Россия. М., 2000. Т.1.
25. Pinto H. S., Gomez-Perez A., Martins J. P. Some Issues on Ontology Integration // Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August 2, 1999.
26. Noy N.F., Musen M.A. An Algorithm for Merging and Aligning Ontologies: Automation and Tool Support // AAAI-99 Workshop on Ontology Management. Also, SMI Technical Report SMI-99-0799.
27. Visser P., Tamma V. An Experience with Ontology-based Agent Clustering // Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5) Stockholm, Sweden, August 2, 1999.
28. Смирнов. Онтологический анализ предметных областей моделирования // Известия Самарского научного центра РАН, 2002. Т.3. №1.
29. Евгенев. Онтология инженерных знаний // Информационные технологии. 2001. №6.
30. Farquhar A., Fikes R., Rice J. Ontolingua Server: a Tool for Collaborative Ontology Construction// International Journal of Human-Computer Studies 46, 707-727, 1997.
31. Farquhar A., Fikes R., Rice J. Building a large Knowledge Base from a Structured Source: The CIA World Fact Book // IEEE Intelligent Systems. January/February. 1999. Vol.14. №1.
32. Noy N.F., Musen M.A. SMART: Automated Support for Ontology Merging and Alignment // Twelfth Banff Workshop on Knowledge Acquisition, Modeling, and Management; Banff, Alberta, Canada; 1999. Also, SMI Technical Report SMI-1999-0813.
33. Guarino N., Masolo C., Vetere G. OntoSeek: Content-Based Access to the Web // 70 1094-7167/99, IEEE INTELLIGENT SYSTEMS.
34. Krichel T., Левин Д., Паринов С. Активный информационный робот как сетевой агент исследователя (на примере сети онлайновых ресурсов по экономике RePEc/RuPEc) // Труды Первой Всероссийской научной конференции. СПб., 1999.

ONTOLOGY DEVELOPMENT METHODS AND TOOLS FOR INTELLECTUALIZATION OF INTERNET RESOURCES

© 2002 S.V. Batishchev¹, T.V. Iskvarina¹, P.O. Skobelev²

¹Samara Branch of Physics Institute named for P.N. Lebedev of Russian Academy of Sciences
²Institute for the Control of Complex Systems of Russian Academy of Sciences, Samara

The problem of Internet resource intellectualization and "Semantic Web" development is analyzed. New generation of Internet applications (based on the concept of "ontology" for semantic description of Internet resources and cooperative activities) are overviewed. Different approaches to "ontology" concept definition are compared. Ontology development criteria and basis of primary concepts are given. Methods for formal ontology definition and integration (together with relevant programming language tools for ontology description) are described. Software tools requirement are proposed. Examples of successful implementations of different ontology-based web projects are given.

МЕТОД КОМПЕНСАЦИЙ ДЛЯ ПОДДЕРЖКИ ПРОЦЕССОВ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ДИНАМИЧЕСКОМ РАСПРЕДЕЛЕНИИ РЕСУРСОВ

© 2002 П.О. Скobelев

Институт проблем управления сложными системами РАН, г. Самара

Рассматривается метод компенсаций для поддержки принятия решений о динамическом распределении ресурсов в системах, функционирующих в реальном времени. Предлагается создание мультиагентных систем с миром агентов заказов и ресурсов, способных оперативно воспринимать ситуацию, принимать решения, коммуницировать и устанавливать связи между собой. Для принятия и пересмотра решений агенты получают возможность предлагать компенсации, что приводит к установлению нового баланса интересов за счет перераспределения прибыли между вновь приходящими и имеющимися агентами системы. Рассматриваются достоинства и ограничения метода и показывается возможность его применения для широкого круга приложений.

Введение

Традиционно проблема распределения ресурсов рассматривается и решается в случае, когда все ресурсы и потребности в этих ресурсах известны заранее. Однако существующие рыночные условия настолько нестабильны, что, как правило, спрос на ресурсы и предложение ресурсов подчиняются весьма изменчивым и непредсказуемым процессам, а сами ресурсы и заказы могут появляться и исчезать в заранее непредвиденные моменты времени. При таких условиях алгоритмы размещения и реализации заказов должны включать в себя возможности для частого и быстрого перераспределения ресурсов, с чем плохо справляются обычные подходы [1-3].

Динамическое распределение ресурсов является проблемой, при которой ресурсы, которые необходимо разместить и/или спрос на эти ресурсы непредвиденным образом изменяются в процессе принятия решений. Мультиагентные системы [4-5] могут рассматриваться как основа для создания более эффективных методов и средств для динамического распределения и перераспределения ресурсов при постоянно изменяющемся спросе и предложении. *Агент* является программным объектом, обладающим способностью автономно существовать и "разумно" общаться с другими объектами, что предполагает наличие возможностей создавать со-

общения; посылать их другим агентам или людям; интерпретировать входящие сообщения; действовать над ними некоторым образом. *Мультиагентная система* (MAC) рассматривается как система, состоящая из агентов *соревнующихся или кооперирующих* между собой, в зависимости от ситуации, с целью дать возможность системе выполнить поставленную задачу.

При этом главной особенностью предлагаемого в статье подхода становиться самоорганизация заказов и ресурсов, при которой вместо централизованного решения задачи и построения полностью оптимального плана действий, каждый заказ и каждый ресурс должен самостоятельно принять решение, какой из возможных вариантов взаимодействия ему подходит, и изменить свое принятые ранее решение, если находится вариант для его улучшения.

Под *самоорганизацией* при этом понимается возможность системы автономно видоизменять уже существующие и/или устанавливать новые связи между ее компонентами с целью повышения значения критерия эффективности ее существования или восстановления после повреждения. В контексте распределения ресурсов любое автономное изменение *связи заказ-ресурс* рассматривается как шаг в самоорганизации.

Примеры различных областей, где требуется постоянное динамическое распределение

ние ресурсов, включают:

1. *Производственная логистика* – установление места и времени нахождения материалов или компонентов на производственных мощностях (сборочные устройства, конвейеры, металлообрабатывающие станки и т.д.) в условиях неопределенности, создаваемой постоянными изменениями спецификаций производимых продуктов и поведением поставщиков и потребителей.

2. *Планирование Работы Персонала* – распределение задач для персонала в условиях быстро развивающегося рынка.

3. *E-комерция* – распределение доступных товаров и услуг (предложение) в соответствии с требованиями клиентов (спрос), когда клиенты и/или поставщики неожиданно появляются или покидают торговую площадку.

4. *Управление Знаниями* – распределение записей по кластерам в ситуациях частого обновления баз данных.

5. *Понимание текста* – размещение (приписывание) смыслов (значений) к словам в процессе непрерывного диалога между человеком и машиной

Мир заказов и ресурсов МАС для поддержки процессов принятия решений

Для поддержки процессов принятия решений в мультиагентных системах строится специальный виртуальный мир "ресурсов-заказов", в котором агенты заказов и ресурсов могут динамически устанавливать и разрывать связи между собой [6-7]. При этом агенты заказов ищут наилучшие варианты реализации своих заданий с наилучшим качеством, минимальной ценой и в заданный срок, а агенты ресурсов стремятся повысить коэффициент своего использования и, наоборот, получить максимально возможную прибыль. В результате агенты заказов и ресурсов составляют диалектическую пару противоположностей, которая находится в постоянной "борьбе и единстве": агенты заказов могут конкурировать за агента некоторого ресурса или наоборот агенты ресурсов могут конкурировать между собой за агента некоторого заказа. С другой стороны, несколько агентов ресурсов могут скооперироваться для выполн-

нения некоторого сложного заказа, равно как и несколько агентов заказов могут скооперироваться для уменьшения цены некоторого ресурса. Наконец,шедшие друг друга заказы и ресурсы, вступившие по обоюдному согласию в связь между собой, представляют собой временно образование, существующее до момента исполнения некоторого проекта (работы), которое может иметь довольно сложную иерархическую структуру, поскольку ресурсы для обеспечения своей жизнедеятельности могут в свою очередь генерировать новые заказы на субподрядные работы и т.д.

Виртуальный мир предприятия состоит из агентов заказов и ресурсов, способных взаимодействовать между собой. Состояние агентов, зафиксированное в мире в некоторый момент времени, будем называть сценой мира. Агенты в сцене виртуального мира могут быть свободными (располагать ресурсом и не иметь связи), и в этом состоянии они активно ищут связи с другими агентами, постоянно осуществляя поиск взаимного соответствия (матчинг) с потенциальными партнерами (как мужские и женские особи, разноименные заряды и т.д.). Для этого они находят потенциальных кандидатов, сопоставляют варианты по методу четкого (полного) или нечеткого матчинга (частичное совпадение свойств), принимают решения о выгодности того или иного варианта и делают предложения выбранному партнеру. Однако связь устанавливается только в том случае, если партнер также согласен и для него данный вариант является лучшим из всех имеющихся в его положении в данный момент времени. При этом для агента устанавливается состояние "доволен" (что будет отображаться на рисунках далее выражением его символического "лица"), если ценность его варианта выше, чем например средний показатель удовлетворенности агентов на рынке, и "не доволен", если наоборот, ниже этого показателя (для нечеткого матчинга это также может быть степень соответствия агентов друг другу – полностью или частично совпало три свойства из пяти, или же все пять). Степень неудовлетворенности агента является одним из стимулов к пересмотру вариантов.

Если свободный активный агент исследовал весь рынок и обнаружил, что для него нет никаких возможностей для установления связи, он переходит в состояние покоя или бездействия ("сна") и будет ожидать прихода сообщения с предложением от другого агента или появления любого другого важного события, связанного с каким-либо изменением в сцене (когда такое происходит, агент "пробуждается" и становится активным, а вновь "успокаивается" и "засыпает", когда вся цепочка переговоров других агентов, вызванная некоторым событием, например, приходом нового пользователя, затихает). Кроме того, в состояние покоя переходят и агенты, установившие отношения с другими агентами, которые заводят "будильник" и ждут инициализации по таймеру до момента возобновления своей активности – для агента детали в логистике, например, когда придет срок ее поставки на конвейер. В случае "пробуждения" по сообщениям к классу агентов в первую очередь будятся свободные агенты, во вторую – недовольные агенты, и в третью – довольные (для возможного пересмотра и их вариантов по методу компенсаций).

Рассмотрим взаимодействие агентов на простом примере (рис.1). Здесь имеются 7 агентов, которые пытаются осуществить матчинг и установить отношения между собой. Здесь белым цветом выделены агенты заказов, а серым – ресурсов. Видно, что два агента слева нашли друг друга, удовлетворены условиями сотрудничества (улыбаются) и установили отношение (непрерывная стрелка), три агента в группе в центре рисунка также удовлетворены условиями (улыбаются) и находятся в стадии принятия решения и установления отношений (пунктирная стрелка), а группа справа – в стадии принятия решения (пунктирная стрелка), но не удовлетворены условиями (грустят), возможно, что и вовсе не подходят друг другу и вынуждены будут расстаться. Обратим внимание, что здесь также показана возможная активность с каждой из сторон, как от заказов к ресурсам, так от ресурсов к заказам (направление стрелки).

В мире "заказов-ресурсов" основные действия агентов связаны с процедурами мат-

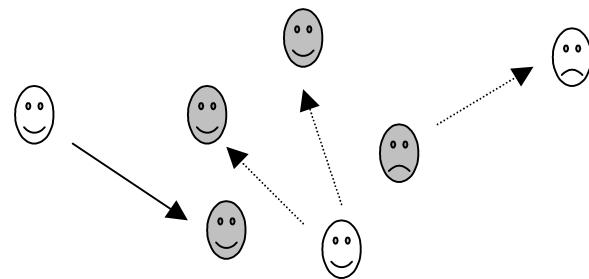


Рис.1. Сцена простейшего виртуального мира:
матчинг агентов заказов (белые)
и ресурсов (серые)

чинга (поиска соответствия между собой) и принятия или пересмотра решений, установления и разрыва связей. Для этого также агенты должны иметь возможность запрашивать друг у друга значения необходимых параметров, находить в онтологии правила принятия решений и методы вычислений значений сложных параметров, проводить требуемые расчеты значений этих параметров, а также создавать вспомогательных агентов заказов или ресурсов, если это необходимо.

Метод компенсаций для динамичного распределения ресурсов

Предположим, что мультиагентная система решает задачу постоянного размещения производственных ресурсов по заказам. При этом заказы прибывают в систему одна за другим и время прибытия заказов и их характеристики непредсказуемы, более того, некоторые заказы могут отзываться из системы. Также и ресурсы могут оставаться постоянными или изменяться и даже исчезать из системы в силу потери или поломки.

Каждый ресурс R характеризуется:

- рядом характерных черт f (тип изделия, материал и т.д.);
- его цена c , которая выражается в условных денежных единицах (условных единиц - у.е.) и назначается самим ресурсом в зависимости от ситуации на рынке;
- себестоимость ресурса s (не изменяется, причем $c > s$).

Прибыль ресурса $Profit(R)$ рассматривается как сумма прибылей от выполнения каждого заказа $Profit(R) = \sum(c(O_i) - s(O_i))$, где $c(O_i)$ – стоимость ресурса для заказа O_i , а $s(O_i)$ – себестоимость выполнения этого заказа.

Накапливая прибыль, ресурс может принимать заказы и по цене, меньше чем себестоимость.

Каждый заказ O характеризуется:

- набором характерных черт g (желаемый тип изделия, материал и т.д);
- его покупательная способность r выражается денежными единицами (у.е.);
- объемом средств r , которые он может потратить на получение ресурса в данный момент времени, $r < p$.

Прибыль заказа $Profit(O)$ определяется как сумма прибыли от каждого подзаказа, т.е. разница между объемом средств, имеющихся на покупку данного ресурса и его текущей ценой $Profit(O) = \sum (r(Ri) - c(Ri))$, где $r(Ri)$ – резерв на покупку ресурса Ri для данного заказа, а $c(Ri)$ – принятая заказом цена данного ресурса.

Пусть каждый заказ и ресурс имеют определенные критерии для принятия решений, например, максимизируют свою прибыль. Тогда правило принятия решения, т.е. выбора ресурса со стороны заказа, есть правило выбора предложения с минимальной ценой $Min(c)$, а правило для ресурса, в случае, когда для него есть много предложений – как выбор заказа с максимальным предложением $Max(r)$. Принятие решения при этом будет простейшим, состоящим в том, чтобы выбрать минимум или максимум по одному из параметров. Заметим, что в дальнейшем агенты могут применять и сколь угодно более сложные критерии и методы оптимизации, если для этого есть время. Главная проблема, решаемая в настоящем подходе, состоит не в том, как именно принять оптимальное решение (проблема выбора – лишь третий этап), а в том, чтобы реализовать за заданное время все основные этапы процесса принятия решений, о которых писал Саймон [8].

Введем теперь общий критерий $Value(i)$, отражающий "общественную ценность" варианта для системы на i -ом шаге. Этот критерий призван частично погасить "эгоистические" устремления агентов, локально оптимизирующих свою прибыль без учетов своего окружения. В простейшем случае, это может быть, например, общая прибыль системы, или скажем срок, на который станок должен

быть выведен из оборота для его ремонта или регламента. Агенты смогут устанавливать новые связи и принимать или пересматривать свои решения только в том случае, если $Value$ не уменьшается, а увеличивается, т.е. $Value(i) > Value(i+1)$.

Общий алгоритм метода компенсаций для динамического распределения ресурсов тогда выглядит следующим образом:

1. Как только заказ входит в систему, ему назначается агент. Агент заказа посыпает сообщение всем агентам, отвечающим за имеющиеся в наличии ресурсы, заявляя, что ему необходим такой-то ресурс с определенными характеристиками g , и что он может заплатить за этот ресурс определенное количество денежных единиц r .

2. Все агенты, обладающие ресурсами с указанными свойствами g , или лишь несколькими из указанных свойств (в случае частичного матчинга), и со стоимостью, которая меньше или равна названному количеству денежных единиц, начинают предлагать эти ресурсы Агенту потребности, т.е. $r > c$ и при этом каждый ресурс выбирает предложение с $Max(r)$.

3. Из получаемого множества предложений, Агент заказа выбирает наиболее подходящий ему ресурс в соответствие с заданными ему критериями и методом принятия решений, например, определяя $Min(c)$.

4. Если подходящего ресурса нет в наличии, Агент заказа старается заполучить в свое распоряжение ресурс, который был уже отдан другому заказу, предлагая тому заказу компенсацию, максимальный размер которой k не превосходит $p - c$.

5. Агент заказа, которому предлагается компенсация k , рассматривает это предложение. Компенсация принимается только в том случае, если она позволяет заказу приобрести другой устраивающий его ресурс, для чего необходимо новое резервирование ресурсов на сумму, и в то же время повысить общую ценность системы (например, итоговую прибыль системы). Если предложение принято, Агент заказа реорганизует связи – ранее установленные отношения между заказом и ресурсом разрушаются, а новые отношения между другим заказом и свободным ресурсом

устанавливаются, тем самым, повышая общую ценность системы.

6. Процесс, описанный выше, повторяется до тех пор, пока все ресурсы не становятся связанными с заказами, и никакое дополнительное изменение во владении ресурсом не сможет более увеличить ценность системы *Value* или до тех пор, пока не закончится время, отведенное для распределения.

7. Если заказ или ресурс удаляются из системы, они посылают сообщение и разрывают связи, освобождая партнера от обязательств, вследствие чего освободившийся партнер начинает заново осуществлять поиск вариантов своей реализации.

Этот метод реализует важные принципы самоорганизации, позволяющие децентрализованным образом осуществлять динамическое распределение ресурсов в любой открытой системе, в которой имеется постоянный приток новых заказов, от транспортной и производственной логистики – до понимания смысла фраз предложений на естественном языке.

Пример применения метода в транспортной логистике

Рассмотрим пример организации перевозок грузов транспортными средствами с использованием механизма самоорганизации и варианта переговоров с взаимными компенсациями. Пусть имеется некоторая сеть городов и дорог, а также два транспортных средства, осуществляющие перевозки в данной сети (рис.2). Один из этих перевозчиков является более мощным (имеет грузоподъемность 100 тонн) и более тихоходным (скорость 10 км/час), а второй - менее мощным, но более скоростным (с грузоподъемностью 10 тонн при скорости 15 км/час). Соответственно стоимость перевозки первым перевозчиком равна 7 условных единиц за 1 км, а вторым – 3 условные единицы.

Ситуация 1.

Пусть 15 сентября в систему приходит первый заказ на перевозку груза в 5 тонн из пункта А в пункт D на расстояние 500 км с ценой 15 000 условных единиц и дедлайном 20 сентября. Предположим, что перевозчик 1 в текущий момент уже свободен и уже на-

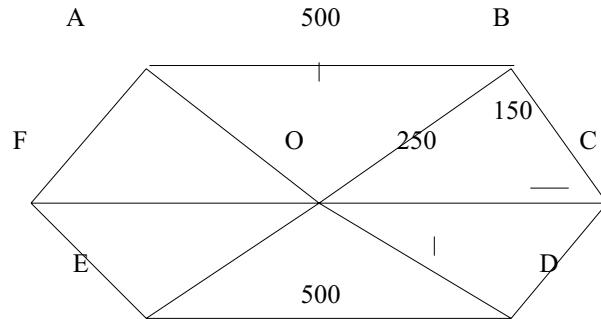


Рис.2. Сеть городов и дорог
(числа – условные расстояния)

ходится в пункте Е, а перевозчик 2 завершает исполнение принятого ранее заказа в пункт С.

Для планирования исполнения этого заказа создается Агент заказа 1. Этот Агент посыпает запрос на перевозку указанного груза обоим перевозчикам. В ответ каждый из перевозчиков рассматривает запрос и посыпает Агенту заказа свои предложения.

Первый перевозчик готов приступить к исполнению запроса немедленно, он свободен и ждет заказов. Для этого перевозчика предварительно потребуется перемещение из пункта Е в пункт А (300 км). Стоимость исполнения заказа при этом составит $(300 + 500) \cdot 7 = 5600$ у.е., а срок исполнения – $(300+500)/10=80$ часов, т.е. примерно 4 суток.

Второй перевозчик также вполне успевает прилететь в пункт А и исполнить данный заказ, но для этого ему потребуется сначала доисполнить текущий заказ. Если в данный момент перевозчик 2 находится в пункте О, то ему потребуется пройти маршрут О – С – О – А и его срок исполнения заказа составит $(250 + 250 + 250 + 500)/15=83,3$ часов,

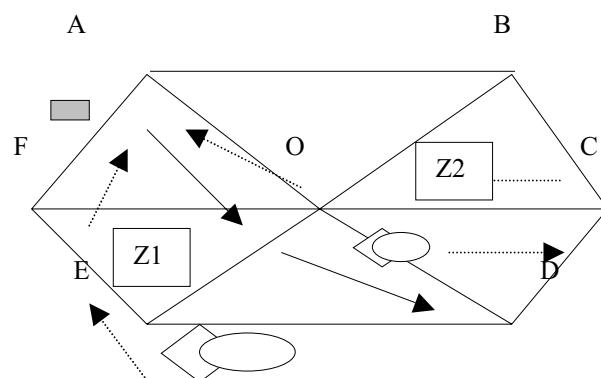


Рис.3. Начальное состояние сети и маршруты перемещения дирижаблей в пункт исходного расположения груза

т.е. также примерно 4 суток, при цене $(250 + 250 + 500) \cdot 3 = 3000$ у.е. Прибыль в случае второго перевозчика составит 12000 у.е., а в случае первого - 10 400 у.е.

Очевидно, что по критерию прибыльности Агент заказа 1 выберет второго перевозчика, хотя ему и дальше лететь к пункту А. Заметим, что в случае другого критерия, например, по критерию равномерности распределения перевозчиков по сети или их средней загрузке или необходимости своевременного профилактического осмотра у второго перевозчика, мог бы быть выбран и первый перевозчик.

Таким образом, каждый из перевозчиков рассмотрел ситуацию для себя и рассчитал свою определенную цену и сроки в зависимости от своих параметров, и второй перевозчик, который более легкий и компактный, предложил и лучшую цену. Агент заказа, в свою очередь, принял это предложение с учетом своих требований и предложенной наилучшей цены. В результате между заказом и перевозчиком (ресурсом) в системе возникла первая связь (рис.4), обусловленная бронированием перевозчика для исполнения первого заказа.

Ситуация 2.

Предположим теперь, что через час в систему пришел новый срочный заказ на транспортировку груза, в соответствие с которым требуется перевести груз весом 8 тонн из пункта С в пункт В на расстояние 150 км с ценой 20 000 условных единиц и дедлайном 17 сентября.

В этой ситуации из переговоров агентов быстро выясняется, что только второй перевозчик может выполнить заказ. Действительно, первому потребуется $(500 + 300)/10 = 80$ часов, т.е. примерно 3,5 суток, что больше срока дедлайна. Второму же перевозчику нужно будет $(250 + 300)/15 = 36,6$ часов, т.е. около 1,5 суток. Однако перевозчик 2 уже забронирован заказом 1. В данной ситуации обращение заказа 2 к перевозчику 2 вызывает "волну" (цепочку) переговоров: перевозчик 2 должен обратиться к заказу 1 с предложением разрешить ему реализовать заказ 2 и определить размер компенсации, который он должен получить за такое разрешение, который покро-

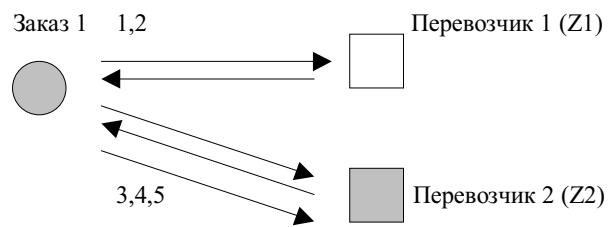


Рис.4. Обнаружение заказом двух перевозчиков (ресурсов) и бронирование второго перевозчика для транспортировки груза: 1 – запрос к первому перевозчику, 2 – ответ и предложение от первого перевозчика, 3 – запрос второму, 4 - ответ и предложение от второго, 5 – выбор варианта второго перевозчика и его бронирование

ет все его расходы на бронирование первого и бронирование второго вариантов перевозок, если таковой найдется.

Для этого заказ 1 вновь обращается на рынок перевозчиков со своим предложением – и в ответ получает уже рассчитанное ранее предложение от перевозчика 1, которое вполне устраивает заказ 1 по срокам исполнения. Вместе с тем, прибыль по данному варианту оказывается существенно меньше той, что была ранее (10400 против 12000). Резонно, что заказ 1 попросит у заказа 2 по возможности компенсировать разницу (1600), что заказ 2 вполне может сделать, поскольку его прибыль от перевозки груза составляет $20\ 000 - (250 + 300) \cdot 3 = 18\ 350$. Заметим, что здесь запрошенный размер компенсации первому заказу существенно меньше, чем прибыль второго заказа – если бы эти суммы были бы сопоставимы, то могла бы быть задействована любая другая стратегия (например, деления общей прибыли пополам или пропорционально стоимости заказов и т.д.). В результате достижения "договоренности" устанавливается новый баланс интересов и происходит разбронирование перевозчика 2 и бронирование перевозчика 1 для первого заказа. Теперь перевозчик 2 получает возможность исполнить заказ 2.

Ситуация 3.

Рассмотрим теперь несколько иную ситуацию. Пусть первый перевозчик исполняет полученный заказ А – D (и пройдено уже 100 км), а второй перевозчик только что освободился в точке В (сразу заметим, что в условиях быстрых изменений планировать на

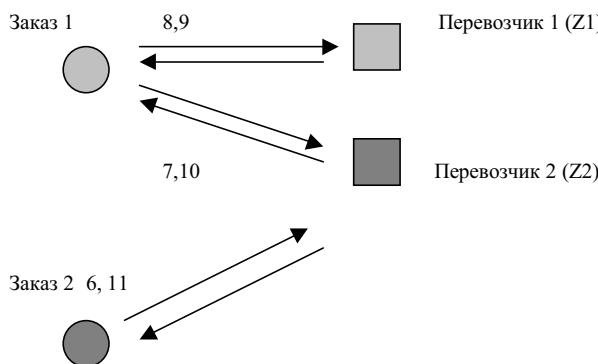


Рис.5. Приход второго заказа и перебронирование второго перевозчика для второго заказа с выплатой компенсации: 6 – запрос ко второму перевозчику, 7 – переадресация предложения первому заказу, 8 – запрос к первому перевозчику, 9 – положительный ответ и предложение, 10 – разрешение на перебронирование второго перевозчика, 11 – согласие и предложение с суммой компенсации первому заказу

большой период времени вперед не эффективно – поэтому агенты в данном примере планируют свою загрузку на один контракт (шаг вперед). Никаких новых заказов пока нет.

Освободившийся ресурс (перевозчик 2) немедленно начинает искать новые возможности для приложения своих сил. Теперь уже он проявляет инициативу и активность и посыпает запрос всем заказам (в данном случае заказу 1): есть ли возможность еще повысить его прибыль? В ответ заказ 1 обнаруживает точку О возможной пересадки и посыпает запрос на перевозку из О в D для второго перевозчика.

Второй перевозчик делает свои расчеты по перевозке груза первого перевозчика из О в D. Такая перевозка будет стоить $(250 + 250) \cdot 3 = 1500$ у.е. Та же перевозка на первом перевозчике потребует $250 \cdot 7 = 1750$ у.е. Если перегрузка ничего не стоит, то можно сэкономить 250 у.е., хотя придется гнать второй

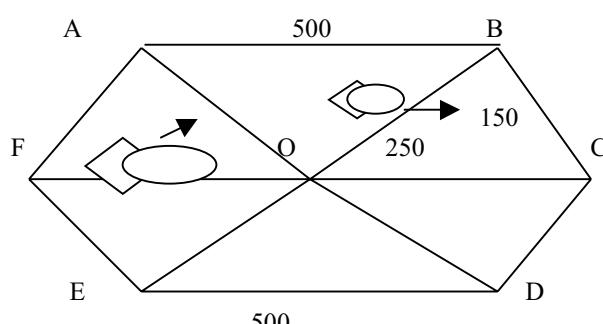


Рис.6. Новое начальное положение дирижаблей

перевозчик лишних 250 км. Вариант принимается и общая ситуация компании улучшается еще на 250 долларов – хотя первый дирижабль при этом "лично" страдает в угоду повышения общей прибыли компании (в определенных случаях полученная дополнительная прибыль может быть и поделена между перевозчиками, отчасти компенсируя первому перевозчику лишение части заказа).

Выводы и заключение

Рассмотренный метод предназначен для построения мультиагентных систем массового обслуживания различных заказов в режиме реального времени, причем в условиях высокой неопределенности и большой динамики изменений, когда на получение полностью оптимальных решений не достаточно времени. Предполагается, что в любой такой системе непрерывно происходят следующие важные изменения, осуществляемые в заранее не предсказуемые моменты времени:

- приходит новый заказ, для которого требуется найти подходящие ресурсы, удовлетворяющие как заказчика, так и всех имеющихся исполнителей,
- появляется новый свободный ресурс (причем вводится новый или забронированный ранее вдруг становится свободным),
- "на ходу" меняются цели, предпочтения или ограничения заказчика, например, пользователь может переменить свой приоритет с поиска вариантов минимальной цены на скорейший срок получения товара,

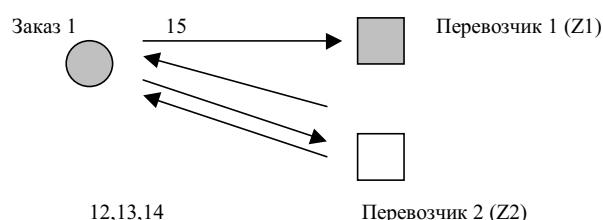


Рис.7. Освобождение ресурса также вызывает пересмотр ситуации и перепланирование ресурсов в ходе исполнения заказа: 12 – освободившийся перевозчик инициирует выполняющийся заказ, 13 – выясняется, что перегрузка на второго перевозчика в точке О выгодна для первого заказа, 14 – дается подтверждение и бронирование второго перевозчика для перевозки с точки О груза в D, 15 – первому перевозчику дается извещение об изменении маршрута в точке О и необходимости перегрузки груза на второго перевозчика

• исчезает ранее забронированный или свободный ресурс (ресурс неожиданно выводится из системы или ломается и выходит из строя и т.д.),

• отменяется существующий, возможно уже даже запущенный в исполнение, заказ,

• изменяются критерии функционирования системы в целом (например, для сети производства автомобилей критерий получения максимальной прибыли на несколько месяцев может уступить место критерию высвобождения одной из линий конвейера для проведения ее ремонта и модернизации).

Суть предлагаемого метода состоит в том, чтобы ускорить процесс нахождения варианта за счет выбора первого подходящего варианта, повышающего общую ценность системы – но при сохранении возможности пересмотра ранее принятых решений при возникновении новых ситуаций, в которых возникает конфликт интересов между новыми и старыми заказами (решениями). Это означает установление некоторой структуры отношений между агентами и пересмотр этой структуры в ходе работы системы, причем этот конфликт и пересмотр отношений может осуществляться как внутри самой системы, так и требовать обращения к одному из потребителей, например, для удлинения сроков поставки за счет сокращения цены и т.д. Метод предусматривает выплату компенсаций за разрыв отношений, что означает динамического перераспределение прибыли между вновь приходящими и имеющимися заказами.

Из вышеприведенного описания ясно, что в процессе распространения "волны" переговоров и ре-переговоров (при пересмотре уже ранее заключенных договоренностей) мультиагентная система осуществляет постоянное динамическое перераспределение ресурсов между заказами и находит решения, которые:

1) постоянно учитывают возможное изменения ситуации, как-то появление новых заказов или отзыв уже имеющихся, возможный выход из строя существующих ресурсов, появление новых ресурсов на рынке и т.д.;

2) удобны для работы в условиях наращиваемых распределенных сетевых сред, в

которых участники могут взаимодействовать по принципам "каждый с каждый" и "равный с равным";

3) по возможности не ухудшают положение участников, в то же время обеспечивая улучшение ситуации (решения) компании в целом (ранее был только один заказ – теперь удалось разрешить конфликт и перераспределить ресурс и в результате прибыль по одному заказу дополняется прибылью по другому заказу). Система переходит от одного атTRACTора, состояния устойчивого равновесия, к другому (в данной ситуации новые приходящие заказы могут долго не затрагивать существовавшие договоренности, и будет расти общая прибыльность);

4) первый вариант решения находится очень быстро (в простейшем случае берется первый подходящий), поскольку не требуется глобальная оптимизация, что особенно важно с ростом числа заказов и ресурсов; если же время на принятие решений имеется – возможно дальнейшее исследование более подходящих вариантов;

5) находится некоторый временный баланс интересов и компромисс между "личными" и "общественными" интересами каждого участника, заменяющий, как правило, недостижимый глобальный оптимум.

Таким образом, в рамках предлагаемого метода проблема неопределенности и динамического окружения решается за счет введения механизмов самоорганизации, связанной с обеспечением полной децентрализованности и распределённости в процессах принятия решений путем введения автономных программных агентов и поддержкой двусторонней активности агентов, реализуемой как со стороны агентов заказов, так и агентов ресурсов, возможностью динамического установления связей и их пересмотра с согласия всех заинтересованных сторон.

При этом достоинства системы связаны не только с возможностью работы в условиях динамического окружения, но и высокой скоростью работы системы при неполной загрузке системы. Скорость системы замедляется лишь при полном пересмотре и переорганизации всех связей заказов и ресурсов, т.е. при переходе от одного атTRACTора (со-

стояния устойчивого равновесия) к другому. Такие процессы связаны с лавинообразно развивающимися процессами по типу катастроф различного рода, с той лишь разницей, что система в результате восстанавливается и переходит в новое более устойчивое равновесие. Другим потенциальным недостатком системы является квазиоптимальность принимаемых решений, что во многих случаях является приемлемой платой за возможность работы в высоко неопределенных и динамических условиях, характерных для современной глобальной экономики.

Данные модели, методы и средства взаимодействия агентов реализованы в инструментальных программных средствах [5] и успешно применяются в целом ряде различных приложений [9].

СПИСОК ЛИТЕРАТУРЫ

1. Джонсон Р., Каст Ф., Розенцвейг Д. Системы и руководство: Теория систем и руководство системами. М.: Советское радио, 1971.
2. Емельянов С.В., Наппельбаум Э.Л. Методы управления сложными системами. Принципы рациональности коллективного выбора // Техническая кибернетика. М.: ВИНИТИ, 1978. Т.10.
3. Саати Т., Кернс К. Аналитическое планирование. Организация систем. М.: Радио и связь, 1991.
4. Vittikh V.A., Skobelev P.O. Multi-agent systems for modelling of self - organization and cooperation processes // Proceedings of the XIII International Conference on the Application of Artificial Intelligence in Engineering, Ireland, Calway, 1998.
5. Batishev S.V., Ivkushkin C.V., Minakov I.A., Rzevski G.A., Skobelev P.O. MagentA Multi-Agent Systems: Engines, Ontologies and Applications // Proc. of the 3rd Intern. Workshop on Computer Science and Information Technologies CSIT'2001, Ufa, Russia, 21-26 September, 2001. Ufa State Aviation Technical University – Institute JurInfoR-MSU. Regular Papers. 2001. Vol.1.
6. Виттих В.А., Ржевский Г.А., Скобелев П.О. Мультиагентные модели взаимодействия в процессах принятия решений // Труды 4-й Международной конференции по проблемам управления и моделирования сложных систем, Самара, 17-24 июня 2002. Самара: СНЦ РАН, 2002.
7. Андреев В., Батищев С., Ивкушкин К., Минаков И., Ржевский Г., Сафонов А., Скобелев П., Шамашов М. Принципы построения открытых мультиагентных систем для поддержки процессов принятия решений // Труды 4-й Международной конференции по проблемам управления и моделирования сложных систем, Самара, 17-24 июня 2002. Самара: СНЦ РАН, 2002.
8. Simon H. The New Science of Management Decision. Harper and Row Publishers Inc., New York, 1960.
9. Скобелев П.О. Открытые мультиагентные системы для поддержки процессов принятия решений при управлении предприятиями // Известия Самарского научного центра РАН. Т3. 2001. №1.

METHOD OF COMPENSATIONS FOR DECISION MAKING SUPPORT IN DYNAMIC RESOURCE ALLOCATION PROBLEMS

© 2002 P.O. Skobelev

Institute for the Control of Complex Systems of Russian Academy of Sciences, Samara

Method of compensations for dynamic resource allocation in real time systems is described. The development of multiagent system with the world of demand and supply agents (which are able to percept current situation, make decisions, communicate and establish the relations between each other) is proposed as a solution. To negotiate and re-negotiate decisions the agents are able to propose compensations to each other, which establish the balance of interests as a result of income redistribution between existing and newly generated agents. Advantages and limitations of method are analyzed. The prospective applications for different problem domains are proposed.

УДК 336.01

ФРАКТАЛЬНО-КЛАСТЕРНАЯ ТЕОРИЯ И ТЕРМОДИНАМИЧЕСКИЕ ПРИНЦИПЫ УПРАВЛЕНИЯ СЛОЖНЫМИ СИСТЕМАМИ

© 2002 В.Т. Волов

Самарский научный центр РАН

В статье изложены основы фрактально-кластерной теории, включающей фрактально-кластерные соотношения, динамические уравнения эволюции фрактально-кластерной системы и критерии управления. Показано, что критерий эффективности фрактально-кластерной системы является более чувствительным индикатором качества структурного управления сложной самоорганизующейся системой, чем полная эффективность и информационная энтропия. Проведен анализ устойчивости состояния и трансформирования фрактально-кластерных соотношений.

Основой метода управления, анализа эффективности управления и функционирования самоорганизующейся системы в исследовании приняты термодинамический метод и фрактально-кластерные соотношения (ФКС) [1].

В результате длительных многолетних исследований [1] удалось доказать, что в любых системах: технических, биологических системах, прошедших эволюционный путь развития, системах машина-человек всегда имеется пять основных кластеров. Это энергетический (K_e), транспортный (K_{tp}), технологический (K_t), экологический (K_e) и информационный (K_i) кластеры, имеющие определенные (идеальные) значения, выраженные в процентах илиолях целого для экстенсивного параметра системы (время, деньги, масса и т.п.). Для энергетического кластера это значение составляет 38%, транспортного – 27%, экологического – 16%, технологического – 13%, информационного – 6%.

Каждый из пяти кластеров имеет пять подкластеров, например: в энергетическом – энергетическая поддержка самой энергетической системы, энергетическая поддержка транспорта, экологии, технологии, информатики и т.д. И остальные подкластеры соответствующим образом подразделяются на пять подкластеров следующего уровня. Для

анализа функционирования биологических, технических и антропогенных систем, как правило, достаточно второго или третьего уровня ФКС.

Такая кластеризация [1] позволила производить оценку функционирования сложной системы. Однако, теории на основе ФКС создано не было. Поэтому целью данного исследования являлась разработка теории структурного управления сложными системами на основе синтеза экономики, ФКС и неравновесной термодинамики.

Представленная фрактально-кластерная теория включает в себя:

- 1) фрактально-кластерные соотношения (ФКС) В.П. Бурдакова [1];
- 2) динамические уравнения эволюции фрактально-кластерной системы [2];
- 3) фрактально-кластерные критерии эффективности управления системой;
- 4) анализ устойчивости сложных самоорганизующихся систем.

Динамические уравнения ФКС

Эволюция любой экоматермической системы n -го уровня из неидеального кластерно-фрактального состояния в идеальное может быть записана в виде следующей системы уравнений [2, 3]:

$$\begin{cases} \bar{K}_i(\bar{t}) = \bar{K}_i^0 + \sum_{j=1}^5 U_{ij}(e_{ij}, \bar{t}) \cdot \bar{K}_{ij}^0; \\ \bar{K}_{ij}(\bar{t}) = \bar{K}_{ij}^0 + U_{ij}(e_{ij}, \bar{t}) \cdot \bar{K}_{ij}^0; \end{cases} \quad (1)$$

$$\begin{cases} \bar{K}_{ijm}(\bar{t}) = \bar{K}_{ijm}^0 + U_{ijm}(e_{ijm}, \bar{t}) \cdot \bar{K}_{ijm}^0; \\ \bar{K}_{ijm...n}(\bar{t}) = \bar{K}_{ijm...n}^0 + U_{ijm...n}(e_{ijm...n}, \bar{t}) \cdot \bar{K}_{ijm...n}^0; \end{cases} \quad (2)$$

$$\begin{cases} \bar{K}_{ijm...n}(\bar{t}) = \bar{K}_{ijm...n}^0 + U_{ijm...n}(e_{ijm...n}, \bar{t}) \cdot \bar{K}_{ijm...n}^0; \\ \sum_{i=1}^5 \sum_{j=1}^5 \sum_{m=1}^5 \dots \sum_{n=1}^5 \bar{K}_{ijm...n} = 1, \end{cases} \quad (3)$$

$$\begin{cases} \bar{K}_{ijm...n}(\bar{t}) = \bar{K}_{ijm...n}^0 + U_{ijm...n}(e_{ijm...n}, \bar{t}) \cdot \bar{K}_{ijm...n}^0; \\ \sum_{i=1}^5 \sum_{j=1}^5 \sum_{m=1}^5 \dots \sum_{n=1}^5 \bar{K}_{ijm...n} = 1, \end{cases} \quad (4)$$

$$\begin{cases} \bar{K}_{ijm...n}(\bar{t}) = \bar{K}_{ijm...n}^0 + U_{ijm...n}(e_{ijm...n}, \bar{t}) \cdot \bar{K}_{ijm...n}^0; \\ \sum_{i=1}^5 \sum_{j=1}^5 \sum_{m=1}^5 \dots \sum_{n=1}^5 \bar{K}_{ijm...n} = 1, \end{cases} \quad (5)$$

$$\text{где } U_{ij} = \left(\frac{\bar{K}_{ij}^{\text{ideal}}}{\bar{K}_{ij}^0} - 1 \right) f_{ij}(\bar{t}) = e_{ij} f_{ij}(\bar{t});$$

$$U_{ijm...n} = \left(\frac{\bar{K}_{ijm...n}^{\text{ideal}}}{\bar{K}_{ijm...n}^0} - 1 \right) f_{ijm...n}(\bar{t}) = e_{ijm...n} f_{ijm...n}(\bar{t}).$$

Кластеры и подкластеры любого уровня есть положительно определенные величины $\bar{K}_i > 0$; $\bar{K}_{ij} > 0$; $\bar{K}_{ijm...n} > 0$.

На функции $f(\bar{t})$ наложены следующие краевые условия:

$$\begin{cases} f_{ij}(\bar{t}) \\ f_{ijm}(\bar{t}) \\ f_{ijm...n}(\bar{t}) \end{cases} \rightarrow 0, \text{ при } \bar{t} \rightarrow 0;$$

$$\begin{cases} f_{ij}(\bar{t}) \\ f_{ijm}(\bar{t}) \\ f_{ijm...n}(\bar{t}) \end{cases} \rightarrow 1, \text{ при } \bar{t} \rightarrow \bar{t}_{\text{fin}}. \quad (6)$$

Безразмерное время эволюции подкластеров определяется:

$$\begin{aligned} \bar{t} &= \frac{t - t_{ij}^0}{t_{ij}^{\text{fin}} - t_{ij}^0}; \\ \bar{t} &= \frac{t - t_{ijm}^0}{t_{ijm}^{\text{fin}} - t_{ijm}^0}; \\ \bar{t} &= \frac{t - t_{ijm...n}^0}{t_{ijm...n}^{\text{fin}} - t_{ijm...n}^0}. \end{aligned} \quad (7)$$

где $t_{ij}^0, t_{ijm}^0, t_{ijm...n}^0, t_{ij}^{\text{fin}}, t_{ijm}^{\text{fin}}, t_{ijm...n}^{\text{fin}}$ - начальные и конечные значения первого, второго и (n-1) уровня.

В уравнениях (1)-(5) безразмерные величины определяются следующим образом:

$$\bar{K}_i = K_i / K_y; \quad \bar{K}_{ij} = K_{ij} / K_y;$$

$$\bar{K}_{ijm} = K_{ijm} / K_y; \quad \bar{K}_{ijm...n} = K_{ijm...n} / K_y.$$

Индексы i, j, m... n изменяются в пределах от 1 до 5 и соответствуют 1 – символу "э" – энергетика, 2 – "тр" - транспорт, 3 – "эк" – экология, 4 – "т" – технология, 5 – "и" – информатика.

$U_{ij}, U_{ijm}, U_{ijm...n}$ – управляющие функции для подкластеров первого, второго и (n-1) уровня, $\bar{K}_{ij}^{\text{ideal}}, \bar{K}_{ijm}^{\text{ideal}}, \dots, \bar{K}_{ijm...n}^{\text{ideal}}$ – идеальные относительные значения подкластеров первого, второго и (n-1) уровня, а $\bar{K}_{ij}^0, \bar{K}_{ijm}^0, \dots, \bar{K}_{ijm...n}^0$ – начальные относительные значения соответствующих подкластеров.

Уравнение (5) представляет собой аналог закона сохранения для фрактальной системы.

Энтропийно-кластерный метод управления структурой сложной системы

Предложенные в [2, 3] методы оптимизации управления ФКС опираются на интуитивные или жестко формализованные понятия и аналогии. В связи с выше сказанным логично сформулировать критерий эффективности управления матрицы ФКМ на основе фундаментальных принципов термодинамики устойчивых состояний.

Рассмотрим матрицу идеальных состояний ФКМ (табл.1).

Первая строка и первый столбец идеальной матрицы дают количественную информацию об общей доле энергетических ресурсов системы, которая составляет $\sim 61,5\%$, то есть является основным определяющим фактором эффективности функционирования системы:

Таблица 1. Таблица идеальных значений

\bar{K}_i		\bar{K}_{ij}				
K_1	0,38	0,1440	0,1026	0,0608	0,0494	0,0228
K_2	0,27	0,1026	0,0729	0,0432	0,0351	0,0162
K_3	0,16	0,0608	0,0432	0,0256	0,0208	0,0960
K_4	0,13	0,0494	0,0351	0,2080	0,0169	0,0780
K_5	0,06	0,0228	0,0169	0,0960	0,0780	0,0036

$$\bar{K}_y^3 = \sum_{j=1}^5 \bar{K}_{lj} + \sum_{i=2}^5 \bar{K}_{il} \approx 61,5\%. \quad (8)$$

Это число очень близко к так называемому "золотому" сечению $H_0 = 0,618$ известному из многочисленных публикаций как основа красоты и гармонии как в природных, так и в антропогенных явлениях.

Фрактально-кластерная матрица (ФКМ) $\{\bar{K}_{ij}\}$

несет информацию об энергетическом состоянии исследуемой системы. В связи с выше сказанным представляется целесообразным энтропийный подход к анализу управления фрактально-кластерной системой.

Связь между элементами ФКМ и информационной энтропией H позволяет найти критерий управления ФКМ с целью оптимальной эволюции из неидеального состояния системы (ФКМ неидеальные) в идеальное состояние (ФКМ – идеальные), при этом сумма элементов ФКМ первого столбца и первой строки (8) переходит в их идеальное значение $= 0,618$ – золотое сечение энтропии H^0 .

$$H = \sum_{j=1}^5 \bar{K}_{lj} + \sum_{i=2}^5 \bar{K}_{il} \Rightarrow U_{ij} \Rightarrow \\ \Rightarrow \sum_{j=1}^5 \bar{K}_{lj}^{ideal} + \sum_{i=2}^5 \bar{K}_{il}^{ideal} = H_0 = 0,618. \quad (9)$$

Таким образом, вводится гипотеза об определении условной энтропии (или квазиэнтропии) ФКС (9), опирающаяся на обобщение опытных данных по эволюционирующим системам [1] и структуре ФКМ.

Структура ФКМ сложной системы фрактальна: это цепочки повторяющихся подкластеров, самоподобных по своей структуре.

Как известно, изображение фракталов получается за счет итерационных процессов. Простейшим итерационным процессом является ряд Фибоначчи.

Оказалось, что ключом к управлению фрактально-кластерной матрицей (ФКМ) является знаменитый ряд Фибоначчи ($0, 1, 1, 2, 3, 5, 8, 13, U_n \dots U_{n+1}$), в котором каждая последующая цифра равна сумме двух предыдущих. Замечательным свойством ряда Фибоначчи является то, что по мере увеличения чисел ряда отношение двух соседних чисел ряда асимметрически приближается к точной пропорции "золотого" сечения – основе красоты и гармонии как в природных, так и в многочисленных антропогенных проявлениях:

$$\lim_{n \rightarrow \infty} \frac{U_n}{U_{n+1}} = H_0 = 0,618. \quad (10)$$

Правило "золотого" сечения было известно еще в Вавилонии и Древнем Египте. Пропорции пирамиды Хеопса, других произведений древнего искусства красноречиво свидетельствуют об этом.

Ряду Фибоначчи соответствуют многообразные явления структурирования в природе и обществе: размещение семян в цветах, сердечная деятельность, структура моллюска Наutilus и т.д.

В связи с выше сказанным возникла гипотеза об оптимальном управлении ФКМ при помощи ряда Фибоначчи. Но ряд Фибоначчи используется не тривиально сам по себе. Для управления ФКМ используется аппроксимация итераций ряда Фибоначчи. При

этом итерации соответствуют отрезкам времени, кратным периоду затухания колебания, то есть аппроксимация итераций ряда Фибоначчи является шаблоном для управления матрицы $\{U_{ij}\}$ (рис.1).

Управляющая матрица $\{U_{ij}\}$ при одинаковых временах начала и конца эволюции $\bar{t}_{ij}^0 = \bar{t}_{ji}^0 = \text{const}$ и $\bar{t}_{ij}^{\text{fin}} = \bar{t}_{ji}^{\text{fin}} = \text{const}_2$ принимает вид:

$$U_{ij} = \left(\bar{K}_{ij}^{\text{ideal}} / \bar{K}_{ij}^0 - 1 \right) \cdot f(\bar{t} - \bar{t}_0). \quad (11)$$

Функция $f(\bar{t} - \bar{t}_0)$ удовлетворяет следующим условиям:

$$f(0) = 0; f(\bar{t}_{\text{fin}} - \bar{t}_0) = 1, \quad (12)$$

где функция $f(\bar{t} - \bar{t}_0)$ кусочно-дифференцируема. Аппроксимация итераций ряда Фибоначчи (рис.1) дает следующее выражение:

$$\begin{aligned} f(\bar{t} - \bar{t}_0) &= \frac{H}{H_0} = 1 + \\ &+ H_0 \cdot \exp(-\alpha(\bar{t} - \bar{t}_0)) \times \cos(p(\bar{t} - \bar{t}_0) + \varphi_0), \end{aligned} \quad (13)$$

при начальной фазе

$$\varphi_0 = 0, H_0 = 0,618, \alpha = 1,05, \bar{t}_0 = 1.$$

Выражение (13) не удовлетворяет начальным условиям при $\bar{t} = \bar{t}_0$. Для удовлетворения второго граничного условия введем

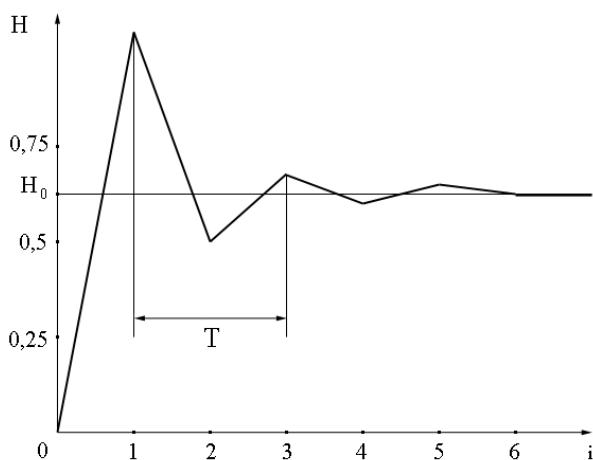


Рис.1. Итерация членов ряда Фибоначчи
i – номер итерации, T = 2 – период,
 $H_0 = 0,618$ – "золотое" сечение

новую управляющую функцию U^* на отрезке времени от нуля до некоторого t :

$$U^*(\bar{t} - \bar{t}_0) = 1 - \exp(-\beta(\bar{t} - \bar{t}_0)), \quad (14)$$

и осуществим сшивку решений для $U(\bar{t} - \bar{t}_0)$:

$$\begin{cases} U_1^*(\bar{t} - \bar{t}_0) = f(\bar{t} - \bar{t}_0) \\ (U^*(\bar{t} - \bar{t}_0))' = f'(\bar{t} - \bar{t}_0) \text{ при } \bar{t} = \bar{t}_{\text{сшивки}} \end{cases}. \quad (15)$$

Видно, что управление $U^*(\bar{t} - \bar{t}_0)$ удовлетворяет условию (12) при $\bar{t} = \bar{t}_0$. После несложных преобразований получаем систему трансцендентных уравнений:

$$\begin{cases} b = \beta - \frac{\ln(-\cos[\pi(\bar{t}_{\text{сшивки}} - \bar{t}_0)] \cdot H_0)}{\bar{t}_{\text{сшивки}} - \bar{t}_0} \\ \ln\left[-\frac{1}{\cos\pi(\bar{t}_{\text{сшивки}} - \bar{t}_0) \cdot H_0}\right] = \\ = \tan(\pi(\bar{t}_{\text{сшивки}} - \bar{t}_0)) \end{cases}, \quad (16)$$

где $\cos(\pi(\bar{t}_{\text{сшивки}} - \bar{t}_0))_0 < 0$.

Численно из уравнений (16) определяется значение $\bar{t}_{\text{сшивки}}$ и показатель β . Оказалось, что $\bar{t} - \bar{t}_0 \approx 1,19$; $\beta \approx 1,53$.

На рис.2 показана эволюция кластеров с управлением по (13, 14).

В связи с данным фактом представляет интерес сравнение решения для управляю-

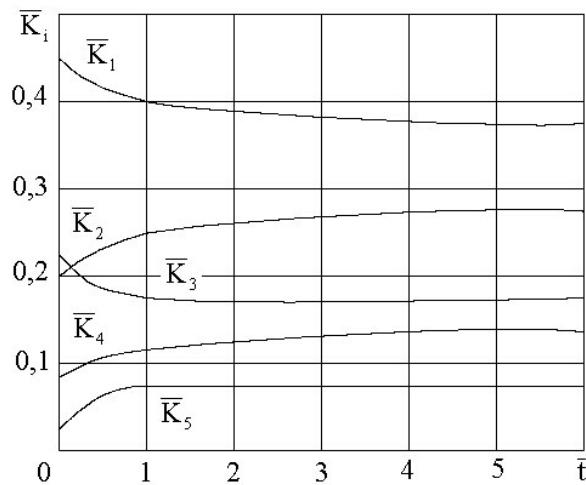


Рис.2. Эволюция кластеров по ряду Фибоначчи
для $U^* \neq 1 - \exp(-b \cdot (\bar{t} - \bar{t}_0))$ и матрица ФКС

щей функции U^* по формуле (17) с аппроксимацией итераций ряда Фибоначчи (13):

$$U(t) = \begin{cases} \bar{t}^2 \cdot \exp\left(-\frac{\beta(\bar{t}-1)}{2\sigma^2}\right) & \\ 0 \leq t < t_{\text{шивки}}, & (17) \end{cases}$$

$$\begin{cases} 1 + H_0 \cdot \exp(-\beta(\bar{t}-1)) \cdot \cos(\pi(\bar{t}-1)) & \\ t_{\text{шивки}} < \bar{t} \leq 1. & (18) \end{cases}$$

Решение системы трансцендентных уравнений определяет $t_{\text{шивки}}$ и в при фиксированном значении $\bar{t} = 1$:

$$\begin{cases} \bar{t}^2 \cdot \exp\left(-\frac{\beta}{2\sigma^2}(\bar{t}_{\text{шивки}} - 1)\right) = \\ = 1 + H_0 \cdot \exp(-\alpha(\bar{t}_{\text{шивки}} - 1)) \cos(\pi(\bar{t}_{\text{шивки}} - 1)), \\ 2\bar{t}_{\text{шивки}} \cdot \exp\left(-\frac{\beta}{2\sigma^2}(\bar{t}_{\text{шивки}} - 1)\right) - \\ - \frac{\bar{t}^2 \beta}{2\sigma^2} \cdot \exp\left(-\frac{\beta}{2\sigma^2}(\bar{t}_{\text{шивки}} - 1)\right) = \\ = H_0 \alpha \cdot \exp(-\alpha(\bar{t}_{\text{шивки}} - 1)) \cdot \cos(\pi(\bar{t}_{\text{шивки}} - 1)) - \\ - H_0 \cdot \sin(\pi(\bar{t}_{\text{шивки}} - 1)) \cdot \exp(-\alpha(\bar{t}_{\text{шивки}} - 1)). & (19) \end{cases}$$

При $\bar{t}_{\text{шивки}} = 1$ первое уравнение (19) обращается в тождество при любых β и σ .

Таким образом, получено энтропийно-кластерное решение по структурному управлению сложной системой. При этом критерием оптимизации управления сложной системы не являются общепринятые критерии – максимизация или минимизация издержек функционирования системы. Критерием оптимизации управления в предлагаемой теории является минимизация отклонения управления системой от решения (17, 18), что соответствует бескризисному развитию сложной системы, основанному на фундаментальных принципах термодинамики и обобщении законов структурной эволюции широкого класса эволюционирующих систем.

Фрактально-кластерные критерии эффективности управления сложной системой

Для определения критерия эффективно-

сти управления $D_{\text{эфф}}$ фрактально-кластерной матрицы использовался подход Ф. Хаусдорфа. В отличие от чисто фрактальных структур фрактально-кластерные n -мерные матрицы ФКМ⁽ⁿ⁾ имеют существенное отличие от геометрических фрактальных структур, так как количественные распределения по подкластерам любого уровня могут отличаться от идеального распределения и тем самым изменяется качество системы. Однако все перераспределения в кластерах и подкластерах любого уровня подчиняются законам сохранения (5).

Поэтому предлагается следующий алгоритм определения критерия эффективности фрактально-кластерной n -мерной матрицы ФКМ⁽ⁿ⁾.

Фрактально-кластерные критерии эффективности управления определяются по формуле:

$$D_{\text{эфф}} = \frac{\log \sum_{i=1}^5 \sum_{j=1}^5 \sum_{K=1}^5 \dots \sum_{m=1}^5 D_{ijk\dots m}}{\log N}, \quad (20)$$

где ФКМ является m -мерная матрица.

В формуле (20) величины $D_i^*, D_{ij}^*, D_{ijk\dots m}^*$ подсчитываются по соотношениям:

$$D_{ijk\dots m}^* = 1 - \sqrt{\left(\frac{\bar{K}_{ijk\dots m}^{\text{ideal}}}{\bar{K}_{ijk\dots m}} - 1\right)^2}. \quad (21)$$

Полная эффективность функционирования фрактально-кластерной системы определяется согласно [1] следующим образом:

$$z^y = 1 - \sum_{i=1}^5 \bar{K}_i^{\text{ideal}} z_i^{n-1} \sqrt{\left(\bar{K}_i^{\text{ideal}} - \bar{K}_i\right)^2}, \quad (22)$$

где η_i – эффективность i -го кластера определяется по формуле:

$$z_i = 1 - \sum_{j=1}^5 \bar{K}_{ij}^{\text{ideal}} \cdot z_{ij}^{n-1} \sqrt{\left(\bar{K}_{ij}^{\text{ideal}} - \bar{K}_{ij}\right)^2}, \quad (23)$$

Расчет подкластерных эффективностей $\eta_{ij\dots m}$ ($m-1$)-уровня начинается с последнего ($m-1$)-уровня. Номер уровня меняется следующим образом:

Таблица 2. Сравнительный анализ управления муниципальными структурами

Наименование структуры	Энтропия	Критерий эффективности $D_{\text{эфф}}$	Полная эффективность z^y	Относительное отклонение от идеального		
				e_H	e_D	e_z
Муниципальные структуры Московской области	1990 г.	0,360	0,1132	0,8300	41,7%	88,6%
	1993 г.	0,564	0,8755	0,9690	8,7%	12,5%
	1996 г.	0,407	0,7227	0,9257	3,4%	27,8%
Municipal Department USA, t. Nessua	1993 г.	0,603	0,9700	0,9900	2,42%	3%
	1994 г.	0,6156	0,9875	0,9957	0,4%	1,2%
						0,4%

$$m - 1 \leq n \leq 1. \quad (24)$$

Для определения эффективности управления предложен обобщенный критерий функционирования самоорганизационной системы:

$$\chi = \frac{H \cdot D_{\text{эфф}} \cdot z^y}{H_0 \cdot D_{\text{эфф}}^{\max}}, \quad (25)$$

где H – информационная энтропия, вычисляемая по (9), $H_0 = 0,618$ – "золотое сечение".

Формула (25) представляет собой кластерно-энтропийный критерий управления фрактально-кластерной системой.

В качестве примера в таблице № 2 приведен фрактально-кластерный анализ управления муниципальными структурами для Московской области и муниципального департамента г. Нешуа США. Из данной таблицы видно, что для американского муниципального департамента ФКС практически

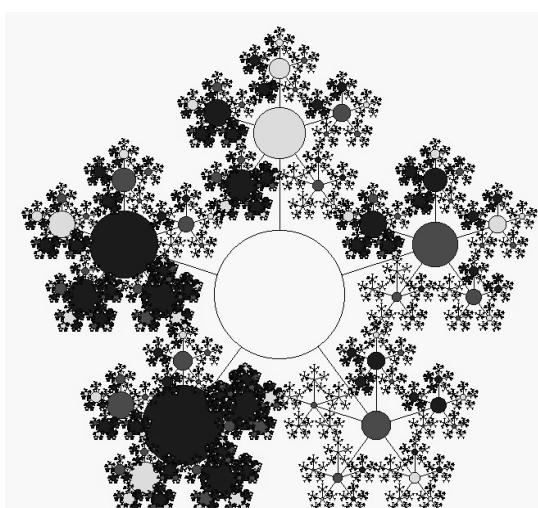
идеальны, критерий эффективности управления $D_{\text{эфф}}$ и полная эффективность системы близки к 100%. Для муниципальных структур Московской области наиболее удачным с точки зрения управления является 1993 г.

Обобщенный критерий χ для г. Нешуа США максимальен, что свидетельствует об оптимальности управления.

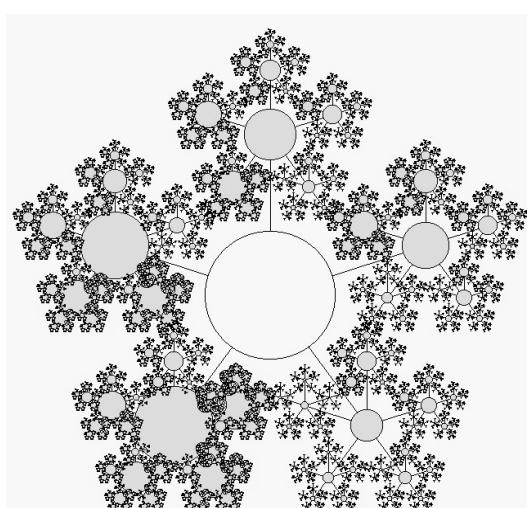
На рис. 3 показаны различные сценарии топологических структур ФКМ для различных временных этапов эволюции ФКМ к ее идеальному значению.

Анализ устойчивости состояния и трансформирования сложной системы

Используя результаты и основные положения термодинамики структуры [4] и неравновесной нелинейной флуктуационно-диссипативной термодинамики [5], проведен анализ устойчивости сложных саморазвивающихся систем.



$t = 0,5; \phi = 72^\circ$



$t = 1; \phi = 72^\circ$

Рис.3. Эволюция ФКС шестого уровня $\phi = 72^\circ$

Рассмотрим кластеры $\{\bar{K}_i\}$ и подкластеры $\{\bar{K}_{ij}\}$, составляющие ФКС, как случайные внутренние параметры $K_i(t)$, $K_{ij}(t)$, то есть изменяющиеся флюктуационным образом. Если система изолирована, то информационная энтропия $H(\{\bar{K}_{ij}\})$ не убывает со временем. Однако флюктуационная составляющая информационной (условной) энтропии $H(\bar{K}_{ij}(t))$ может убывать на величину, не превышающую k ($k = 1,38 \cdot 10^{-23} \frac{\text{Дж}}{\text{КГ}}$) (k - постоянная Больцмана).

$$\sqrt{dH(\bar{K}_{ij}(t))^2} < k. \quad (26)$$

Выражение (26) представляет собой микронарушение II начала термодинамики для ФКС.

Усредненные значения подкластеров по некоторому интервалу времени ϕ равны:

$$\sum_{j=1}^5 \langle \bar{K}_{ij} \rangle = \left\langle \sum_{j=1}^5 \bar{K}_{ij} \right\rangle = \int_0^\phi \bar{K}_{ij}(t) dt / \phi. \quad (27)$$

Условная энтропия $H(A(t))$ ФКС в случае несимметричной ФКМ имеет следующий вид:

$$H(A(t)) = \left\langle \bar{K}_1 + \sum_{j=2}^5 \bar{K}_{ij} \right\rangle = \langle \bar{K}_1 \rangle + \sum_{j=2}^5 \langle \bar{K}_{ij} \rangle, \quad (28)$$

где $A(t)$ – усредненные внутренние параметры, $\langle \rangle$ – знак усреднения по некоторому промежутку времени ϕ существенно меньшему времени эволюции T из начального состояния $\{\bar{K}_{ij}^0\}$ в конечное (идеальное) состояние

$$\text{системы } \{\bar{K}_{ij}^0\} \text{ в конечное (идеальное) состояние} \\ \phi \ll T \quad \langle \bar{K}_{ij}^0 \rangle \Rightarrow \langle \bar{K}_{ij} \rangle^{\text{fin(ideal)}}. \quad (29)$$

В симметричном случае подкластеры $\langle \bar{K}_{ij} \rangle$ определяются по соотношениям: $\langle \bar{K}_{ij} \rangle = \langle \bar{K}_{ji} \rangle$ и $\langle \bar{K}_{ij} \rangle = \langle \bar{K}_i \rangle \cdot \langle \bar{K}_j \rangle$, т.е. $\langle \bar{K}_{ij} \rangle = \langle K_i \rangle^2$.

Условная энтропия в этом случае равна:

$$H = 2\langle \bar{K}_1 \rangle - \langle \bar{K}_1 \rangle^2. \quad (31)$$

В соответствии с критерием термодинамической устойчивости [4] определим второй дифференциал условной энтропии H для симметричного случая:

$$d^2H = \frac{\partial^2 H}{\partial \langle \bar{K}_1 \rangle^2} (\Delta \langle \bar{K}_1 \rangle)^2 = -2(\Delta \langle \bar{K}_1 \rangle)^2 \leq 0. \quad (32)$$

Таким образом, для состояний, близких к ветви термодинамического равновесия в симметричном случае ФКМ, второй дифференциал энтропии d^2H отрицателен, то есть ФКС устойчива.

Граница потери устойчивости для симметричной ФКМ $d^2H = 0$ реализуется только при $\Delta \langle \bar{K}_1 \rangle = 0$, то есть при полном отсутствии флюктуаций энергетического кластера $\langle \bar{K}_1 \rangle$.

Во всех остальных случаях при симметричной ФКМ при состояниях, близких к ветви термодинамического равновесия, $d^2H < 0$.

Рассмотрим несимметричный случай ФКМ. Второй дифференциал условной энтропии в несимметричном случае ФКМ имеет вид:

$$\begin{aligned} d^2H(\langle \bar{K}_1 \rangle, \langle \bar{K}_{21} \rangle, \langle \bar{K}_{22} \rangle, \langle \bar{K}_{23} \rangle, \langle \bar{K}_{24} \rangle) = \\ = \frac{\partial^2 H}{\partial \langle \bar{K}_1 \rangle^2} (\Delta \langle \bar{K}_1 \rangle)^2 + \sum_{j=2}^5 \frac{\partial^2 H}{\partial K_{j1}^2} (\Delta \langle \bar{K}_{j1} \rangle)^2 + \\ + 2 \frac{\partial}{\partial \langle \bar{K}_1 \rangle} \sum_{j=2}^5 \left(\frac{\partial H}{\partial \langle \bar{K}_{j1} \rangle} \cdot \Delta \langle \bar{K}_{j1} \rangle \right) \Delta \langle \bar{K}_1 \rangle + \\ + 2 \sum_{i=2}^5 \sum_{j>i}^5 \frac{\partial^2 H}{\partial \langle \bar{K}_{ii} \rangle \partial \langle \bar{K}_{jj} \rangle} \Delta \langle \bar{K}_{ii} \rangle \Delta \langle \bar{K}_{jj} \rangle. \end{aligned} \quad (33)$$

Второй дифференциал условной энтропии d^2H определяется в случае независимости энергетического кластера $\langle \bar{K}_1 \rangle$ и энергетических подкластеров $\langle \bar{K}_{12} \rangle$, $\langle \bar{K}_{13} \rangle$, $\langle \bar{K}_{14} \rangle$ и $\langle \bar{K}_{15} \rangle$ следующим образом:

$$d^2H = 0, \quad (34)$$

то есть даже при наличии флюктуаций имеет место нейтральная устойчивость эволюции сложной системы.

В случае линейной зависимости $\langle \bar{K}_1 \rangle$ и энергетических кластеров $\langle \bar{K}_{12} \rangle, \langle \bar{K}_{13} \rangle, \langle \bar{K}_{14} \rangle$ и $\langle \bar{K}_{15} \rangle$ также имеет место нейтральная устойчивость.

В случае нелинейной зависимости подкластеров $\{K_{ij}\}$ ($i > 1$) от энергетического кластера могут возникать как устойчивые, так и неустойчивые режимы эволюции фрактально-кластерной матрицы ФКМ сложной системы, то есть:

$$d^2H \begin{cases} < 0 & \text{устойчивый режим} \\ = 0 & \text{нейтральная устойчивость} \\ > 0 & \text{неустойчивый режим} \end{cases}. \quad (35)$$

Проведенный выше анализ структурной устойчивости сложной системы на основе обобщенной термодинамики необратимых процессов И. Пригожина [4] и предлагаемой фрактально-кластерной теории относится к состояниям, близким к термодинамической ветви равновесия, то есть к линейной термодинамике необратимых процессов.

Критерием устойчивости для сложных систем, соответствующих понятию "диссиpативных структур" И. Пригожина, является квадратичная знакопеременная форма, называемая производством приращения энтропии [4]. Для устойчивых диссиpативных структур производство избыточной энтропии есть величина положительно определенная:

$$P[dH] > 0, \quad (36)$$

$$\text{где } dH = \frac{dH}{d\langle \bar{K}_1 \rangle} d\langle \bar{K}_1 \rangle + \sum_{j=2}^5 \frac{dH}{d\langle K_{1j} \rangle} d\langle K_{1j} \rangle.$$

Как отмечено в [4] однозначно знак производства избыточной энтропии в общем случае определить невозможно. Для конкрет-

ных систем необходимо использование феноменологических законов для определения знака $P[dH]$.

Рассмотрим вопрос об устойчивости траектории перевода сложной системы из произвольного в идеальное состояние в соответствии с основными положениями фрактально-кластерной теории. Очевидно, что через две точки в фазовой плоскости энтропия – время ($H-t$) могут проходить как устойчивые, так и неустойчивые траектории трансформирования системы с точки зрения фрактально-кластерной теории.

Рассмотрим фрактально-кластерную структуру сложной системы, находящейся в состоянии, близком к термодинамической ветви равновесия, то есть при анализе эволюции можно использовать линейную термодинамику неравновесных процессов.

В соответствии с данным фактом можно использовать теорему о минимуме производства энтропии [4].

Для простоты рассмотрим симметричную фрактально-кластерную матрицу (ФКМ) топологической структуры сложной системы, тогда энтропия системы определяется по (31).

Используя теорему И. Пригожина о минимуме производства энтропии [4], определим вид функции $f(\bar{t})$ из условия нейтральной устойчивости:

$$\frac{dP}{d\bar{t}} = 0, \text{ где } P = \frac{dH}{d\bar{t}}. \quad (37)$$

Выражение трансформационной функции $f(\bar{t})$, осуществляющей перевод из произвольного в идеальное состояние фрактально-кластерной структуры сложной системы, по траектории нейтральной устойчивости имеет следующий вид:

$$f(\bar{t}) = \{\exp(\beta) - 1\}^{-1} (\exp[\beta \cdot \bar{t}] - 1). \quad (38)$$

В общем случае ненулевой правой части в выражении для производства энтропии получаем следующее выражение для трансформационной функции $f(\bar{t})$:

$$f(\bar{t}, e) = -\frac{e}{\alpha} \bar{t} + \left(1 + \frac{e}{\alpha}\right) \cdot (\exp(\alpha) - 1)^{-1} (\exp(\alpha \cdot \bar{t}) - 1). \quad (39)$$

При параметре e , стремящемся к нулю, $f(\bar{t}, e)$ переходит в решение для $f(\bar{t})$ при нейтральной устойчивости:

$$\lim_{e \rightarrow 0} f(\bar{t}, e) = f(\bar{t}). \quad (40)$$

Выражение для функции $f(\bar{t}, e)$ соответствует следующим качественно различным режимам трансформирования топологической фрактально-кластерной структуры сложной системы из неидеального в идеальное состояние:

$$e \begin{cases} = 0 & \text{трансформирование сложной} \\ & \text{системы по траектории} \\ & \text{нейтральной устойчивости} \\ > 0 & \text{неустойчивая траектория} \\ & \text{трансформирования} \\ & \text{сложной системы} \\ < 0 & \text{устойчивая траектория трансфор-} \\ & \text{мирования сложной системы} \end{cases} \quad (41)$$

В качестве первого примера иллюстрации полученных результатов по анализу устойчивости процесса трансформирования сложной системы из произвольного состояния фрактально-кластерной структуры рассмотрим рис.4, где изображены три различных траектории трансформирования структуры экономических систем:

- 1) с большими градиентами (шоковые реформы) $\left(\frac{dP}{dt} < 0\right)$;
- 2) с переменными градиентами (переходные экономики) $\left(\frac{dP}{dt} > 0\right), \left(\frac{dP}{dt} < 0\right)$;
- 3) с небольшими градиентами (медленные реформы) $\left(\frac{dP}{dt} > 0\right)$.

Первый режим ($dP/dt < 0$) с точки зрения термодинамики устойчив, второй – име-

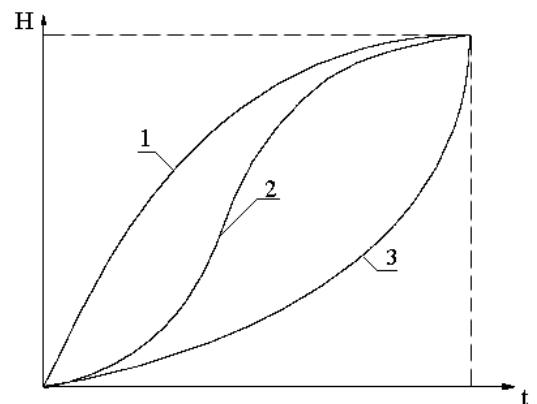


Рис.4. Управление ЭС как функция Н от t

ет устойчивый ($dP/dt < 0$) и неустойчивый ($dP/dt > 0$) характер, а третий режим трансформирования системы неустойчив.

Таким образом, медленное градуированное структурное реформирование с точки зрения фрактально-кластерной теории является неустойчивым.

Данный факт имеет подтверждение на примере рыночных преобразований стран бывших участников Варшавского пакта и стран СНГ.

Математическое моделирование (рис.5) и статистические данные [6] по ВНП, бюджетному дефициту, темпам инфляции показывают, что в странах с шоковой терапией экономики (Польша, Чехия, Словения, Эстония, Латвия) имеет место быстрое выздоровление экономики – за 4 года преобразований возникает профицит бюджета, то есть темпы инфляции снижаются до приемлемых цифр. У стран, проводящих реформы в виде неинфляционного градуализма (Венгрия, Словакия) имеет место медленный темп выздоровления экономики по сравнению со странами

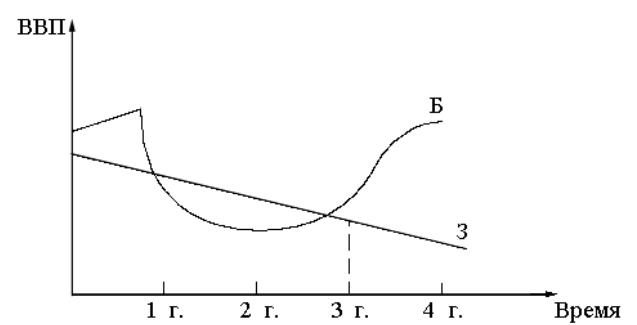


Рис.5. Зависимость ВВП от времени для быстрых (Б) и медленных (З) реформ

первой группы (Польша, Чехия, Словения, Эстония и Латвия).

У стран третьей группы (Россия, Болгария, Украина), проводящих реформирование экономики в режиме инфляционного градуализма, имеет место ухудшение экономической ситуации (падение ВНП, гиперинфляция).

СПИСОК ЛИТЕРАТУРЫ

1. *Бурдаков В.П.* Эффективность жизни. М.: Энергоиздат., 1997.
2. *Волов В.Т.* Фрактально-кластерная теория управления образовательными структурами. Казань: Казанский гос. ун-т, 2000.
3. *Волов В.Т.* Экономика, флуктуации и термодинамика. Самара: СНЦ РАН, 2001.
4. *Гленсдорф П., Пригожин И.* Термодинамическая теория структуры, устойчивости и флуктуаций. М.: Мир, 1973.
5. *Стратонович Р.Л.* Нелинейная неравновесная термодинамика. М.: Наука, 1985.
6. *Дж. Санс.* Рыночная экономика и Россия. М.: Экономика, 1994.

FRACTAL-CLUSTER'S THEORY AND THERMODYNAMIC PRINCIPLES OF THE COMPLEX SYSTEMS CONTROL

© 2002 V.T. Volov

Samara Science Centre of Russian Academy of Sciences

This theory presents the original fractal-cluster's theory for the completed selforganized systems. This theory includes the Burdakov's fractal-cluster correlations [1], the fractal-cluster dynamic equations, the fractal-cluster criterions of the selforganized systems controlling and the analysis of the ones.

УДК 621.3.078.001

РАСПОЗНАВАНИЕ ЗАКОНОМЕРНОСТЕЙ ВЗАИМОВЛИЯНИЯ МЕЖДУ ДВУМЯ МНОГОМЕРНЫМИ ПРОСТРАНСТВАМИ В СЛОЖНЫХ СИСТЕМАХ ЧЕРЕЗ ИХ ДИНАМИКУ

© 2002 Л.С. Бекасов¹, Ю.Н. Лазарев¹, А.М. Гриненко², Ю.С. Муратов¹

¹ Самарский государственный технический университет

² Самарский государственный медицинский университет

Представлены наиболее простые алгоритмы для определения характера взаимосвязи между различными координатами многомерных пространств, принадлежащих сложным системам. Эти алгоритмы позволяют обосновать критерии их классификации и прогнозировать реакции систем на внешнее возмущение информационного характера. Наряду с классическими подходами, основанными на линейной и нелинейной регрессиях, предложены оригинальные подходы с использованием понятия силы информационного возмущения, а также нового метода, базирующегося на теории множеств.

В работе представлены методы, при помощи которых в сложном объекте выявляется характер связи между отдельными подсистемами, представленными многомерными пространствами. В объекте выделяются две подсистемы А и В с пространствами Р и S соответственно. Координаты вектора состояния подсистемы А инвариантны к внешнему возмущению информационного характера, тогда как вектор состояния подсистемы В изменяет свои координаты при воздействии внешнего возмущения на сложный объект в целом.

На наш взгляд, взаимовлияние удобнее выявить через динамику, т.е. когда возмущение F информационного характера представляется через приращение параметров пространства В.

Характер связей выявляется при помощи следующих подмножеств параметров:

- $\{P_m\}_i$ - параметры, определяющие состояние подсистемы А;
 - $\{S_s\}_i$ - параметры состояния подсистемы В до возмущения;
 - $\{S^*_s\}_i$ - параметры состояния подсистемы В после внешнего возмущения F;
 - $\{D_s\}_i$ - параметры, отражающие динамику подсистемы В (разница между спокойным и возмущенным состояниями);
- где i - объем выборки (число сложных систем), m- число параметров подсистемы А, s- число параметров подсистемы В.

С целью повышения достоверности полученных результатов использовались четыре метода исследования вышеуказанных пространств, представленных множествами параметров $\{P_m\}_i$ и $\{D_s\}_i$.

Первый метод. Определялись коэффициенты корреляции между множествами $\{P_m\}_i$ и $\{D_s\}_i$ методом линейной регрессии [1]. Применительно к двум рядам наблюдений Y и X, регрессия Y по X представлена следующей зависимостью:

$$E(Y|X) = \beta_0 + \beta_1(X), \quad (1)$$

где β_0 , β_1 - коэффициенты регрессии, которые находятся по формулам:

$$\beta_0 = m_Y - \rho \frac{\sigma_Y}{\sigma_X} m_X; \quad \beta_1 = \rho \frac{\sigma_Y}{\sigma_X} \quad (2)$$

с учетом того, что ρ - коэффициент корреляции X и Y, $m_X = E(X)$, $m_Y = E(Y)$, $\sigma^2_X = D(X)$, $\sigma^2_Y = D(Y)$. За регрессию, в данном случае принята зависимость средних арифметических параметров $d_{j,i}$ от элементов множества $\{P_m\}_i$. Таким образом, если имеется параметр $p_{m,i}$ и наблюдается j_i значений $d_{i,1}, \dots, d_{i,j_i}$ случайной величины D, то зависимость средних арифметических

$$d_{j,i} = \frac{(d_{i,1} + \dots + d_{i,j_i})}{j_i} \quad (3)$$

этих значений от p_m является регрессией.

На основании соотношений (1), (2) и (3) разработана структурная схема алгоритма вычисления коэффициентов корреляции с последующей реализацией его на Delphi как в табличном, так и в графическом виде.

Второй метод. Коэффициент ранговой корреляции Спирмена [2] позволил выявить характер связей при помощи следующего подхода: реакцию подсистемы В на внешнее возмущение можно представить как некоторую аддитивную силу F_S параметров $d_{j,i}$

$$F_{S,i} = \sum_{j=1}^s d_{j,i} \quad (4)$$

где i - номер объекта, j - номер параметра пространства S .

Соотношение Спирмена применялось к двум множествам $\{F_s\}_i$ и $\{P_m\}_i$:

$$R_{m,i} = 1 - \frac{6}{n(n-1)} \sum_{i=1}^n (A_i - B_i), \quad (5)$$

где A_i - место, занимаемое ранжированными величинами $F_{s,i}$, B_i - место, занимаемое соответственно величинами $P_{m,i}$.

Первое из этих множеств было ранжировано по убыванию. Между ними находились коэффициенты $R_{m,i}$ корреляционной связи – связи m -го параметра пространства $P_{m,i}$ объекта по отношению к силе $F_{s,i}$ реакции сложной системы.

Однако формула (2) непосредственно не позволяет находить коэффициенты $R_{m,i}$, отражающие характер связности между $\{P_m\}_i$ и $\{D_s\}_j$. Что касается выявления связи в отдельно взятом сложном объекте, то значимость этой связи можно оценить, например, через отношение коэффициентов $R_{m,i}$ и $R_{m,i}^*$. Коэффициент $R_{m,i}^*$ находится следующим образом: силу F по соотношению (1) можно использовать "не полностью", исключив одну из ее составляющих $d_{j,i}$ (по j). Тогда она представится соотношением

$$F_{s,i}^* = \sum_{j=1}^{k-1} d_{j,i}, \quad (6)$$

после чего ранжирование производится над множествами $\{F_s^*\}_i$ и $\{P_m\}_i$. Коэффициенты A_i и B_i из соотношения (2) модифицируются (обозначим их соответственно $A_{i,j}^*$, $B_{i,j}^*$), а

само соотношение приобретает вид:

$$R_{m,i}^* = 1 - \frac{6}{n(n-1)} \sum_{i=1}^n (A_{i,j}^* - B_{i,j}^*). \quad (7)$$

Весомость исключенного параметра $d_{j,i}$ (по j) представится коэффициентом φ через зависимость

$$\varphi = 1 - \frac{R_{m,i}^*}{R_{m,i}}. \quad (8)$$

Применение этого понятия (силы F) предоставляет возможность оценивать вес каждой из составляющих множества $\{P_m\}$ в "раскачке" подсистемы В.

Третий метод. Более вероятно то, что в сложных системах характер взаимовлияния между $\{P_m\}_i$ и $\{D_s\}_i$ определяется нелинейной регрессией. Для подтверждения справедливости такого предположения смоделированы $\{P_m\}_i$ и $\{D_s\}_i$ в виде таблиц, которые были применены в алгоритме корреляционного отношения $\eta_{x/y}$ [3], [4], представленного формулой

$$\eta_{y/x} = \frac{\sigma_{y/x}}{\sigma_y}, \quad (9)$$

где $\sigma = \sqrt{\frac{1}{n} \sum (y_x - \bar{y})^2}$ - среднее квадратичное отклонение для ряда значений $(y_x - \bar{y})^2$;

$y_x = 1/K \sum_{i,j} (y_i n_{ij})$ – условная средняя; n_{ij} - частота повторений значения в интервалах одного признака в комбинации с определенными значениями в интервалах другого признака;

y_i - значение y , соответствующее n_{ij} (y_i - средний интервал в корреляционной таблице); i - номер строки в корреляционной таблице; j - номер столбца в корреляционной таблице; K - итоги значения частот в столбцах корреляционной таблицы;

$y = 1/n \sum_i y_i l_i$ - общая средняя; x - пределы интервалов в корреляционной таблице;

n - итог значений частот во всех столбцах (строках) корреляционной таблицы;

$\sigma = \sqrt{\frac{1}{n} \sum (y_x - \bar{y})^2}$ - среднеквадратичес-

кое отклонение для ряда у;

1 - итоги значений частот в строках корреляционной таблицы.

Подчеркнем, что таблица отражает зависимость между случайными величинами $\{P_m\}_i$ и $\{D_s\}_j$ с той особенностью, что в ней эти величины представлены интервалами по строкам и столбцам. Координаты таблицы (число строк и столбцов соответственно представлены индексами i, j , а сами переменные обозначены как x и y). Алгоритм (9) реализован программой, которая выдала таблицы со значениями корреляционных отношений применительно ко всем элементам множеств $\{P_m\}_i$ и $\{D_s\}_j$.

Четвертый метод. Авторами данной статьи разработан упрощенный метод исследования количественного влияния параметров множества А на изменение параметров множества В. Метод позволяет легко получать вероятностную характеристику степени связи параметров и определять коэффициенты взаимосвязи между множествами. Его основу составляет анализ подмножеств, получаемых путем пересечения выборок, формируемых пороговой обработкой рядов наблюдений. При этом смещение мощностей подмножеств по сравнению со статистически ожидаемым значением говорит о наличии взаимовлияний наблюдаемых величин [5].

Метод включает следующие этапы :

1. Исследуемая совокупность рядов или один ряд наблюдений i -го параметра, относящегося к множеству А, разбивается относительно медианы на два приблизительно равных по мощности подмножества F_{+i} и F_{-i} . Аналогично некоторая совокупность или один ряд j -го параметра, относящегося к множеству В, также разбивается на два подмножества P_{+j} и P_{-j} .

2. Определяются мощности пересечений подмножеств $M(F_{+i} \cap P_{+j})$ и $M(F_{-i} \cap P_{+j})$. Если взаимовлияние параметров отсутствует, то мощности пересечений приблизительно равны. В противном случае будет наблюдаться смещение мощности пересечений. Однако, сумма мощностей пересечений в любом случае равна $M(P_{+j})$. Для количественной оценки относительных смещений мощностей пересечений нами введен коэффици-

ент смещений K_{cm}

$$K_{cm}(i,j) = \frac{M(F_{+i} \cap P_{+j}) - M(F_{-i} \cap P_{+j})}{M(P_{+j})}. \quad (10)$$

Этот нормированный коэффициент может принимать значения от -1 до $+1$. При отсутствии связи между изучаемыми совокупностями рядов наблюдений его значение стремится к 0 .

3. Для оценки достоверности обнаружения взаимного влияния двух качественно различных сторон изучаемого объекта, характеризуемых двумя вариационными рядами А и В, предполагается отсутствие этого взаимного влияния, то есть выдвигается "нулевая гипотеза". Затем, в качестве реперных точек вычисляются вероятности отклонения K_{cm} от нуля, чем вводится вероятностная шкала оценки. Определяется вероятность отклонения K_{cm} .

Вероятность получения значения $M(F_{+i} \cap P_{+j}) = MM$ составляет

$$P_{MM} = \frac{C(0,5N, MM) C(0,5N, (0,5N-MM))}{C(N, 0,5N)},$$

$$\text{где } C(x,y) = \frac{x!}{y!(x-y)!},$$

N - число изученных объектов.

Практическое применение. Представленные методы применены в исследовании моделей сложных систем, в частности, человека.

Независимой подсистемой А служила психологическая структура личности, описываемая пространством Р, с координатами - ипохондрия, депрессия, истерия, психопатия, шкала мужских и женских черт характера, паранойи, психастении, шизофрении, гипомании, социальной интраверсии, представленных шкалами теста MMPI.

Подсистема В и ее пространство S - это сердечно-сосудистая система (ССС) с координатами: частота сердечных сокращений (ЧСС), артериальное давление систолическое (АДС), артериальное давление диастолическое (АДД), ударный индекс (УИ), сердечный индекс (СИ), удельное периферическое сосудистое сопротивление (УПСС), мощность сокращения левого желудочка (МСЛЖ), рас-

ход энергии на перемещение 1 л минутного объема крови (РЭ), тройное произведение (ТП). Таким образом были получены два множества вариационных рядов параметров : множество А включало 10 перечисленных психологических и множество В - 9 физиологических параметров.

Возмущающим фактором, определяющим значения элементов множества $\{Ds\}_j$ (динамику параметров ССС) служил экзаменационный стресс.

Применительно к последнему методу для каждого испытуемого определялись значения y_i :

$$y_i = \frac{2y_{i \text{ после}} - y_{i \text{ до}}}{y_{i \text{ после}} + y_{i \text{ до}}},$$

где y_i – значение i -го параметра, отражающее физиологическую динамику;

$y_{i \text{ после}}$ и $y_{i \text{ до}}$ – значения i -ых физиологических параметров после и до возмущения;

i – число пар рядов измерений.

Результаты вычисления вероятности появления определенных значений K_{cm} при $N = 85$ приведены в таблице.

При значении $K_{cm} = 0,371$ вероятность статистической независимости рядов наблюдений составляет менее 0,001. Это означает, что с высокой степенью достоверности можно утверждать наличие взаимовлияний.

Первые три метода были также апробированы на вышеобозначенном статистическом материале. Данные сведены в таблицу коэффициентов корреляции между психоло-

Таблица. Результаты вычисления вероятности

ММ	13	17	21	25	29
K_{cm}	-0,38	-0,19	0	0,19	0,38
P_{mm}	0,000	0,032	0,172	0,046	0,0006

гическими параметрами и изменениями показателей гемодинамики в условиях внешнего возмущения. Ввиду большого объема этой таблицы приводится ее фрагмент, отражающий взаимовлияние психологического фактора депрессии и частоты сердечных сокращений .

При линейной регрессии оно составило 0,571, а при ранговой корреляции 0,549.

Сопоставление результатов обработки, полученных всеми четырьмя методами дают расхождение не хуже 4-12%. Поскольку данная работа посвящена, в основном, описанию стратегий решения поставленной задачи, мы не приводим здесь полностью цифровой материал в силу его большого объема.

СПИСОК ЛИТЕРАТУРЫ

1. Крамер Г. Математические методы статистики. М.: Мир, 1975.
2. Корн Г., Корн Т. Справочник по математике. М.: Наука, 1970.
3. Гурман В.Е. Теория вероятности, математическая статистика. М.: Высшая школа, 1977.
4. Зенкин А.Н. О математических методах прогнозирования. М.: Знание, 1976.
5. Феллер В. Введение в теорию вероятностей и ее приложения. Т.1. М.: Мир, 1984.

THE RECOGNIZING OF THE MUTUAL INFULANCE REGULARITIES OF TWO MULTIDIMENSIAL SPACES IN THE COMPLICATED SYSTEM BY THEIR DYNAMICS

© 2002 L.S. Bekasov¹, Yu.N. Lazarev¹, A.M. Grinenko², Yu.S. Muratov¹

¹Samara State Technical University

²Samara State Medical University

This paper describes more simple algorithmus for the determining of correlation character between different coordinates of the multidimensional spaces in the complicated systems. These algoritms let to base the criterions of their classifications and forecasting of their responses on the outward information disturbance. This papir suggests the original approaches founded on new conception of the disturbance force and using some principles of the theory of Sets in common with the classical approaches founded on linear and nonlinar regressions.

УДК 621.375.826

КОМПЬЮТЕРНАЯ СИСТЕМА УПРАВЛЕНИЯ ТЕХНОЛОГИЧЕСКИМИ ПРОЦЕССАМИ ЛАЗЕРНОЙ И КОМБИНИРОВАННОЙ ОБРАБОТКИ МАТЕРИАЛОВ

© 2002 С.П. Мурzin

Самарский государственный аэрокосмический университет

Создана система бесконтактного контроля температурного поля в зоне высокоэнергетической обработки материалов. Операции лазерной термической, комбинированной плазменно-лазерной и лазерно-дуговой обработки автоматизированы с помощью специально разработанных аппаратно-программных средств, введенных в состав лазерного технологического комплекса ЛК-1300 "Хебр-1А". Разработана компьютерная система управления технологическими процессами.

Технологические процессы обработки материалов концентрированными энергетическими потоками только тогда раскрывают свои преимущества и становятся экономически целесообразными, когда позволяют выполнять операции с высокой производительностью, качественно и автоматизированно [1]. Выбор режимов обработки на основе данных экспериментального и теоретического моделирования должен сочетаться с текущим контролем и управлением в ходе проведения технологических операций. Разработка аппаратных и программных средств для контроля, оптимизации, прогнозирования и управления лазерными технологическими процессами обработки материалов посвящены работы [2, 3 и др.]. В них рассматриваются различные аспекты автоматизации технологических комплексов для обеспечения гарантии качества и воспроизводимости процессов газолазерной резки, лазерной сварки и маркировки изделий. Однако вопросы разработки информационно-измерительных систем диагностики и средств управления технологическими процессами лазерной термической и комбинированной обработки материалов отражены далеко не в полной мере.

Известно, что несоответствие температурно-скоростных режимов обработки тонкостенных деталей заданному изменению состояния технологических объектов приводит уже на стадии обработки к образованию различных дефектов, таких как пережоги, крупнозернистость, локальные оплавления и ко-

робление листовых заготовок из-за неравномерного тепловыделения по ширине зоны термического влияния. Для объемных деталей характерны следующие дефекты: неравномерное распределение механических свойств по ширине зоны термического влияния; неравномерная глубина обработки; локальное оплавление; повышенная хрупкость изделия вследствие перегрева центральной и недостаточная твердость в результате недогрева периферийных областей энергетического воздействия [4, 5]. Использование в составе оптических систем формирования излучения элементов компьютерной оптики (фокусаторов излучения) [6-8] создает предпосылки для устранения перечисленных негативных последствия технологических процессов. Обработка деталей полосовым тепловым источником имеет ряд технологических преимуществ и позволяет обеспечить наиболее однородное распределение механических свойств по ширине зоны термического влияния. Увеличение интенсивности излучения по краям полосы компенсирует повышенные тепловые потери периферийных участков [9].

Применение средств для бесконтактного контроля температурного поля и обнаружения дефектов обработанной поверхности деталей в составе систем компьютерного управления создает предпосылки для реализации сложных законов управления высокоэнергетическими технологическими процессами, в том числе и законов управления с

адаптацией.

Поскольку специфика взаимодействия высококонцентрированных потоков энергии с деталью делает невозможным использование контактных способов измерения температуры, то контроль температурных полей целесообразно осуществлять оптико-электронными системами анализа потоков теплового излучения. При температуре поверхности от 700 до 1500°C большая часть энергии излучения с поверхности технологических объектов приходится на спектральный диапазон $d\lambda = (1\dots6)\cdot10^{-3}$ м. В этой области ИК-спектра излучательная способность материалов в меньшей степени зависит от шероховатости их поверхности, т.к. длина волн излучения становится больше размеров самих неровностей. Использование тепловизоров (оптико-электронных сканирующих инфракрасных радиометров пассивного типа) позволяет обеспечить однозначность и пропорциональность значения сигнала отклика контролируемому параметру при достаточно высоких пространственно-энергетических и временных разрешениях. Для проведения экспериментальных исследований быстроменяющихся температурных полей в реальном масштабе времени и применения в составе систем управления технологическими процессами создана автоматизированная система бесконтактной диагностики нестационарных тепловых процессов.

Периодичность поступления видеоинформации при использовании телевизионного стандарта достаточна, поэтому элементная база из ИК-радиометра - блока оптической визуализации температурного поля в рабочей зоне тепловизора "Радуга", устройства ввода-вывода видеоизображения и персонального компьютера в состоянии обеспечить создание автоматизированных систем с использованием программных средств обработки изображений.

Визуальное наблюдение нестационарных процессов в их развитии, неразрушающая диагностика и термографические исследования поверхности объекта, нагретой до температуры $T = (300\dots1500)^{\circ}\text{C}$, проводятся при использовании ослабителя теплового

излучения. Соответствие излучательной способности поверхности исследуемого объекта значениям уровня и диапазона исследуемых температур устанавливаются в режиме калибровки сигнала. Энергия излучения преобразуется в пропорциональный ей электрический сигнал, который управляет яркостью луча, перемещающегося по экрану электроннолучевой трубы видеоконтрольного устройства. Распределение температуры по поверхности объекта фиксируется в виде изображения с различной яркостью участков или в условных цветах палитры RGB.

Обработка информации о нестационарных тепловых процессах в зоне лазерного воздействия осуществляется с использованием разработанного программного обеспечения, функционирующего в среде Windows 95/98 и имеющего многооконный интерфейс. Анализ термоизображения проводится как в автоматическом режиме работы, так и в командном интерактивном. Реализуются различные комбинации воспроизведения, записи и сравнения термограмм. Например, изображение или его часть может фиксироваться в режиме "стоп-кадр" с увеличением отдельных участков. В этом случае, на одной части экрана может быть воспроизведена эталонная термограмма, а на другой – в реальном масштабе времени визуализируется характер распределения температур объекта. Предусмотрено формирование полутоновой или цветной шкалы, а также цифровой информации о номере кадра, диапазоне и уровне исследуемых температур. На экране также отображается оптический клин с градациями яркости или шкалой цвета.

Для устранения влияния диффузно-рассеянного лазерного излучения, уровень мощности которого может значительно превышать тепловое излучение нагретой области, применяются германиевый и сапфировый фильтры, ограничивающие спектральный диапазон пропускания полосой $d\lambda = (2\dots7)\cdot10^{-3}$ м. При работе с сильнозашумленными термоизображениями модуль преобразователя кода переводится в режим цифровой фильтрации, при котором действие шумов ослабляется. Для установившегося (квазистацио-

нарного) состояния, которое наступает как предельное при длительном действии подвижного источника, разброс показаний от действия шума ослабляется с увеличением времени измерения. Программное обеспечение позволяет осуществить пространственную и временную фильтрацию помех, а также сигнализировать о выходе температуры за заданные границы в выбранном участке термоизображения.

На основе тепловизора ТВ-03 разработана система контроля пространственного распределения интенсивности лазерного излучения.

Схема регистрации нестационарных тепловых процессов в зоне лазерного воздействия и пространственного распределения интенсивности излучения представлена на рис.1. Излучение 1, отразившись от фокусирующего оптического элемента (фокусатора излучения) 2, попадает на ответвитель 3. Ответвленная часть пучка (1% от полной мощности) направляется на зеркало механиз-

мов кадровой и строчной развертки оптического блока 4 тепловизора ТВ-03. Чувствительная площадка пирометрического приемника лазерного излучения располагается в плоскости, оптически сопряженной с анализируемой. Основная часть излучения поступает на обрабатываемую деталь 5. Для повышения эффективности энерговклада используются специальные поглощающие покрытия 6. Тепловое излучение от исследуемых участков объекта отражается поворотным зеркалом 7, установленным под углом 35...40° к оптической оси падающего излучения. Германиевый и сапфировый фильтры 8 ограничивают спектральный диапазон пропускания. Распределение температурного поля регистрируется приемником излучения в оптическом блоке тепловизора "Радуга" 9. Ирисовая диафрагма позволяет изменять относительное отверстие объектива, уменьшая входной поток излучения при исследовании высокотемпературных объектов. Угловое искажение термоизображений исправляется программными средствами автоматизированной системы бесконтактной диагностики. В состав электронных блоков тепловизоров входят аналогово-цифровые преобразователи 10. Цифровые сигналы через устройства ввода/вывода видеоИзображения поступают в персональный компьютер 11.

Разработано лазерное устройство обнаружения поверхностных дефектов обработанной поверхности деталей. Методы прямой голографической интерферометрии и голографических топограмм для технической дефектоскопии поверхности деталей с нанесенными пористыми композиционными газотермическими и модифицированными покрытиями применять нецелесообразно. Сущность разработанного способа заключается в определении изменения интенсивности диффузно отраженного излучения, рассеивающееся на участках трещинообразования и отслаивания покрытия. Лазерное устройство позволяет обнаружить поверхностные повреждения, порядок величин линейных размеров которых соответствует диаметру сканируемого сфокусированного пучка полупроводникового или Не-Не-лазера. К достоин-

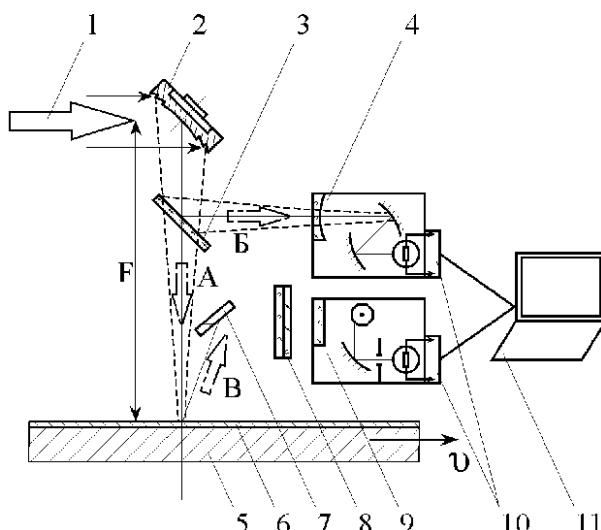


Рис.1. Схема регистрации нестационарных тепловых процессов в зоне лазерного воздействия: 1 – поток лазерного излучения; 2 – фокусатор излучения; 3 – ответвитель; 4 – оптический блок тепловизора ТВ-03; 5 – обрабатываемая деталь; 6 – поглощающее покрытие; 7 – поворотное зеркало; 8 – германиевый и сапфировый фильтры; 9 – оптический блок тепловизора "Радуга"; 10 – аналогово-цифровые преобразователи; 11 – персональный компьютер. А – лазерное излучение, падающее на обрабатываемую деталь; Б – ответвленная часть лазерного излучения; В – тепловое излучение от исследуемых участков объекта

ствам разработанного средства контроля следует отнести возможность автоматизации экспресс-диагностики, количественной оценки дефектов и построения профилограмм, а также достаточную простоту и высокую надежность устройства, состоящего из относительно недорогостоящих элементов.

Разработаны аппаратно-программные средства автоматизации операций лазерной термической, комбинированной плазменно-лазерной и лазерно-дуговой обработки, введенные в состав лазерного технологического комплекса ЛК-1300 "Хебр-1А". Создана компьютерная система управления технологическими процессами (рис.2), состоящая из следующих элементов:

1. Системы автоматизированной подготовки программ (САПР) управления исполнительными элементами технологического модуля (ТМ) лазерного комплекса (рис.3). Одними из элементов САПР являются программные средства обеспечения автоматизированного преобразования данных, подготовленных в стандартных САПР (AutoCAD, CherryCAD) и пакетах иллюстративной векторной графики (CorelDRAW, Fractal Design

Painter) в программу управления исполнительными элементами систем позиционирования, подачи вспомогательных газов, транспортировки излучения и изменения профиля его интенсивности.

2. Системы контроля пространственно-распределения интенсивности лазерного излучения, формируемого динамическими фокусаторами (элементами компьютерной оптики).

3. Системы бесконтактного контроля температурного поля в зоне высокоэнергетической обработки материалов.

4. Средства обнаружения поверхностных дефектов обработанной поверхности деталей.

5. Устройства сопряжения блока управления технологическим модулем лазерного комплекса с персональным компьютером.

6. Программного обеспечения обмена данными и трансляции управляющих команд.

Основные задачи системы компьютерного управления:

- преобразование данных, подготовленных в стандартных САПР и пакетах иллюстративной векторной графики, в программу

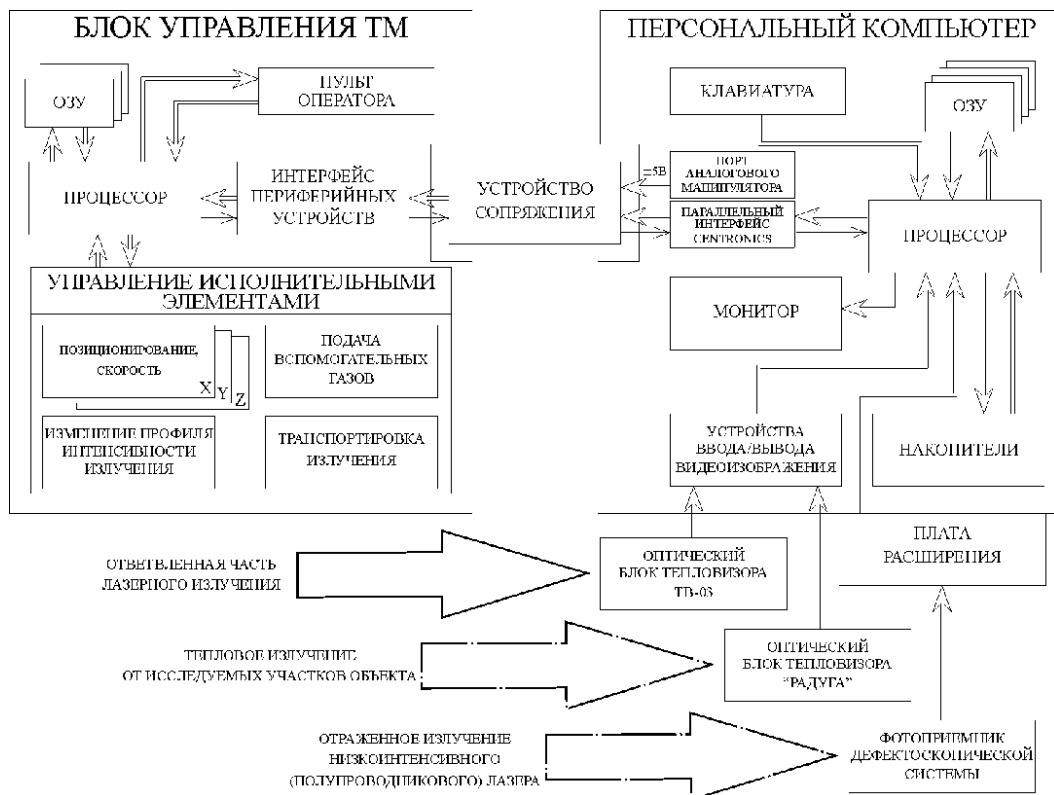


Рис. 2. Структурная схема компьютерной системы управления

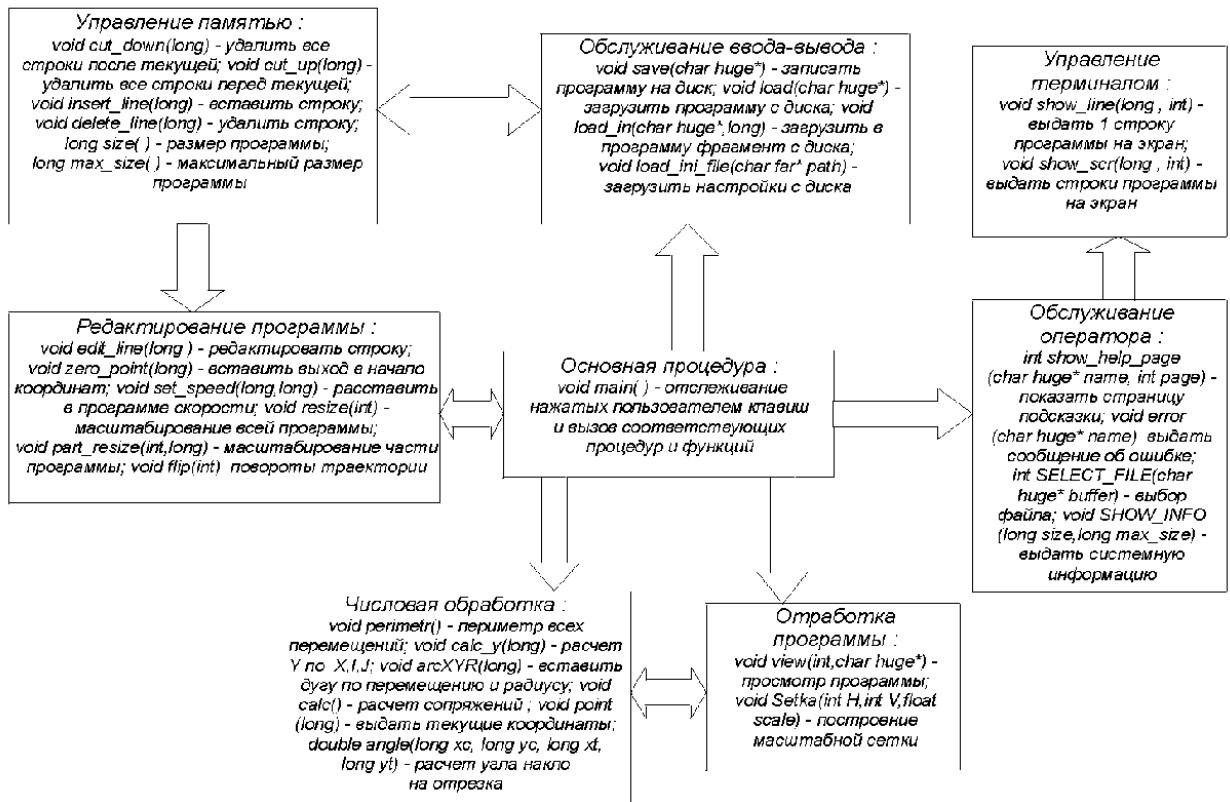


Рис. 3. Структурная схема системы автоматизированной подготовки программ (САПП) управления исполнительными элементами технологического модуля лазерного комплекса

управления исполнительными элементами систем технологического модуля лазерного комплекса: позиционирования, подачи вспомогательных газов, транспортировки излучения и изменения профиля его интенсивности;

- синтез программ управления исполнительными элементами технологического модуля лазерного комплекса и проведение сопутствующих вычислений;
- отображение получаемого контура в системе координат плоскости обработки, определение длины траектории перемещения и задание скорости позиционирования;
- редактирование программ и их хранение в виде электронных документов;
- трансляция команд управления на исполнительные элементы систем технологического модуля лазерного комплекса: позиционирования, подачи вспомогательных газов, транспортировки излучения и изменения профиля его интенсивности;
- передача программ управления в оперативное запоминающее устройство системы

управления технологическим модулем для хранения и последующего исполнения;

- контроль и изменение профиля интенсивности лазерного излучения, формируемого динамическими фокусаторами (элементами компьютерной оптики);
- бесконтактный контроль температурного поля на поверхности материалов в зоне воздействия концентрированных энергетических потоков;
- обнаружение поверхностных дефектов обработанной поверхности деталей;
- обеспечение оператора информацией о контролируемых параметрах технологического процесса;
- изменение параметров энергетического воздействия для проведения коррекции температурно-скоростных режимов обработки.

Таким образом, созданная система контроля температурного поля в зоне высокоэнергетической обработки материалов является элементом компьютерной системы управления технологическими процессами лазерной термической, комбинированной плав-

менно-лазерной и лазерно-дуговой обработки, проводимыми на лазерном технологическом комплексе ЛК-1300 "Хебр-1А".

Компьютерная система управления технологическими процессами обработки материалов концентрированными энергетическими потоками является универсальной: имеется возможность доработать находящееся в эксплуатации оборудование различных фирм-производителей. Разработанные аппаратно-программные средства автоматизации позволяют реализовать сложные законы управления высокоэнергетическими технологическими процессами при проведении операций лазерной и комбинированной обработки различных конструкционных материалов, в том числе и законы управления с адаптацией.

СПИСОК ЛИТЕРАТУРЫ

1. Абильситов Г.А., Гонтарь В.Г., Новицкий Л.А. и др. Технологические лазеры. Справочник / Под общ. ред. Г.А. Абильситова. М.: Машиностроение. 1991. Т.2.
2. Лапота В.А. Система подхода к разработке САПР и АСУ лазерных технологических комплексов // Лазерная технология и средства ее реализации. Материалы науч.-технич. конф. Санкт-Петербург: СПбГТУ, 1997.
3. Golubev V.S., Panchenko V.Ya. Applications of lasers in machine-building and metallurgy // Progress in Research and Development of High-Power Industrial CO₂ Lasers. Editors, Proceedings of SPIE. 2000. Vol.4165.
4. Лазерная и электронно-лучевая обработка материалов. Справочник / Н.Н. Рыкалин, А.А. Углов, И.В. Зуев, А.Н. Кокора. М.: Машиностроение, 1985.
5. Григорьянц А.Г. Основы лазерной обработки материалов. М.: Машиностроение, 1989.
6. Golub M.A., Sisakian I.N., Soifer V.A. Infra-red radiation focusators // Optics and lasers in engineering. 1991. Vol.15. №5.
7. Prokhorov A.M., Sisakian I.N., Soifer V.A., Shorin V.P., Mordasov V.I., Murzin S.P. et al. Device for laser treatment of an object: Pat. USA 5, 103, 073. 7.04.1992. Int. cl.⁵ B 23 K 26/00.
8. Методы компьютерной оптики / Под ред. В.А. Сойфера. М.: Физматлит, 2000.
9. Мурзин С.П. Оптимизация поля температур при лазерной обработке материалов с применением фокусаторов излучения // Компьютерная оптика, МЦНТИ, 2001. Вып.22.

THE COMPUTER SYSTEM CONTROL OF THE TECHNOLOGICALLY PROCESSES OF THE LASER AND COMBINED MATERIALS TREATMENT

© 2002 S.P. Murzin

Samara State Aerospace University

The system of temperature field noncontact monitoring in a high-energy material-processing zone is designed. The laser thermal treatment, combined plasma-laser and laser-arc processing are automated with a help special hardware/software module, included in a laser technological complex LK-1300 "Chebr-1A". The computer system control of the technologically processes is created.

УДК 681.3.016.075

ИНТЕЛЛЕКТНАЯ СИСТЕМА УПРАВЛЕНИЯ ИННОВАЦИОННЫМ ПРОЦЕССОМ ИНТЕГРИРОВАННОГО ПРОИЗВОДСТВЕННОГО КОМПЛЕКСА

© 2002 А.В. Барвинок¹, А.Н. Коптев²

¹ Ленинградский металлический завод

² Самарский государственный аэрокосмический университет

Рассматривается проблема управления инновационно-инвестиционным процессом в интегрированных производственных комплексах и на базе теоретических моделей предлагаются структуры интеллектной и экспертной системы управления инновационным процессом на предприятии.

Проведенные исследования показали значительный интерес в стране к проблеме инновационного развития и, особенно, к решению задач реализации механизма регулирования и управления этим развитием.

В работе на базе теоретических моделей предложена структура интеллектной системы управления инвестиционно-инновационным процессом в интегрированных производственных комплексах, включающая две инновационные подсистемы. Одна инновационная подсистема реализует проект предventивного обслуживания, т.е. аккумулирует исходные средства G и контролирует текущие средства M и, как следствие, стимулирует отношение "процесс-структура" путем выделения средств \tilde{M} , поступивших в распоряжение субъекта в результате осуществления предшествующих этапов целенаправленной деятельности. Кроме этого в задачи подсистемы входит развитие партнерских отношений с венчурными фондами, финансовыми структурами страны и отрасли для содействия размещения базового и венчурного капитала в подобранные экспертной комиссией отрасли проекты. Вторая инновационная подсистема реализует инновационные технологии - набор методов, средств и последовательных мероприятий, обеспечивающих реализацию нововведения.

Подобные подсистемы требуются по ряду причин. Во-первых, необходимая гибкость систем проектирования и реализации проектов в части охвата большой номенклатуры проектов и наличие распределенных

информационных источников идей требует децентрализации некоторых управляющих функций по реализующим проектно-производственным подразделениям. Во-вторых, использование различных современных CAD/CAM/CAE систем обуславливает необходимость предотвращения задержек из-за нестыковок проектной документации с возможностями производства, так как задержки в реализации инноваций связаны с эффективностью проектов.

Вышеприведенное обсуждение предполагает, что проблема управления инновационным процессом должна решаться посредством "отдельных процессов целенаправленной деятельности", в рамках децентрализованной структуры принятия решений, способной оценивать не только результаты реализации выбранных проектов, но и выходы всех промежуточных уровней этого процесса.

Этой цели можно достичь, если на каждом уровне реализации (разбиваем множество PUGUMUCUV показателей, характеризующих цели субъекта деятельности, его средства и среду целенаправленной деятельности, на уровни в соответствии с отношением "процесс-структура") будет установлена интеллектная система контроля показателей, функция которой моделируется с теоретической точки зрения как решение задачи адаптивного управления. Для процессов целенаправленной деятельности подобная модель управления реализацией инновационных проектов требует:

- оценки текущего уровня средств ($M = \{m_k\}$ показателей) для реализации проекта путем контроля $\{m_k\}$ в интервалах Δm_k с точностью δm_k в пространстве локализации R целенаправленной деятельности на этапах реализации;

- решения по стратегии обслуживания, необходимой для предотвращения случаев выхода этих показателей за установленные пределы, базируясь на динамической модели, описывающей временную эволюцию текущего расходования средств.

На практике все это означает, что желательно иметь контролирующую систему, которая включает:

- подсистему контроля бюджета инновационного проекта, предназначенную для сканирования статей расхода проекта на каждом этапе его реализации в соответствии с масштабом времени Δt_i , описываемых показателями $\{a_i\}$;

- анализатора показателей n -го уровня (подмножество $\{a_p^{(n)}\}$, находящихся друг с другом в отношениях взаимодействия, и подмножества $\{a_s^{(n)}\}$, находящихся в отношениях воздействия к $\{a_p^{(n)}\}$. При этом $\{a_s^{(n)}\}$ является структурой, определяющей проте-

кание процессов взаимосвязанного изменения $\{a_p^{(n)}\}$.

Основываясь на этих оценках параметров, процедура регулирования расхода средств должна обеспечить (рис.1):

- оценку текущего уровня расхода бюджета проекта в виде значений показателей (используя обсервер расхода);

- описание будущей эволюции расходов посредством динамической модели затрат (предиктор расхода бюджета);

- решение о стратегии обслуживания, которую необходимо выбрать для предотвращения расхода бюджета выше установленного уровня (основываясь на прогнозировании эволюции), путем решения задачи выбора обслуживания с гарантированной надежностью.

Важными для оценивания расхода бюджета являются следующие параметры: стоимость денег во времени (инфляция, процент, риск, измерение текущей и будущей стоимости одной денежной меркой), сложный процент, дисконтирование - процесс роста основной суммы вклада в проект (текущая и будущая стоимость).

Любое решение по превентивному обслуживанию должно базироваться на этих параметрах расхода, как на их текущих зна-



Рис.1. Структура интеллектной системы управления инновационным процессом

чениях, так и на предсказаниях их эволюции в будущем (динамические модели партнеров и взаимодействие между уровнями).

Определение динамического предикатора расхода бюджета проекта, т.е. динамику его расхода, можно описывать марковскими моделями при следующих допущениях.

1. Для уровней разрушения бюджета допускается существование дискретной матрицы вместе с преобразованием пространства параметров расхода к такой матрице уровней разрушения. Это означает, что на основе множества идентифицированных в данный момент параметров расхода могут быть оценены соответствующие разрушения бюджета.

2. Считается, что уровни разрушения будут упорядочены в порядке увеличения степени разрушения. Количество уровней разрушения является таким, что переходы между соседними уровнями могут быть смоделированы посредством случайных потрясений.

При этих допущениях уровни разрушения определяют N_T состояний для марковской модели следующего типа:

$$\dot{P}(t) = \Lambda P(t); \quad P(t_0) = W, \quad (1)$$

где $P(t) = [P_1(t), \dots, P_j(t), \dots]^T$ является N_L -мерным вектором, чей j -й элемент дает вероятность того, что разрушение бюджета достигает уровня j в момент времени t ; Λ - постоянной матрицей перехода; t_0 - начальным временем процедуры предсказания разрушения бюджета проекта (т.е. момент времени выделения данного бюджета); W - вектором начальных вероятностей (т.е. вектор вероятностей, ассоциируемый с оцененным начальным уровнем разрушения) и $P_{NL}(\cdot)$ является поглощающим состоянием, соответствующим отказавшему бюджету.

Идентификацию матрицы перехода можно выполнить, воспользовавшись методом максимума правдоподобия и опираясь на следующее решение марковской модели:

$$P(t) = \exp(\Lambda(t - t_0))W. \quad (2)$$

Если принять τ за фиксированную задержку между последовательными оценками уровня разрушения, то функция правдоподобия сводится к следующей форме:

$$n(\Lambda) = \sum_{i,k} M_{ik} \ell n H_{ik}(\Lambda, \tau), \quad (3)$$

где M_{ik} означает число наблюдаемых переходов с уровня k на уровень i и H_{ik} является (i, k) - м элементом матрицы $H = \exp(\Lambda\tau)$. Поскольку, по определению, H является одноступенчатой матрицей перехода дискретной марковской цепи, то

$$H_{ik}(\cdot) \geq 0; \quad \sum_i H_{ik}(\cdot) = 1. \quad (4)$$

Следовательно, функция правдоподобия может быть переписана следующим образом:

$$\Pi(\Lambda) = \sum_k \Pi_k(\cdot); \quad n_k(\cdot) = \sum M_{ik} \ell n H_{ik}(\cdot) \quad (5)$$

и каждая подфункция $\Pi_k(\cdot)$ может быть максимизирована независимо. Результатирующим решением задачи оптимизации является

$$H_{ik}^* = M_{ik} / \sum_j M_{jk}. \quad (6)$$

Каждый раз после контроля этапа проекта значения параметров динамической модели разрушения (6) могут быть скорректированы. Последовательные вычисления динамической модели позволяют предсказывать эволюцию будущего разрушения.

Исходя из вышеизложенного, а также из полученных в работе теоретических результатов исследований динамического управления связями между уровнями инновационно-инвестиционной системы и известного опыта построения оптимальных процедур обслуживания [1], необходимо создание экспертной системы для контроля и обслуживания инновационно-инвестиционного процесса.

Оптимальный процесс обслуживания инновационных проектов существенно связан с процессом воспроизведения функций системы, т.е. способностью системы контролировать свои изменения через изменения множества показателей $\{S_i\}$ диапазона допустимых изменений ΔS_i . Каждый количественный показатель характеризуется порогом чувствительности системы δS_i к его изменениям. При выходе значения показателя S_i за пределы ΔS_i система теряет способность

к самовоспроизведению. Поэтому эта способность системы определяется ее способностью удерживать значения S_i в интервале ΔS_i . В целом эти процедуры обслуживания основываются на компромиссе между расходами на обслуживание инновационного проекта и риском возникновения отказа в таком обслуживании. Однако такой подход нуждается в некотором улучшении. Контролируя состояние расхода бюджета, можно точно определить не один уровень разрушения, а множество наиболее вероятных уровней разрушения (поскольку динамика разрушения бюджета всегда является марковским процессом). Более того, когда принимаются решения по будущему обслуживанию, важную роль играет знание, приобретенное из предыдущих наблюдений разрушения бюджета и соответствующих им процедур обслуживания бюджета.

Этот подход к решениям по обслуживанию является подходом эксперта. В нашем случае аналогичный подход может быть использован для решений по обслуживанию, генерируемых экспертной системой, которая должна уметь выполнять следующие задачи.

1. Основываясь на контроле бюджета и

оценивании показателей расхода, может быть определено подмножество прошлых образцов разрушения бюджета, которые похожи на образцы, взятые в текущее время. Эти прошлые образцы разрушений бюджета хранятся в базе знаний. Образцы вместе с вектором вероятностей $P(t^*)$, связанным с оцененным текущим разрушением бюджета и оцененной матрицей Λ предикатора динамики (рис.2), используются в задаче 2 (t^* является временем текущего контроля бюджета).

2. На основе результатов задачи 1 должна быть выбрана стратегия обслуживания инновационного проекта $V(S)$, $S \geq t^*$, гарантирующая, что процесс реализации проекта будет иметь требуемую надежность до времени следующей проверки.

3. Результатирующую стратегию следует подтвердить путем сравнения со стратегиями, запомненными в базе знаний. Это означает, что выбирается та стратегия, которая будет наиболее подходящей к ожидаемым событиям будущего.

Для того, чтобы выполнить вышеуказанные задачи, разработанная экспертная система организована в виде системы управления с разомкнутой обратной связью с двумя сле-



Рис.2. Структура экспертной системы управления инновационным процессом

дующими отдельными действиями (рис.2):

1) обучением, т.е. распознаванием разрушения бюджета, как инструмента реализации инновационного проекта;

2) управлением, т.е. планированием обслуживания инструмента реализации инновационного проекта.

Первая проблема в настоящее время практически решается путем использования процедуры, сравнивающей (в каждый момент контроля) воздействия многих допустимых стратегий обслуживания, принятых априори. В частности, предположим, что вероятностный вектор $P(t)$ можно расчленить следующим образом:

$$P(t) = [P_R(t), P_F(t)]^T,$$

где $P_R(t)$ и $P_F(t)$ обозначают состояние надежной работы и состояние отказа, соответственно.

Аналогично, матрица Λ может быть расчленена и представлена в следующем виде:

$$\Lambda = \begin{bmatrix} \Lambda_R & 0 \\ 0 & \Lambda_F \end{bmatrix}.$$

Тогда марковскую модель можно записать следующим образом:

$$\dot{P}_R(t) = \Lambda_R P_R(t),$$

$$\dot{P}_F(t) = \Lambda_F P_F(t), \quad (7)$$

где $P(t^*)$ считается известным.

Таким образом, на базе теоретических моделей предложена структура интеллектной системы управления инвестиционно-инновационным процессом в интегрированных производственных комплексах.

СПИСОК ЛИТЕРАТУРЫ:

1. Cumani A., Villa A., Ltv R. Diskrete decription of multi-edge cutter conditions - Model evalnation and parameter identificion for servise life control. Annuls of the CIRP.1984, 33(1).

INTELLECT'S SYSTEM FOR CONTROL OF INNOVATION PROCESS OF INTEGRATED MANUFACTURING SYSTEM

© 2002 A.V. Barvinok¹, A.N. Koptev²

¹Leningrad Metal Plant

²Samara State Aerospace University

We discuss the problem methods of Innovation- investment control in Integrated manufacturing systems, and on the base of theoretical models we suggest the structures of Intellect's and expert system for control of Innovation process at enterprise.

ОБЛАСТИ ДОСТИЖИМОСТИ ПРИ МНОГОКАНАЛЬНОМ УПРАВЛЕНИИ ТРАЕКТОРИЯМИ ЭКСПЕРИМЕНТАЛЬНОГО СУБОРБИТАЛЬНОГО САМОЛЕТА

© 2002 Ю.Н. Лазарев¹, Т.А. Баяндина²

¹ Самарский научный центр РАН

² Самарский государственный аэрокосмический университет

Исследуются маневренные возможности суборбитального самолета при движении в атмосфере. Построены области достижимости для двухканального и трехканального управления с учетом ограничений на управление, режимы движения в атмосфере и терминальные условия. Показаны управляющие зависимости по каналам угла атаки, угла крена и тяги двигателя, обеспечивающие приведение суборбитального самолета в характерные точки областей достижимости.

Введение

Успешное создание многоразовой космической транспортной системы связано с опережающей разработкой экспериментального летательного аппарата, предназначенного для проведения летных экспериментов по отработке наиболее сложных технологий. Обоснованы проекты разработки демонстраторов, способных как совершать суборбитальные полеты для проведения летных исследований, так и выводить на орбиту спутника Земли небольшие полезные нагрузки. В статье в качестве объекта управления рассматривается экспериментальный суборбитальный самолет МАКС-Д – демонстратор технологий перспективных авиационно-космических систем [1].

Самостоятельное движение суборбитального самолета начинается с момента окончания его выведения на суборбитальную траекторию на высоте около 70 км при положительном угле наклона траектории и скорости аппарата значительно меньше крутовой. После достижения высоты 85...90 км суборбитальный самолет совершает управляемый спуск в атмосфере до начала участка предпосадочного маневрирования на высоте около 20 км.

Суборбитальные траектории характеризуются наличием начального восходящего участка и относительно небольшим уровнем кинетической энергии аппарата. Недостаточная скорость не позволяет суборбитальному

самолету реализовать на участке спуска в атмосфере режим квазистационарного плавирования. В зависимости от программ изменения углов атаки и крена могут быть реализованы как крутые траектории с быстрой потерей высоты и большими тепловыми и инерционными нагрузками, так и рикошетирующие траектории, характеризующиеся отражениями аппарата от плотных слоев атмосферы.

Каналы управления и ограничения

Управление движением суборбитального самолета при спуске в атмосфере осуществляется по двум каналам — по каналу угла атаки α и каналу скоростного угла крена γ_a . На управляющие зависимости накладываются ограничения. Минимальные и максимальные значения угла атаки заданы в зависимости от числа Маха M : при $1 \leq M \leq 10$ $\alpha_{\min} = 1^\circ \times M$, $\alpha_{\max} = 3,8^\circ \times M$; при $M > 10$ $\alpha_{\min} = 20^\circ$, $\alpha_{\max} = 45^\circ$. Ограничения на максимальный угол крена на участках снижения $\gamma_{a\max} = 60^\circ$, на участках, прилегающих к точкам рикошета $\gamma_{a\max} = 90^\circ$.

Расширение маневренных возможностей суборбитального самолета возможно при использовании тяги двигателя, в этом случае к двум рассмотренным добавляется третий канал — канал управления секундным массо-

вым расходом топлива β . Необходимым условием реализации трехканального управления является наличие на борту самолета запаса топлива, предназначенного для использования на участке спуска в атмосфере.

Траектории спуска с рикошетами неустойчивы в том смысле, что могут значительно изменять свой профиль при небольшом изменении управляющих зависимостей в процессе управления спуском [2]. Поэтому в качестве номинальных (расчетных) траекторий обычно принимаются траектории без отражений от плотных слоев атмосферы или с небольшим числом отражений и ограничениями на величину отражений. В рассматриваемой задаче на зависимость изменения высоты полета от времени наложено ограничение: траектория спуска может иметь только одно отражение от плотных слоев атмосферы (рикошет), величина которого не должна превышать допустимого значения. Поскольку в процессе формирования номинальной траектории количество рикошетов может меняться, а убирать следует все рикошеты, начиная со второго, то это условие можно сформулировать следующим образом: разность максимальной высоты после отражения и минимальной высоты до отражения не должна превышать трех километров для первого рикошета и не должна быть больше нуля на последующих участках траектории, т.е.

$$H_j^{\max} - H_j^{\min} \leq \Delta H_{j\ don},$$

где $\Delta H_{j\ don} = 3$ км ($j = 1$), $\Delta H_{j\ don} = 0$ ($j = 2, \dots, J$), J – общее число рикошетов на траектории.

Кроме этого условия при спуске в атмосфере нормальная перегрузка n_y на всей траектории спуска не должна превышать заданного значения: $n_y \leq 3,5$.

В начале участка предпосадочного маневрирования отклонения терминалных значений скорости и угла наклона траектории должны находиться внутри области допустимых значений: их требуемые значения и допустимые отклонения от них равны для скорости $V_{\text{треб}} = 500 \text{ м/с}$ при $\Delta V(T) = \pm 30 \text{ м/с}$,

для угла наклона траектории $\theta_{\text{треб}} = -10^\circ$ при $\Delta\theta(T) = \pm 1^\circ$.

Задача управления, условия и метод решения

Задача построения областей достижимости сводится к формированию оптимального многоканального управления, обеспечивающего приведение суборбитального самолета на их границы. Таким образом, требуется сформировать оптимальные управляющие зависимости по каналам углов атаки и крена, а также по каналам углов атаки, крена и тяги двигателя, обеспечивающие достижение суборбитальным самолетом границ областей достижимости при спуске в атмосфере с учетом ограничений на управление, режимы движения в атмосфере и терминальные условия. Границы областей состоят из крайних точек, в которые возможно попадание самолета при спуске, поэтому задачи формирования соответствующих программ управления формулируются как оптимизационные: максимизировать и минимизировать продольную дальность без ограничений и при заданных ограничениях на боковую дальность; максимизировать боковую дальность без ограничений и при заданных ограничениях на продольную дальность. Все оптимизационные задачи включают условия выполнения ограничений на режимы движения в атмосфере, управляющие зависимости и терминальные условия.

В качестве поверхности приведения принималась сфера с центром в центре Земли, проходящая на высоте 20 км над экватором. В модели движения суборбитального самолета учитывалась несферичность поля тяготения Земли и ее вращение вокруг собственной оси. Считалось, что опорная траектория суборбитального движения совпадает с плоскостью экватора.

Параметры атмосферы и аэродинамические характеристики самолета задавались таблично, максимальное значение аэродинамического качества на гиперзвуковых скоростях движения в атмосфере равнялось 1,8. Масса самолета без топлива принималась равной

15000 кг, а масса топлива на управление при спуске в атмосфере – 3000 кг.

Построение областей достижимости проводилось при следующих начальных условиях суборбитального движения: высота 70 км; скорость 4456 м/с (соответствует числу Маха, равному 15); угол наклона траектории $5,5^{\circ}$; значения широты, долготы и угла курса принимались равными нулю.

При решении задач оптимизации номинального управления использовался численный метод формирования многоканального управления движением в атмосфере [3], позволяющий учитывать ограничения на функционалы задачи и управляющие зависимости. Численный метод разработан на основе метода последовательной линеаризации [4], который является типичным методом спуска в пространстве управлений и сводится к построению минимизирующей последовательности управлений.

Области достижимости

Области достижимости суборбитального самолета на поверхности приведения при ограничениях на управление и режимы движения для двухканального (тонкая линия) и трехканального (жирная линия) управления изображены на рис.1. Точки, находящиеся на границе областей достижимости были полу-

чены в результате решения серии оптимизационных задач. На рис.1 отмечены характерные точки с верхним индексом ⁽²⁾ для двухканального управления, а верхним индексом ⁽³⁾ – для трехканального: $1^{(2)}$ и $1^{(3)}$ обозначены точки, соответствующие максимальной продольной дальности; точки с максимальной боковой дальностью обозначены $2^{(2)}$ и $2^{(3)}$; точки с минимальной продольной дальностью обозначены $3^{(2)}$ и $3^{(3)}$.

Как видно из рис.1, использование в качестве третьего канала управления тяги двигателя позволяет значительно расширить область достижимости. В абсолютном измерении в наибольшей степени область достижимости увеличивается в продольном направлении: на 565 км возрастает максимальная продольная дальность, на 30 км уменьшается минимальная продольная дальность. Максимальная боковая дальность возрастает с 390 км до 510 км. В относительном измерении область достижимости в продольном направлении увеличивается в два раза, а в поперечном направлении расширяется приблизительно на 30%.

Траектории

На рис.1 штриховыми линиями показаны трассы полета – проекции на поверхность приведения траекторий, приводящих субор-

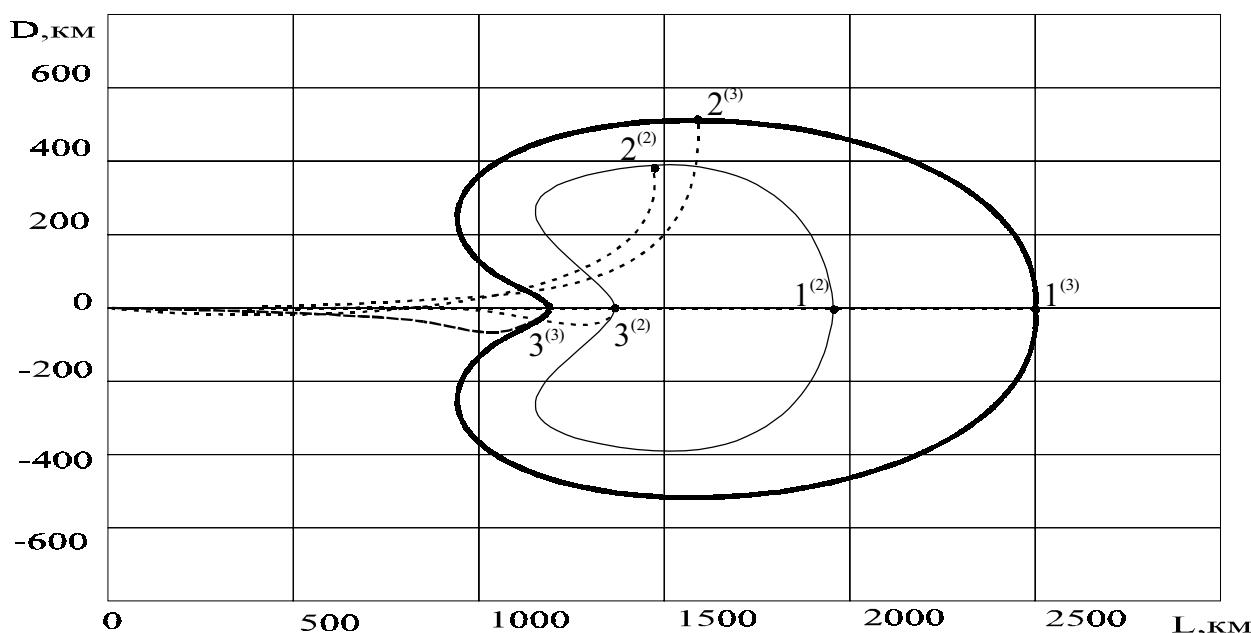


Рис.1. Области достижимости

битальный самолет на границу областей достижимости в характерные точки с максимальной боковой дальностью, а также с минимальной продольной дальностью. Трассы полета на максимальную продольную дальность на большей части траектории практически совпадают с осью симметрии областей достижимости. Несовпадение вызвано отличием угла крена от нуля на участках траектории, на которых управление формируется их условия выполнения ограничений на режимы движения в атмосфере и терминалные условия.

На рис.2 показаны профили этих траекторий – зависимости высоты полета суборбитального самолета от времени. Из рис.2 видно, что полученные траектории обеспечивают спуск самолета с одним отражением от плотных слоев атмосферы, ограниченным по высоте.

Рис.3 и 4 иллюстрируют выполнение в процессе спуска ограничений на удельный тепловой поток в критической точке аппарата с условным радиусом кривизны 1 м и нормальной перегрузки для траекторий полета на максимальную продольную и максимальную боковую дальность.

Из рисунка следует, что использование канала управления тягой двигателя позволяет не только расширить область достижимости аппарата при движении по суборбитальной траектории спуска, но и уменьшить тепловые и инерционные нагрузки на аппарат за счет изменения профиля траектории.

Программы управления

На рис.5-7 показаны программы управления по каждому из каналов, приводящие

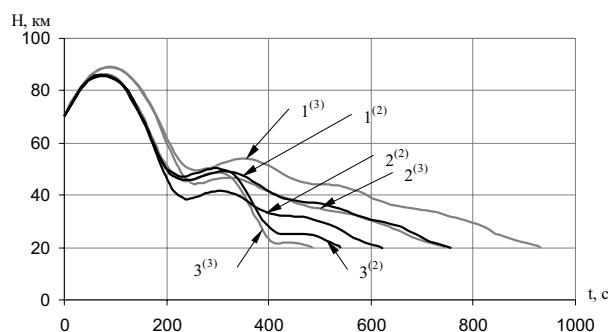


Рис.2. Зависимости высоты полета от времени

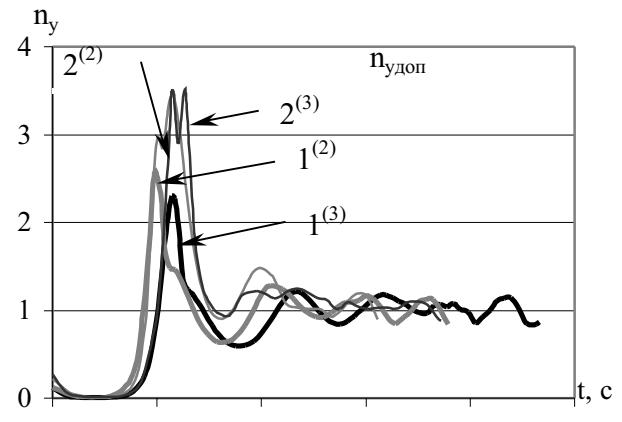


Рис.3. Изменение удельного теплового потока по времени

суборбитальный самолет в характерные точки областей достижимости.

При полете на минимальную продольную дальность программы управления углом атаки (рис.5, кривые $3^{(2)}$ и $3^{(3)}$) обеспечивают минимальное возможное значение аэродинамического качества при максимально допустимом лобовом сопротивлении, для этого управляющие зависимости приближены к предельно допустимым значениям за исключением тех участков траектории, где управление сформировано из условия выполнения ограничения на нормальную перегрузку.

Использование тяги двигателя в качестве дополнительного канала управления приводит к более плавному изменению скоростного угла крена по сравнению с двухканальным управлением (рис.6, кривые $3^{(2)}$ и $3^{(3)}$). При этом структура программы угла атаки изменяется незначительно за исключением участка, где существенно меняется угол курса (от

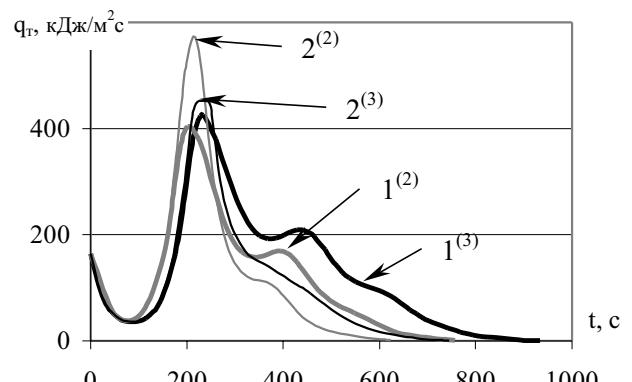


Рис.4. Изменение нормальной перегрузки по времени

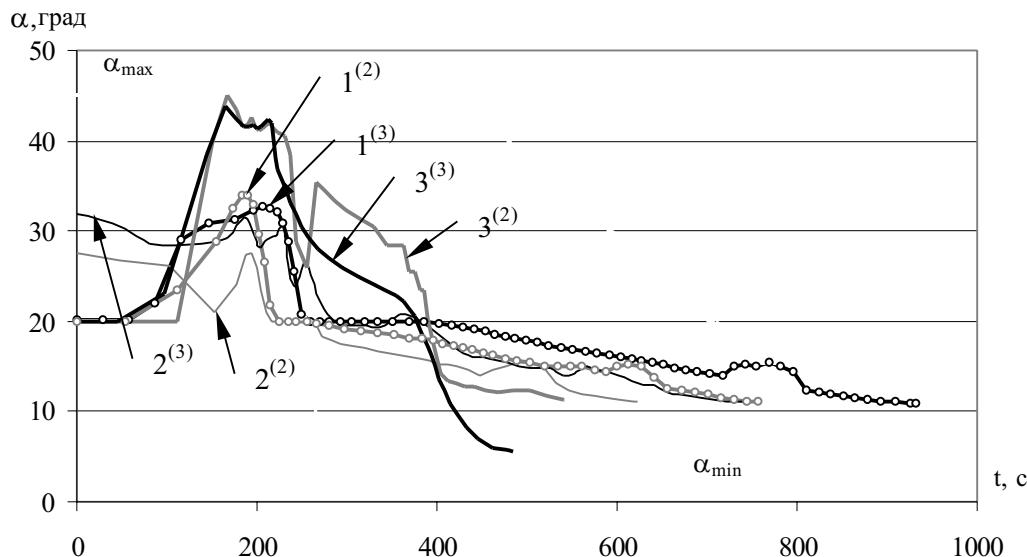


Рис.5. Программы угла атаки

300-й до 400-й секунды), на этом участке угол атаки становится меньше максимально допустимого значения, что позволяет лететь с большим аэродинамическим качеством.

При полете на максимальную продольную дальность программы управления углом атаки (рис.5, кривые $1^{(2)}$ и $1^{(3)}$) обеспечивают максимальное значение аэродинамического качества, кроме участков траектории, на которых управление формируется из условий выполнения ограничений на режимы движения в атмосфере, а также участка, на котором угол крена отличен от нуля в связи с необходимостью уменьшить величину первого отражения аппарата после погружения в плотные слои атмосферы. При трехканальном управлении увеличивается продолжительность

участка спуска с углами атаки, близкими к наивыгоднейшему значению.

При полете на максимальную боковую дальность программы управления углом атаки (рис.5, кривые $2^{(2)}$ и $2^{(3)}$) на большей части траектории обеспечивают наибольшее значение аэродинамического качества, а программы управления углом крена (рис.6, кривые $2^{(2)}$ и $2^{(3)}$) обеспечивают плавное изменение угла курса. Регулярность изменения управляющих зависимостей нарушается на тех участках траектории, где управление формируется из условия выполнения ограничений на режимы движения.

На рис. 7 показаны программы управления массовым секундным расходом топлива для траекторий, приводящих в характерные

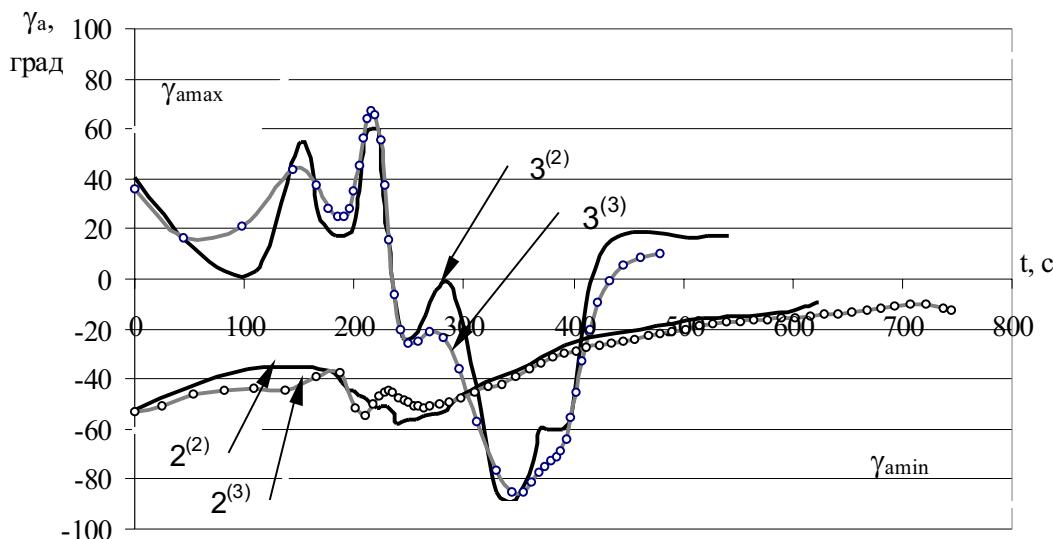


Рис.6. Программы скоростного угла крена

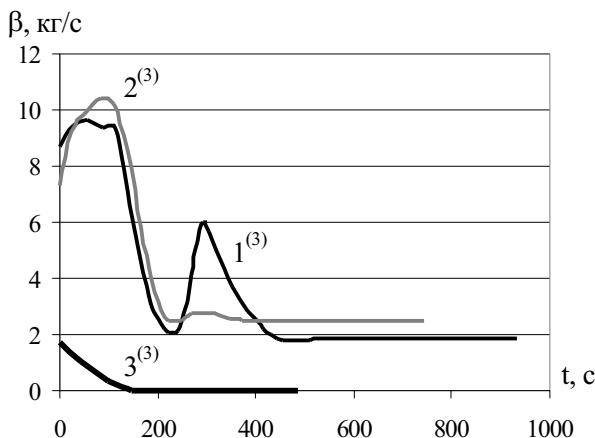


Рис.7. Программы секундного расхода топлива

точки границы области достижимости. Для всех рассмотренных траекторий суборбитального движения тяга двигателя имеет наибольшее значение в начале траектории. При полете на максимальные дальности запас топлива расходуется полностью, при полете на минимальную продольную дальность расходуется небольшая часть топлива.

Заключение

Области достижимости характеризуют предельные маневренные возможности суборбитального самолета при двухканальном и трехканальном управлении и заданных ограничениях. С помощью используемых при построении областей достижимости численных методов на основе последовательной

линеаризации можно сформировать многоканальное управление движением, приводящее суборбитальный самолет во внутренние точки областей достижимости к началу участка предпосадочного маневрирования. Использование совместно с каналами управления углами атаки и крена канала управления тягой двигателя расширяет маневренные возможности суборбитального самолета, обеспечивая возможность приведения в гораздо более обширную область конечных условий движения.

СПИСОК ЛИТЕРАТУРЫ

1. Скороделов В.А. Экспериментальная авиационно-космическая система МАКС-Д // Авиационно-космические системы / Под ред. Г.Е. Лозино-Лозинского и А.Г. Братухина. М.: МАИ, 1997.
2. Ярошевский В.А. Вход в атмосферу космических летательных аппаратов. М.: Наука, 1988.
3. Лазарев Ю.Н. Численный метод формирования многоканального управления движением аэрокосмических аппаратов в атмосфере // Известия Самарского научного центра РАН. 1999. №1.
4. Федоренко Р.П. Приближенное решение задач оптимального управления. М.: Наука, 1978.

LANDING AREAS AND MULTI-CHANNEL TRAJECTORY CONTROL OF EXPERIMENTAL SUBORBITAL AIRCRAFT

© 2002 Yu.N. Lazarev¹, T.A. Bajandina²

¹ Samara Science Centre of Russian Academy of Sciences
² Samara State Aerospace University

Maneuverability of suborbital aircraft motion in atmosphere is discussed. Landing areas for two-channel and three-channel control are determined, with regard to constraints on control, motion modes and terminal conditions. Control programmes for the channels of the angle of attack, angle of roll and engine trust are given, which bring a suborbital aircraft to specific points of landing areas.

РАСЧЕТ МАНЕВРОВ КОРРЕКЦИИ СЛАБОЭЛЛИПТИЧЕСКИХ И КРУГОВЫХ ОРБИТ С ДВИГАТЕЛЕМ МАЛОЙ И КОНЕЧНОЙ ТЯГИ

© 2002 А.А. Храмов, С.А. Ишков

Самарский государственный аэрокосмический университет

Рассматривается задача оптимальной коррекции слабоэллиптических и круговых орбит с двигателями малой и конечной тяги в нецентральном гравитационном поле Земли. Использование модели движения КА в оскулирующих элементах для компонентов вектора Лапласа позволило получить аналитические соотношения для расчета параметров управления, энергетических затрат и времени маневра, справедливые для орбит с малыми эксцентриситетами вплоть до круговых. Показано, что методика расчета может использоваться при более высоких уровнях ускорения тяги. Для оценки методической погрешности решений проведено численное моделирование движения КА.

Значительная часть используемых космических аппаратов (КА) функционируют на круговых и слабоэллиптических орбитах. Как правило, доставка КА с орбит выведения на рабочие орбиты, осуществление маневров коррекции и схода с орбиты реализуется с помощью двигательных установок (ДУ) на жидкостных ракетных двигателях (ЖРД). Для ДУ такого типа практически уже достигнуты предельные значения параметров, характеризующих конструктивное совершенство и энергетические возможности. В связи с этим весьма перспективным представляется использование в качестве маршевых ДУ и двигателей коррекции реактивных двигателей, обладающих большими значениями удельного импульса тяги.

Наряду с электрореактивными двигателями малой тяги (ДМТ), обладающими высоким удельным импульсом, в настоящее время разрабатываются ДУ, которые по уровню тяги занимают промежуточное положение между ДМТ и ЖРД. Это так называемые двигатели конечной или средней тяги. В частности, ведется разработка солнечной тепловой двигательной установки [1] с подогревом рабочего тела и частичным его дожиганием, принцип работы которой заключается в следующем. В течение освещенного участка орбитального полета с помощью солнечных батарей и электронагревателя происходит преобразование световой энергии в тепловую и её накопление в тепловом аккумулято-

ре. Перед подачей компонентов топлива в камеру сгорания с целью повышения удельного импульса осуществляется подогрев горючего (водорода) в тепловом аккумуляторе. Имеется возможность работы двигателя без дожигания водорода в режиме "малой" тяги.

Задача об оптимальных маневрах формирования и коррекции слабоэллиптических орбит с малой тягой рассматривалась в работах [2-4]. В [2, 4] с использованием принципа максимума Понтрягина и метода усреднения получена оптимальная структура управления на витке, численно решена задача оптимальной коррекции солнечно-синхронной орбиты, получены аналитические решения, соответствующие приближенно-оптимальному переходу в центральном гравитационном поле.

В настоящей работе рассматриваются компланарные маневры коррекции круговых и слабоэллиптических орбит ($0 \leq e \leq 0,1$) с двигателями малой и конечной тяги в нецентрическом гравитационном поле Земли. Под коррекцией понимается маневр КА с незначительным изменением большой полуоси орбиты по сравнению с её начальным значением. Управление движением центра масс КА осуществляется с помощью последовательности чередующихся активных и пассивных участков полета. На активных участках реализуется включение ДУ, создающей постоянную по величине тягу. Суммарная продолжительность пассивных участков на витке

считается заданной и постоянной в течение маневра.

Математическая постановка задачи

Для описания орбитального движения КА в плоскости орбиты в центральном гравитационном поле используется математическая модель в оскулирующих элементах [5]:

$$\begin{aligned} \frac{dA}{dt} &= 2\sqrt{\frac{A^3}{\mu}} w_T \frac{R}{1 - q^2 - k^2}, \\ \frac{dq}{dt} &= \sqrt{\frac{p}{\mu}} \left\{ w_S \sin u + w_T \left(\frac{q + \cos u}{R} + \cos u \right) \right\}, \\ \frac{dk}{dt} &= \sqrt{\frac{p}{\mu}} \left\{ -w_S \cos u + w_T \left(\frac{k + \sin u}{R} + \sin u \right) \right\}, \quad (1) \\ \frac{du}{dt} &= \sqrt{\frac{\mu}{p^3}} R^2, \\ \frac{dV_x}{dt} &= \sqrt{w_S^2 + w_T^2}, \end{aligned}$$

где $p = A(1 - q^2 - k^2)$ – фокальный параметр орбиты, A – большая полуось, q и k – компоненты вектора Лапласа, u – аргумент широты, V_x – характеристическая скорость перелета, $w_T = w \cos \alpha$ и $w_S = w \sin \alpha$ – компоненты вектора ускорения тяги ДУ вдоль трансверсального и радиального направлений соответственно, $w = P/m$ – ускорение силы тяги, P – тяга ДУ, m – текущая масса КА, α – угол между вектором тяги ДУ и трансверсалю, μ – гравитационный параметр, $R = 1 + q \cos u + k \sin u$. Компоненты вектора Лапласа связаны с эксцентриситетом e и аргументом перигея ω орбиты следующими соотношениями: $q = e \cos \omega$, $k = e \sin \omega$.

С целью упрощения математической модели с учетом специфики решаемой задачи вводятся следующие ограничения на управление и параметры орбиты:

- рассматривается ДУ с нерегулируемой тягой и скоростью истечения;
- вектор ускорения ориентируется вдоль трансверсали $w_S = 0$;
- вследствие малого расхода рабочего тела в процессе коррекции ускорение тяги

считается постоянным $w_T = \delta w_0$, где $\delta \in \{-1, 0, 1\}$ – функция включения тяги, w_0 – ускорение тяги в начальный момент времени;

– орбиты движения определяются эксцентриситетами $0...0,1$;

– предусматривается наличие пассивных участков на витке заданной продолжительности, обусловленные техническими особенностями ДУ, выполнением целевой задачи и др.

После перехода к новой независимой переменной u с учетом введенных ограничений, уравнения движения центра масс КА запишутся в виде:

$$\begin{aligned} \frac{dA}{du} &= 2 \frac{A^3}{\mu} \delta w_0, \\ \frac{dq}{du} &= 2 \frac{A^2}{\mu} \delta w_0 \cos u, \quad (2) \\ \frac{dk}{du} &= 2 \frac{A^2}{\mu} \delta w_0 \sin u, \\ \frac{dV_x}{du} &= \sqrt{\frac{A^3}{\mu}} |\delta| w_0. \end{aligned}$$

В качестве критерия оптимальности траектории перехода принимается минимум затрат характеристической скорости $V_x \rightarrow \min$. Отметим, что при постоянной величине реактивного ускорения ДУ и постоянной суммарной длительности пассивных участков на витке этот критерий эквивалентен критерию минимального времени перелета. Границные условия орбитального перехода записываются как:

$$\begin{aligned} u = 0, \quad A = A_0, \quad k = k_0, \quad q = q_0. \\ u = u_k, \quad A = A_k, \quad k = k_k, \quad q = q_k. \quad (3) \end{aligned}$$

Характерные для рассматриваемых уровней ускорений ($w_0/g < 10^{-3}$) многовитковые межорбитальные перелеты позволяют упростить математическую модель движения с помощью метода усреднения. Указанный метод предполагает разделение управления на периодическую (быстро меняющуюся) и

вековую (медленно меняющуюся) составляющие и отдельную их оптимизацию [4]. Таким образом, поиск оптимального управления состоит из двух этапов. Первый этап – определение оптимального управления в пределах витка (локальная оптимизация); второй – построение усредненной модели движения и определение оптимального управления вековым изменением параметров орбиты.

Выбор структуры управления на витке

Для решения задачи локальной оптимизации воспользуемся принципом максимума Понтрягина. Гамильтониан для системы (2) и принятого критерия оптимальности записывается в виде:

$$H = \frac{dA}{du} \psi_A + \frac{dq}{du} \psi_q + \frac{dk}{du} \psi_k + \frac{dV_x}{du} \psi_{Vx} - \frac{dV_x}{du} \rightarrow \max,$$

где $\psi_A, \psi_q, \psi_k, \psi_{Vx}$ – сопряженные множители.

После преобразований получаем:

$$H = 2 \frac{A^2}{\mu} \delta w_0 \{ A \psi_A + \psi_q \cos u + \psi_k \sin u \} - \sqrt{\frac{A^3}{\mu}} |\delta| w_0 \{ 1 - \psi_{Vx} \}.$$

Запишем соотношения для сопряженных множителей:

$$\begin{aligned} \frac{d\psi_A}{du} &= -\frac{\partial H}{\partial A} = -4 \frac{A}{\mu} \delta w_0 \times \\ &\times \left\{ \frac{3}{2} A \psi_A + \psi_q \cos u + \psi_k \sin u \right\} + \\ &+ \frac{3}{2} \sqrt{\frac{A}{\mu}} |\delta| w_0 \{ 1 - \psi_{Vx} \}, \end{aligned}$$

$$\frac{d\psi_q}{du} = \frac{d\psi_k}{du} = \frac{d\psi_{Vx}}{du} = 0, \quad (4)$$

$$\psi_q = \text{const}, \quad \psi_k = \text{const}, \quad \psi_{Vx} = \text{const}.$$

Из уравнений (2) и (4) видно, что скоро-

сти изменения параметров A, q, k, V_x и ψ_A пропорциональны малому параметру w_0 и, таким образом, в гамильтониан входит единственная быстрая переменная – аргумент широты u . Оптимальное управление δ_{opt} находится из условия максимума гамильтониана как релейная функция аргумента широты. Точки включения и выключения ДУ будут определяться уравнением:

$$2 \frac{A^2}{\mu} |A \psi_A + \psi_q \cos u + \psi_k \sin u| - \{ 1 - \psi_{Vx} \} = 0,$$

а направление ускорения тяги активных участков:

$$\delta_{opt} = \text{sign} \{ A \psi_A + \psi_q \cos u + \psi_k \sin u \}.$$

Анализ полученного решения показывает, что в общем случае структура оптимального управления на витке будет включать два активных участка с разными знаками тяги, центры которых разнесены по аргументу широты на 180° и разделены двумя одинаковыми пассивными участками (рис.1). Под витком здесь и далее понимается угловой интервал по аргументу широты в 360° , отсчитываемый от центра одного из активных участков. В частном случае один из активных участков может отсутствовать. Подобная структура управления, включающая разгонный и тормозной участки, была получена в работах

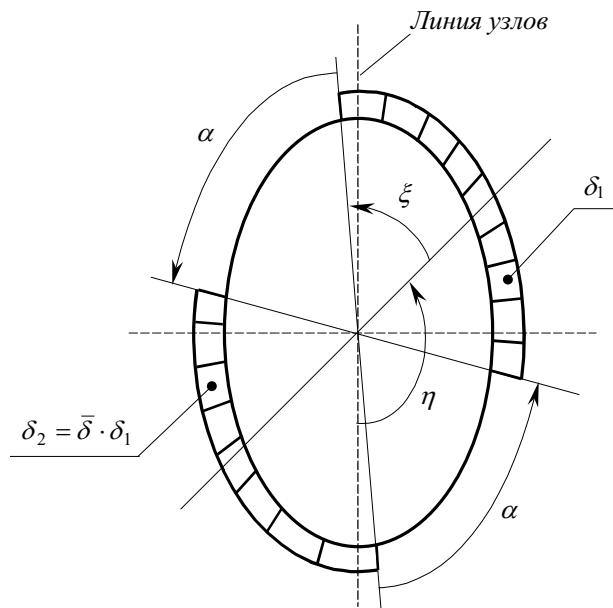


Рис.1. Структура управления на витке

[2-4]. Однако, как показали исследования, при определенных граничных условиях и наличии пассивных участков на витке коррекция орбиты с активными участками противоположных знаков невозможна. В этом случае маневр может быть осуществлен если на витке расположить два активных участка одного знака. Для принятой в данной работе структуры управления, являющейся универсальной относительно граничных условий и длительности пассивных участков на витке, описанные варианты являются частными случаями.

Построение усредненной модели движения и вывод аналитических решений

Используя полученную структуру управления, перейдем к построению модели векового изменения параметров орбиты. В качестве параметров управления вводятся следующие величины: ξ – половина ширины первого активного участка со знаком тяги δ_1 и аргументом широты его центра η , $\bar{\delta}$ – отношение знаков тяг активных участков (рис.1). Ширина одного пассивного участка α задана и постоянна в течение маневра. Размеры участков измеряются по аргументу широты.

Усредня систему уравнений движения (2) с учетом введенных параметров управления и переходя к независимой переменной t , получаем:

$$\begin{aligned} \frac{dA}{dt} &= 2\sqrt{\frac{A^3}{\mu}} \frac{w_0}{\pi} \delta_1 [\xi + \bar{\delta}(\pi - \alpha - \xi)], \\ \frac{dq}{dt} &= 2\sqrt{\frac{A}{\mu}} \frac{w_0}{\pi} \delta_1 [\sin \xi - \bar{\delta} \sin(\xi + \alpha)] \cos \eta, \quad (5) \\ \frac{dk}{dt} &= 2\sqrt{\frac{A}{\mu}} \frac{w_0}{\pi} \delta_1 [\sin \xi - \bar{\delta} \sin(\xi + \alpha)] \sin \eta, \\ \frac{dV_x}{dt} &= \left(1 - \frac{\alpha}{\pi}\right) w_0. \end{aligned}$$

Известно, что на вековое изменение параметров орбиты сильное влияние оказывает нецентральность гравитационного поля

Земли. Если в разложении геопотенциала учесть только вторую зональную гармонику, то вековые возмущения будут испытывать элементы q и k , равные [6]:

$$\left(\frac{dq}{dt}\right)_g = -\frac{B}{A^{3,5}} k, \quad \left(\frac{dk}{dt}\right)_g = \frac{B}{A^{3,5}} q,$$

$$\text{где } B = \frac{\varepsilon(5 \cos^2 i - 1)}{2\sqrt{\mu}}, \quad \varepsilon = 2,6333 \cdot 10^{10} \frac{\text{км}^5}{c^2}.$$

С целью получения аналитических решений в правых частях уравнений (5) будем пренебрегать изменением большой полуоси орбиты, величина которой приближенно равна:

$$A \approx A_{cp} = \frac{A_0 + A_k}{2}.$$

Таким образом, упрощенная математическая модель вековой прецессии слабоэллиптической орбиты в поле земного сферида будет иметь вид:

$$\begin{aligned} \frac{dA}{dt} &= 2\sqrt{\frac{A_{cp}^3}{\mu}} \frac{w_0}{\pi} \delta_1 [\xi + \bar{\delta}(\pi - \alpha - \xi)], \\ \frac{dq}{dt} &= 2\sqrt{\frac{A_{cp}}{\mu}} \frac{w_0}{\pi} \delta_1 \times \\ &\times [\sin \xi - \bar{\delta} \sin(\xi + \alpha)] \cos \eta - \frac{B}{A_{cp}^{3,5}} k, \quad (6) \\ \frac{dk}{dt} &= 2\sqrt{\frac{A_{cp}}{\mu}} \frac{w_0}{\pi} \delta_1 \times \\ &\times [\sin \xi - \bar{\delta} \sin(\xi + \alpha)] \sin \eta + \frac{B}{A_{cp}^{3,5}} q. \end{aligned}$$

Найдем оптимальный по быстродействию закон управления параметрами орбиты в течение длительных интервалов времени (более одного витка). В соответствии с принципом максимума Понтрягина гамильтониан для системы (6) имеет вид:

$$\begin{aligned} H &= -1 + a A_{cp} \delta_1 \{ \xi + \bar{\delta}(\pi - \alpha - \xi) \} \psi_A + \\ &+ a \delta_1 \{ \sin \xi - \bar{\delta} \sin(\xi + \alpha) \} \{ \psi_q \cos \eta + \psi_k \sin \eta \} + \\ &+ b (\psi_k q - \psi_q k) \rightarrow \max, \end{aligned}$$

$$\text{где } a = 2 \sqrt{\frac{A_{cp}}{\mu} \frac{w_0}{\pi}}, \quad b = \frac{B}{A_{cp}^{3,5}}.$$

Уравнения для сопряженных множителей получаются в виде:

$$\frac{d\psi_A}{dt} = 0, \quad \frac{d\psi_q}{dt} = -b\psi_k, \quad \frac{d\psi_k}{dt} = b\psi_q,$$

откуда:

$$\begin{aligned} \psi_A &= \psi_{A_0} = \text{const}, \\ \psi_q &= \sqrt{\psi_{k_0}^2 + \psi_{q_0}^2} \cos(\eta_0 + bt), \\ \psi_k &= \sqrt{\psi_{k_0}^2 + \psi_{q_0}^2} \sin(\eta_0 + bt), \end{aligned} \quad (7)$$

где

$$\eta_0 = \arcsin \frac{\psi_{k_0}}{\sqrt{\psi_{k_0}^2 + \psi_{q_0}^2}} = \arccos \frac{\psi_{q_0}}{\sqrt{\psi_{k_0}^2 + \psi_{q_0}^2}}.$$

Закон изменения параметров управления определяются из уравнений для экстремалей:

$$\begin{aligned} \frac{dH}{d\xi} &= \psi_{A_0} a A_{cp} (1 - \bar{\delta}) + \\ &+ a \delta_1 \{ \psi_q \cos \eta + \psi_k \sin \eta \} \times \\ &\times \{ \cos \xi - \bar{\delta} \cos(\xi + \alpha) \} = 0, \end{aligned}$$

$$\begin{aligned} \frac{dH}{d\eta} &= a \delta_1 \{ \sin \xi - \bar{\delta} \sin(\xi + \alpha) \} \times \\ &\times \{ \psi_k \cos \eta - \psi_q \sin \eta \} = 0. \end{aligned}$$

Из последнего уравнения с учетом (7) получаем:

$$\operatorname{tg} \eta_{opt} = \frac{\psi_k}{\psi_q} = \operatorname{tg}(\eta_0 + bt), \text{ или } \eta_{opt} = \eta_0 + bt.$$

Далее для параметра ξ :

$$\begin{aligned} \cos \xi_{opt} - \bar{\delta} \cos(\xi_{opt} + \alpha) &= \\ &= -\psi_{A_0} \frac{1 - \bar{\delta}}{\sqrt{\psi_{k_0}^2 + \psi_{q_0}^2}} = \text{const}, \end{aligned}$$

откуда следует что:

$$\xi_{opt} = \xi_0 = \text{const}.$$

Таким образом, оптимальная по быстродействию коррекция осуществляется при ли-

нейном изменении расположения центра активных участков относительно линии узлов (параметр η) и постоянной их величине (параметр ξ). Система уравнений движения (6) при полученном законе управления интегрируется аналитически, что позволяет записать:

$$A = \frac{A_0}{(1 - \varphi t)^2},$$

$$\begin{aligned} q &= e_0 \cos(\omega_0 + \frac{B}{A_{cp}^{3,5}} t) + \\ &+ vt \cos(\eta_0 + \frac{B}{A_{cp}^{3,5}} t); \end{aligned} \quad (8)$$

$$k = e_0 \sin(\omega_0 + \frac{B}{A_{cp}^{3,5}} t) + vt \sin(\eta_0 + \frac{B}{A_{cp}^{3,5}} t);$$

$$\text{где } \varphi = \frac{w_0}{\pi} \sqrt{\frac{A_0}{\mu}} \delta_1 \{ \xi + \bar{\delta} (\pi - \alpha - \xi) \},$$

$$v = 2 \sqrt{\frac{A_{cp}}{\mu} \frac{w_0}{\pi}} \delta_1 \{ \sin \xi - \bar{\delta} \sin(\xi + \alpha) \},$$

$$e_0 = \sqrt{k_0^2 + q_0^2}, \quad \omega_0 = \operatorname{arctg} \frac{k_0}{q_0} \text{ - эксцентри-}$$

ситет и аргумент перигея начальной орбиты. После преобразования системы (8) с учетом ограничения $0 \leq \xi \leq \pi - \alpha$ параметры управления выражаются через граничные условия, величину пассивных участков на витке и ускорение тяги ДУ:

$$\bar{\delta} = \operatorname{sign} \left\{ 1 - \frac{\pi - \alpha}{|\lambda| \sin \alpha} \right\},$$

$$\delta_1 = \frac{1 + \bar{\delta}}{2} \operatorname{sign}(A_k - A_0) + \frac{1 - \bar{\delta}}{2},$$

$$\xi = \arccos \left\{ -\frac{\pi - \alpha}{2 \lambda \sin \frac{\alpha}{2}} \right\} - \frac{\alpha}{2}, \text{ при } \bar{\delta} = 1, \quad (9)$$

$$\xi = \frac{\pi - \alpha}{2} + \lambda \sin \left(\xi + \frac{\alpha}{2} \right) \cos \frac{\alpha}{2}, \text{ при } \bar{\delta} = - ,$$

$$\eta_0 = \omega_0 + \operatorname{arctg} \frac{e_k \sin \Delta\omega_a}{e_k \cos \Delta\omega_a - e_0}, \quad \text{где}$$

$$\lambda = \frac{2\sqrt{A_{cp}}}{\sqrt{e_k^2 - 2e_k e_0 \cos \Delta\omega_a + e_0^2}} \left\{ \frac{1}{\sqrt{A_0}} - \frac{1}{\sqrt{A_k}} \right\},$$

$$\Delta\omega_a = \omega_k - \omega_0 - \frac{B}{A_{cp}^{3,5}} t_k \quad \text{изменение аргумента перигея под действием тяги ДУ,}$$

$$t_k = \frac{\sqrt{\mu}}{w_0} \left\{ \frac{1}{\sqrt{A_0}} - \frac{1}{\sqrt{A_k}} \right\} \delta_1 \frac{\pi}{\xi + \bar{\delta}(\pi - \alpha - \xi)}.$$

время маневра.

Таким образом, в случае активных участков одного знака ($\bar{\delta} = +$) их длительность определяется в явном виде, в случае противоположных ($\bar{\delta} = -1$) – в результате решения трансцендентного уравнения. Затраты характеристической скорости в соответствии с (5) рассчитываются по формуле:

$$V_x = w_0 \left(1 - \frac{\alpha}{\pi} \right) t_k = \sqrt{\mu} \times \\ \times \left\{ \frac{1}{\sqrt{A_0}} - \frac{1}{\sqrt{A_k}} \right\} \delta_1 \frac{\pi - \alpha}{\xi + \bar{\delta}(\pi - \alpha - \xi)}. \quad (10)$$

На основе полученных решений формируется следующая методика расчета временных и энергетических затрат при оптимальной коррекции. По заданным граничным условиям перехода и величине пассивных участков на витке определяются тип управления $\bar{\delta}$, знак тяги δ_1 и длительность ξ первого активного участка, расположение активных участков в начальный момент времени относительно линии узлов η_0 . Время коррекции определяется при расчете параметров управления, а энергетические затраты по соотношению (10).

Полученные решения можно отобразить графически в виде изолиний затрат характеристической скорости в плоскости корректируемых параметров $\Delta A/A_0$ и Δe для заданных значений $\Delta\omega_a$. Изолинии математически описываются следующими соотношениями:

$$\Delta A = \left\{ \frac{1}{\sqrt{A_0}} \pm \frac{V_x}{\sqrt{\mu}} \right\}^{-2} - A_0, \quad \text{при } \bar{\delta} = +1,$$

$$\Delta e = -e_0(1 - \cos \Delta\omega_a) \pm \\ \pm \sqrt{\left\{ \frac{4}{\pi - \alpha} \sqrt{\frac{A_{cp}}{\mu}} V_x \cos \beta \cos \frac{\alpha}{2} \right\}^2 - e_0^2(1 - \cos^2 \Delta\omega_a)},$$

при $\bar{\delta} = -1$,

$$\text{где } \beta = \frac{\pi - \alpha}{2} \frac{\sqrt{\mu}}{V_x} \left\{ \frac{1}{\sqrt{A_0}} - \frac{1}{\sqrt{A_k}} \right\}.$$

Границы областей управлений с активными участками одного (управление 1-го типа) и противоположных (управление 2-го типа) знаков определяются как:

$$\Delta e = -e_0(1 - \cos \Delta\omega_a) \pm \\ \pm \sqrt{F^2 - e_0^2(1 - \cos^2 \Delta\omega_a)},$$

$$\text{где } F = 2 \left\{ \frac{1}{\sqrt{A_0}} - \frac{1}{\sqrt{A_k}} \right\} \sqrt{A_{cp}} \frac{\sin \alpha}{\pi - \alpha}.$$

На рис.2 построены области корректируемых параметров, определяющие маневренные возможности КА при заданном значении располагаемой характеристической скорости. Тонкие линии разделяют плоскость корректируемых параметров на области управления разных типов. Как видно, области управления 1-го типа вытянуты вдоль оси, соответствующей изменению большой полуоси орбиты. Таким образом, управление на витке будет иметь активные участки одного знака в случае преимущественной коррекции большой полуоси $\Delta A/A_0$ и противоположных знаков при значительных коррекциях эксцентриситета Δe и аргумента перигея $\Delta\omega_a$. Изолинии V_x для управления 1-го типа – прямые линии; в этом случае энергетические затраты зависят только от требуемого изменения большой полуоси орбиты и не зависят от необходимых приращений эксцентриситета и аргумента перигея, коррекция которых осуществляется "автоматически" при соответствующей величине и расположении активных участков на витке относительно линии узлов.

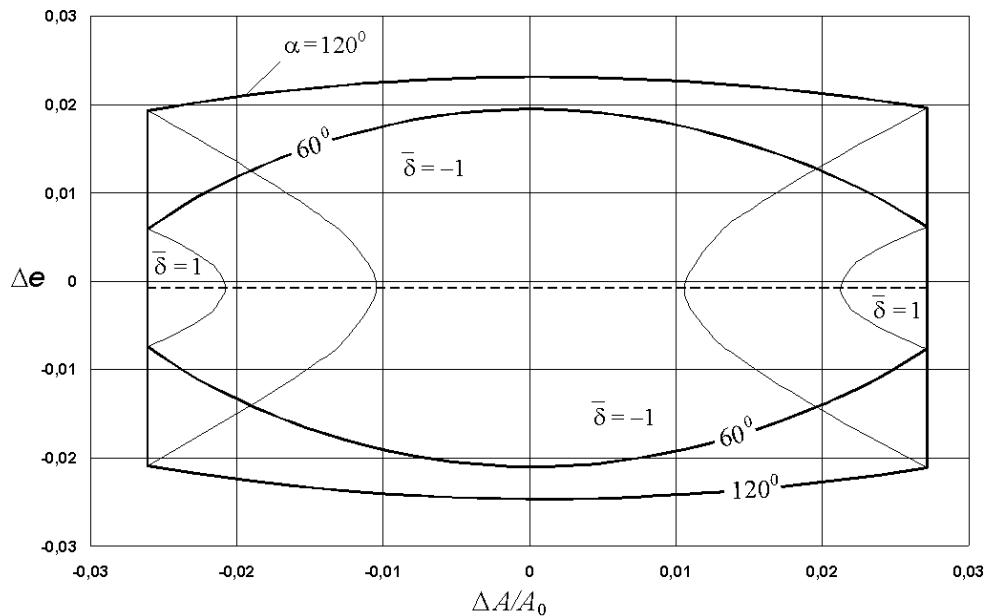


Рис.2. Области корректируемых параметров ($\bar{V}_x = \sqrt{\frac{A_0}{\mu}} V_x = 0,013$; $\Delta\omega_a = 10^0$; $e_0 = 0,01$)

При управлении 2-го типа затраты V_x будут функцией всех трёх корректируемых параметров $\Delta A/A_0$, Δe и $\Delta\omega_a$. Изолинии имеют точки излома, соответствующие решению задачи с одним активным участком на витке.

На рис.3 представлены границы областей в плоскости требуемых изменений эксцентризитета и аргумента перигея, где управление на витке будет состоять из активных участков одного (внутри области) и противоположных (вне области) знаков. Можно

отметить, что тип управления определяется как требуемыми приращениями корректируемых параметров, так и величиной пассивного участка. При увеличении пассивного участка или приращения большой полуоси орбиты управление с активными участками одного знака обеспечит коррекцию больших значений эксцентризитета и аргумента перигея. Если пассивный участок на витке отсутствует или нет необходимости изменять большую полуось орбиты, то область, соответ-

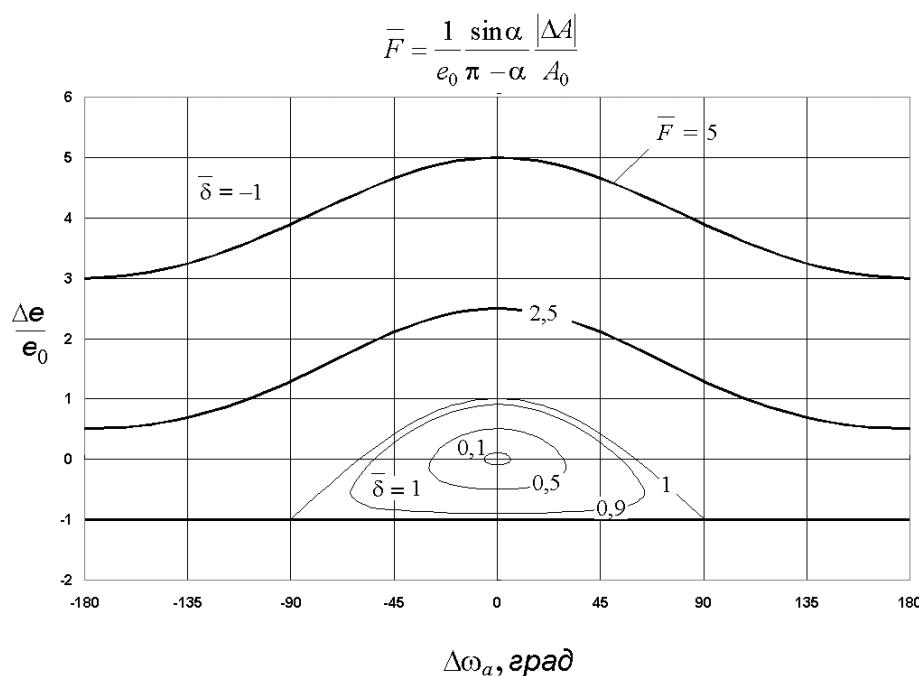


Рис.3. Области, соответствующие управлению разных типов

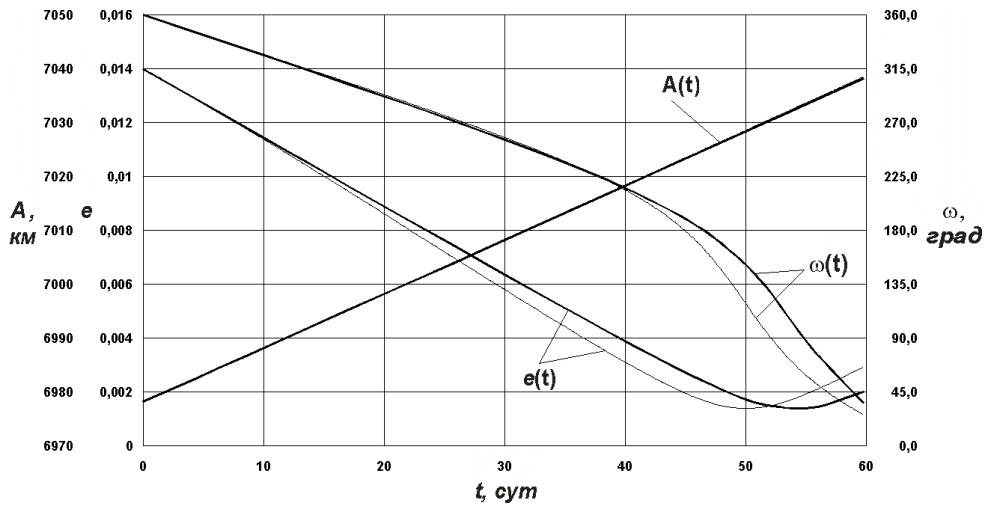


Рис.4. Зависимость параметров орбиты от времени в процессе коррекции ($w_0 = 2 \cdot 10^{-5}$ м/с)

ствующая управлению первого типа преобразуется в точку; в этом случае коррекция возможна лишь при управлении с активными участками противоположного знаков.

Проводилось численное моделирование маневров коррекции солнечно-синхронной орбиты на модели движения в оскулирующих элементах при следующих исходных данных:

$$A_0 = 6978, \quad e_0 = 0,014, \quad \omega_0 = 0 \text{ град.}$$

$$A_k = 7038, \quad e_k = 0,002, \quad \omega_k = 36 \text{ град.}$$

При моделировании учитывались возмущения от нецентральности гравитационного поля Земли (2-ая зональная гармоника). Значения параметров управления вычислялись по уравнениям (9). На рис.4 приведены

зависимости изменения параметров орбиты при коррекциях без пассивных участков на витке, полученные в результате численных расчетов (тонкие линии) и по аналитическим соотношениям (8) при ускорениях тяги $2 \cdot 10^{-5}$ м/с. На рис.5 показаны результаты расчетов при величине ускорения тяги $2 \cdot 10^{-2}$ м/с и суммарной длительности активных участков на витке 5 град. Таким образом, полученные решения могут использоваться и при более высоких уровнях ускорения тяги при некотором уменьшении продолжительности активных участков на витке.

Численные расчеты показывают, что полученные аналитические решения являются достаточно точными и могут быть исполь-

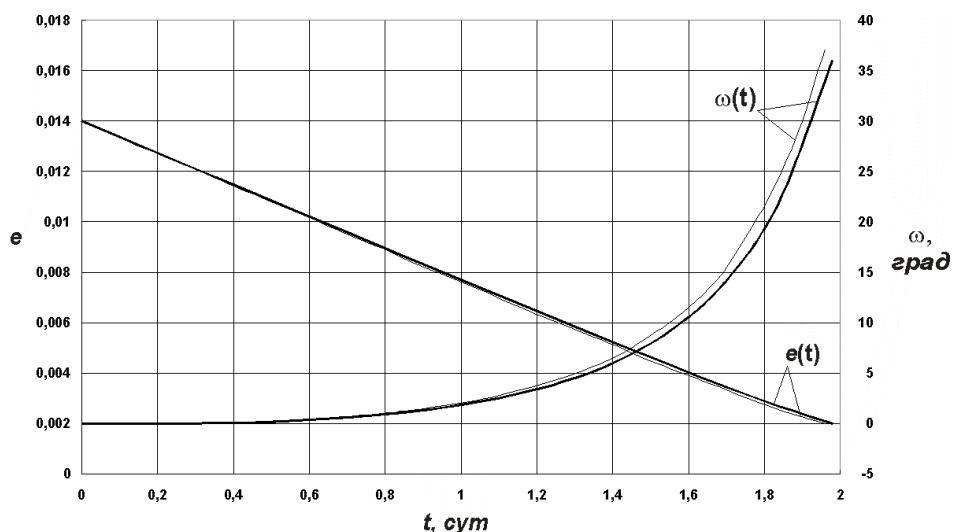


Рис.5. Зависимость параметров орбиты от времени в процессе коррекции ($w_0 = 2 \cdot 10^{-2}$ м/с)

зованы для оценки баллистических характеристик маневров КА с двигателем малой и конечной тяги.

СПИСОК ЛИТЕРАТУРЫ

1. Усталов Ю.М., Храмов А.А. Анализ некоторых схем перевода КА на геостационарную орбиту с использованием солнечной тепловой двигательной установки ограниченной тяги // Сборник трудов девятого Всероссийского научно-технического семинара по управлению движением и навигации ЛА. Ч.1. Самара. 1999.
2. Юрин В.В. Оптимальная коррекция параметров орбиты космического аппарата с двигателем малой тяги // Космич. исслед. 1983. Т.21. №5.
3. Салмин В.В., Соколов В.О. Приближенный расчет маневров формирования орбиты спутника Земли с двигателем малой тяги // Космич. исслед. 1991. Т.29. №6.
4. Салмин В.В. Оптимизация космических перелетов с малой тягой. Проблемы совместного управления траекторным и угловым движением. М.: Машиностроение, 1987.
5. Эльясберг П.Е. Введение в теорию полета искусственных спутников Земли. М.: Наука, 1965.
6. Нариманов Г.С., Тихонравов М.К. Основы теории полета космических аппаратов. М.: Машиностроение, 1972.

CORRECTION MANEUVERS CALCULATION OF QUASI ELLIPTICAL AND CIRCULAR ORBITS BY ENGINE WITH LOW AND LIMIT THRUST

© 2002 A.A. Khramov, S.A. Ishkov

Samara State Aerospace University

The problem of optimal correcting quasi elliptical and circular orbits by engine with low and limit thrust in the Earth non-center gravitational field is considered. Application of Laplace vector components in spacecraft motion model has allowed to get analytical solutions. Such solutions are equitable for calculation of control parameters, energy consumption and maneuver time for orbits with small eccentricity up to circular. It is shown that methodology of calculation can be used under conditions with more high levels of thrust acceleration. For the evaluation of methodical inaccuracy of solutions numerical simulation of spacecraft motion is conducted.

УДК: 519.713.2+681.335

КЛАССИФИКАЦИЯ РЕЛЯТОРОВ ПО ДОМИНАНТНЫМ ПРИЗНАКАМ

© 2002 Л.И. Волгин¹, А.И. Зарукин², А.Б. Клиновский¹

¹ Ульяновский государственный технический университет

² Ульяновский филиал Военной академии тыла и транспорта

Приведена классификация реляторов - логических схемных элементов универсального применения, воспроизводящих бинарные операции ряда алгебраических логик (специальных алгебр и математических логик) в континуальной, дискретной (включая двузначную булеву алгебру логики) областях.

Логико-алгебраические основы реляторной схемотехники

Реляторы являются логическими схемными элементами, воспроизводящими бинарные (базовые) операции импликативной алгебры выбора (ИАВ) [1]

$$Z_1 = \Lambda(y_1, y_2) = y_1 I(x_2 - x_1) + y_2 I(x_1 - x_2), \quad (1a)$$

$$Z_2 = V(y_1, y_2) = y_1 I(x_1 - x_2) + y_2 I(x_2 - x_1) \quad (1b)$$

и сопутствующие им изломные и разрывные элементарные функции.

В(1) $I(x)$ есть единичная функция (оператор Хевисайда) равная нулю при $x < 0$ и единице при $x > 0$, $I(0) \in \{0,1\}$, $I=(I_{12}, I_{21})$ - кортеж (вектор) весовых коэффициентов, $I_{12}=I(x_1-x_2)$, $I_{21}=1-I_{12}=I(x_1-x_2)$, $I_{12}+I_{21}=1$ (условие комплементарности), y_1 и y_2 - предметные переменные (в общем случае любые физические и математические объекты, удовлетворяющие условиям мультиплексивного выделения $y_i \cdot 1 = y_i$ и поглощения $y_i \cdot 0 = 0$), x_1 и x_2 - предикатные переменные (действительные числа).

При отождествлении предметных и предикатных переменных ($y_1 = x_1$, $y_2 = x_2$) операции ИПА-конъюнкции (1a) и ИПА-дизъюнкции (1b) вырождаются в базовые бинарные операции непрерывной логики (НЛ) [2-4]

$$Z_1 = \Lambda(x_1, x_2) = \min(x_1, x_2), \quad (2)$$

$$Z_2 = V(x_1, x_2) = \max(x_1, x_2).$$

Если континуальные переменные в (2) принимают к разрешённых уровням, то выражение (2) вырождается в бинарные операции многозначной (k -значной) логики (МЛ),

которые при $k=2$ являются конъюнкцией

$Z_1 = x_1 \wedge x_2$ и дизъюнкцией $Z_2 = x_1 \vee x_2$ двузначной булевой алгеброй выбора (БА).

В свою очередь ИПА-операции являются частными реализациями бинарных операций предикатной алгебры выбора (ПАВ) [5, 6]

$$Z_1 = \Lambda(y_1, y_2) = y_1 \alpha_2 + y_2 \alpha_1, \quad (3)$$

$$Z_2 = V(y_1, y_2) = y_1 \alpha_1 + y_2 \alpha_2,$$

где $\alpha_1 \in \{0,1\}$, $\alpha_2 \in \{1,0\}$, $\alpha_1 + \alpha_2 = 1$ (условие комплементарности весовых коэффициентов α_1 и α_2), $A = (\alpha_1, \alpha_2)$ - кортеж весовых коэффициентов.

Выражения (1) - (3) воспроизводят различного типа операции альтернативного выбора в континуальной и дискретных (включая двузначную) областях в едином элементном базисе реляторов [7]. В свою очередь выражения (3) при $\alpha_1 + \alpha_2 = 1$ (условие комплементарности) являются базовыми операциями комплементарной алгебры (КА) [5, 6], при этом условии бинарности $\alpha_1, \alpha_2 \in \{0,1\}$ не накладывается.

Указанные алгебраические логики (специальные алгебры и предметные логики) входят в базовую ветвь логико-алгебраической метасистемы, порождаемой функцией-аксиомой взвешенных степенных средних [8].

Принципиальным отличием импликативной алгебры выбора от других алгебраических логик заключается в том, что ИАВ определена на двух множествах - множествах предметных $Y \in \{y_1, y_2, \dots, y_n\}$ и предикатных $X \in \{x_1, x_2, \dots, x_m\}$ переменных, что необозримо расширяет её функциональные и алго-

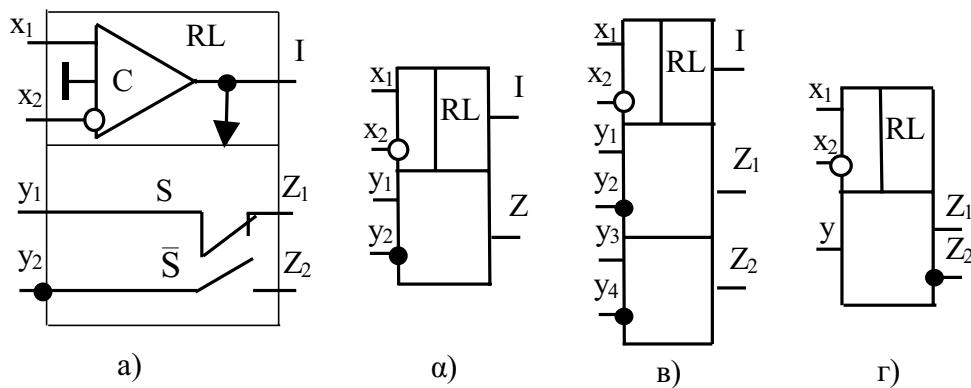


Рис.1. Виды реле

ритмические возможности (включая ИАВ-модели).

Классификация реле

Реле содержат один (рис.1) или два компаратора C_i , которые управляют состоянием группы замыкающих S_i и размыкающих \bar{S}_i аналоговых ключей. На компараторные и переключательные входы реле подаются соответственно предметные y_1, y_2 и предикатные x_1, x_2 переменные.

Номенклатура различного типа реле естественным образом разделяется по признаку физической природы предметных y_i и предикатных x_i переменных - напряжения и (или) тока. В реле используются следующие типы компараторов: нестробируемые, стробируемые, с однобитной памятью (с триггером защелкой), дифференциальные, инвертирующие и неинвертирующие.

В зависимости от вида сигналов (напряжения U_c, U_s , токи I_c, I_s), подаваемых на компараторные и переключательные входы соответственно различают реле: потенциальные $RL(U_c, U_s)$, токовые $RL(I_c, I_s)$, потенциально-токовые $RL(U_c, I_s)$, токово-потенциальные $RL(I_c, U_s)$.

По числу переключательных каналов реле подразделяются на одноканальные (рис.1а, б), двухканальные (рис.1в) и много-канальные [а.с. СССР 1270777]. Зачернёнными кружками на рис.1 обозначены переключательные выводы размыкающих ключей реле, незачернёнными кружками - инвертирующие компараторные входы реле.

По признаку объединения выводов ключей

переключательных каналов различают реле: коммутационные (рис.1а, К-реле), мультиплексорные (рис.1б, в, М-реле) и демультиплексорные (рис.1г, Д-реле) [7].

По признаку наличия основных и дополнительных управляющих функций реле разделяются на четыре группы.

1) Реле с фиксированной воспроизводимой операцией (либо ИАВ-конъюнкция, либо ИАВ-дизъюнкция). Например, К-реле по рис.1а при объединённых переключательных выходах ($Z_1 = Z_2$) воспроизводится ИАВ-дизъюнкция (16).

2) Кодоуправляемые реле [10]. Здесь на управляющий вход реле подаётся логический сигнал $f \in \{0,1\}$, переключающий тип воспроизводимой операции: конъ-

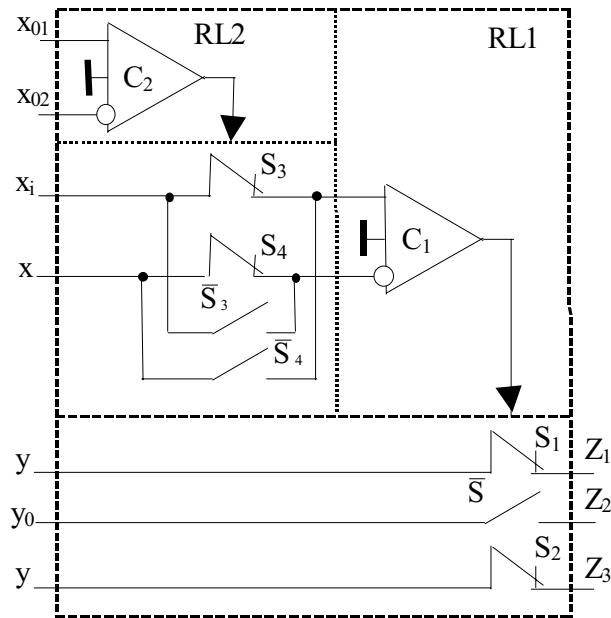


Рис.2. КД-реле

юнкцию (1а) на дизъюнкцию (1б) и обратно.

3) Конъюнктивно - дизъюнктивные (КД) реляторы (рис.2) [11, 12].

Здесь на управляющие (инверсные I) входы релятора RL подаются аналоговые сигналы (напряжения) x_{01} , x_{02} . Принципиальное отличие КД-реляторов от кодоуправляемых заключается в том, что управляющие напряжения x_{01} , x_{02} могут быть использованы в качестве информационных и (или) логических сигналов. Это обеспечивается совместимостью (совпадением) допустимых диапазонов изменения предикатных x_1 , x_2 и управляющих x_{01} , x_{02} сигналов. Указанное существенно расширяет их функциональные и управляющие возможности. Например, если $x_{01} = x_{01}(t)$ и $x_{02} = x_{02}(t)$ являются заданными функциями текущего времени, то переключение выходного процесса $Z = Z(t)$ осуществляется в моменты времени, когда $x_{01}(t)$ и $x_{02}(t)$ пересекаются. Тем самым обеспечивается КД - модуляция выходного процесса $Z(t)$. В частном случае при фиксированных значениях управляющих сигналов (например, при $x_{01}=0$ или $x_{02}=0$) КД-релятор вырождается в кодоуправляемый.

В КД-реляторе (рис.2) на компараторе C_1

и ключах S_1 , \bar{S}_1 и S_2 собран основной релятор RL1 с введённым в его состав дополнительным релятором RL2, содержащим управляющий компаратор и ключи S_3 , S_4 , \bar{S}_3 , \bar{S}_4 . При $x_{01} > x_{02}$ ключи S_3 , S_4 замкнуты, а \bar{S}_3 , \bar{S}_4 разомкнуты. При этом напряжения x_i и \bar{x} поступают соответственно на неинвертирующий и инвертирующий входы компаратора C_1 . При $x_{01} < x_{02}$ имеем обратную картину и напряжения x_i , \bar{x} поступают соответственно на инвертирующий и неинвертирующий входы компаратора C_1 . Тем самым обеспечивается инверсия воспроизведенной ИАП-операции на дуальную.

4) Реляторы с блокировкой переключательных каналов воспроизводят операции

$$Z = [y_1 I(x_1 - x_2) + y_2 I(x_2 - x_1)] \cdot J, \quad (4)$$

где x_1 , x_2 и y_1 , y_2 есть соответственно предикатные и предметные переменные (сигналы), $J=I(x_B)$ - управляющий логический сигнал блокировки, Z_1 , Z_2 - выходные выводы соот-

ветственно замыкающего S и размыкающего \bar{S} ключей переключательного канала.

Согласно (4), при $I(x_B)=1$ операции, воспроизводимые реляторами первой и данной групп, совпадают. Соответственно при $I(x_B)=0$ ключи переключательного канала разомкнуты (режим блокировки) и воспроизводится операция "запрет прохождения сигнала".

При объединении переключательных выходов Z_1 и Z_2 ($Z_1=Z_2=Z$) или входных переключательных входов релятора ($y_1=y_2=y$) приходит соответственно к мультиплексорному или демультиплексорному релятору. В частном случае для построения канала блокировки и коммутации ключей может быть использован мультиплексор KP1531КП7 [13].

5) Пятую группу составляют *реляторы с инверсией (I) и блокировкой (B)* сигналов. Здесь совмещаются операции, воспроизводимые второй, третьей и четвёртой групп: инверсия (взаимозамещение) типа воспроизводимых базовых операций (ИАВ-конъюнкция $\forall\forall$ ИАВ-дизъюнкция) на дуальные и операция блокировки ("Запрет прохождения коммутируемых сигналов").

На рис.3 представлены условные схемные изображения КД-реляторов (рис.3а), B-реляторов (рис.3б) и IB-реляторов (рис.3в).

Воспроизводимые мультиплексорным IB-релятором (выходы Z_1 и Z_2 на рис.3в объединены $Z_1=Z_2=Z$) операции определяются выражением:

$$Z = [y_1 I(\alpha_1 x_1 + \alpha_2 x_2) + y_2 I(\alpha_2 x_1 + \alpha_1 x_2)] I(X), \quad (5)$$

где весовые коэффициенты α_1 , $\alpha_2 \in \{-1, 1\}$, $I(X)=J$ есть единичная функция управляющая

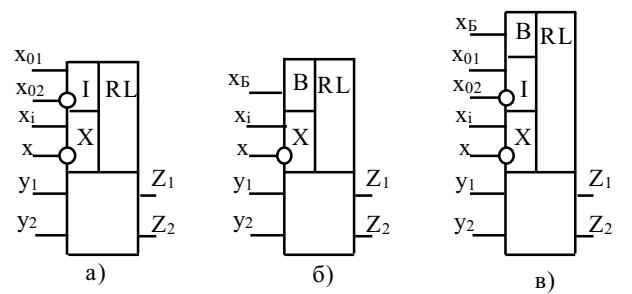


Рис.3. Условные схемы КД-реляторов

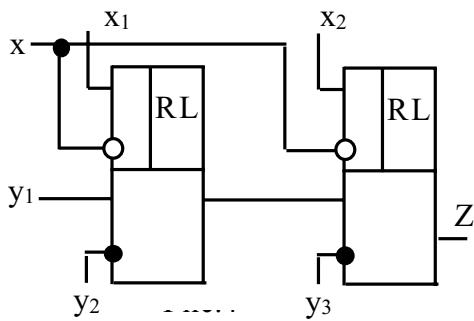


Рис.4. Схема двухпорогового релеятора

включением режима блокировки.

Согласно (5), при $a_1=-1$, $a_2=1$ и $a_3=1$, $a_4=-1$ ИВ-релеяторы при $I(X)=1$ воспроизводят соответственно ИПА - конъюнктивную и ИПА - дизъюнктивную операции.

По признаку числа используемых компараторов релеяторы подразделяются на однокомпараточные (однопороговые) и двухкомпараточные (двухпороговые ДР) [а.с. СССР 1381548, 1501098, 1513480, 1568060] [14, 15], воспроизводимые операции и функции которых имеют соответственно одну и две точки разрыва (излома).

В зависимости от выбранных логических законов управления ключами двухпороговые (двухкомпараточные) релеяторы подразделяются на двух-, трёх- и четырёхпозиционные[14].

В частности, на рис.4 представлена схема двухпорогового релеятора, содержащего два

однопороговых релеятора. В зависимости от заданной комбинации соединения выводов релеятора по рис.4 (коммутационное программирование) и от заданных уровней предметных переменных он воспроизводит ряд линейно-разрывных и линейно-изломных функций. На рис.5 представлены некоторые типовые функции, воспроизводимые релеятором по рис.4.

Релеятор по рис.5а воспроизводит функцию двухпорогового компарирования (функцию допускового контроля), релеяторы по рис.5б,в воспроизводят соответственно функции двустороннего и одностороннего параллельного ограничения входного сигнала x , релеяторы по рис.5г, д являются соответственно режекторным (рис.5г) и селекторным (рис.5д) фильтрами уровней сигнала x .

Релеяторы классифицируются также по признаку использования в различных предметных областях [16-29], что характеризует универсальность и многообразие применений логико-алгебраического (ЛА) аппарата импликативной алгебры выбора и, соответственно, релеяторов и релеяторной схемотехники.

Многообразие типов и номенклатуры не ограничивается потенциальными и токовыми релеяторами. Здесь не рассматривались оптоэлектронные, пневматические релеяторы, релеяторы с трансформаторно-токовыми ком-

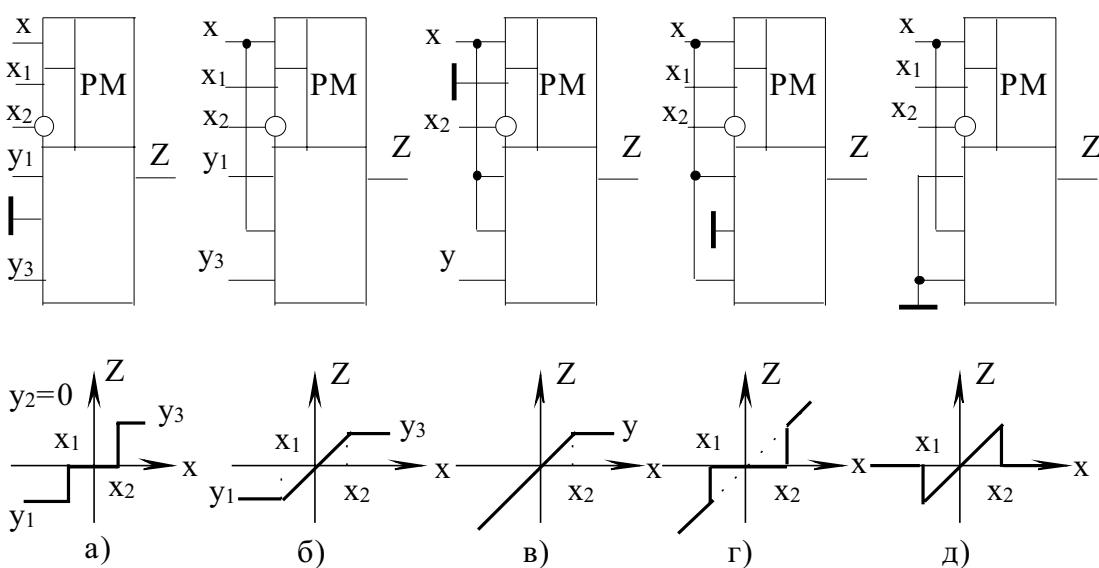


Рис.5. Типовые функции, воспроизводимые релеяторами

параторами, время-импульсные реляторы и др.

Работа выполнена при поддержке гранта Т00-3.3-2659 Минобразования РФ.

СПИСОК ЛИТЕРАТУРЫ

1. Волгин Л.И., Зарукин А.И. Развитие элементного базиса реляторной схемотехники // Датчики и системы. 2002. №3.
2. Гинзбург С.А. Математическая непрерывная логика и изображение функций. М.: Энергия, 1968.
3. Волгин Л.И., Левин В.И. Непрерывная логика. Теория и применения. Таллинн: АН Эстонии, 1990.
4. Волгин Л.И. Непрерывная логика и её схемотехнические применения. Ульяновск: УлГТУ, 1996.
5. Волгин Л.И. Комплементарная алгебра и предикатная алгебра выбора. Ульяновск: УлГТУ, 1996.
6. Волгин Л.И. Свойства и законы комплементарной алгебры // Известия АН ЭССР. сер. Физика, Математика. 1988. №4.
7. Волгин Л.И. Элементный базис реляторной схемотехники. Тольятти: Поволжский технологический институт сервиса, 1999.
8. Волгин Л.И. Метасистема взаимоотношений алгебраических логик и сопутствующих исчислений, порождаемых функцией-аксиомой взвешенных степенных средних // Информационные технологии. 2002. №7.
9. Волгин Л.И., Зарукин А.И. Устройства амплитудной свертки для АЦП // Цифровая информационно-измерительная техника. Пенза: Пенз. политехн. ин-т, 1985. Вып.15.
10. Андреев Д.В. Реляторные коммутационно-логические преобразователи и процессоры ранговой обработки аналоговых сигналов: Автoref. дис...канд. техн. наук / Научн. рук. Л.И. Волгин. Ульяновск: УлГТУ, 1998..
11. Патент РФ №2143730. Конъюнктивно-дизъюнктивный релятор / Л.И. Волгин. Заявлено 16.02.99 // Бюл. 1999. №36.
12. Волгин Л.И. Конъюнктивно-дизъюнктивный релятор // Реляторная схемотехника и средства преобразования аналоговой информации: Труды международной конф. "Методы и средства преобразования и обработка аналоговой информации". Ульяновск: УлГТУ, 1999. Т.2.
13. Волгин Л.И., Зарукин А.И, Ташиков Б.П. Потенциально-токовый релятор с произвольной полярностью коммутируемых токов // Реляторная схемотехника и средства преобразования аналоговой информации: Труды международной конф. "Методы и средства преобразования и обработка аналоговой информации". Ульяновск: УлГТУ, 1999.
14. Волгин Л.И. Синтез устройств для обработки и преобразования информации в элементном базисе реляторов. Таллин: Валгус, 1989.
15. Волгин Л.И. Предикатная алгебра выбора и ее модификации (основы теории и элементный базис) // Опыт, результаты, проблемы: Повышение конкурентоспособности радиоэлектронной аппаратуры. Таллин: Валгус, 1986. Вып.4.
16. Волгин Л.И., Климовский А.Б. Интеллектуализация измерений на основе графа Паскаля // Международная конференция по мягким вычислениям и измерениям. СПб.: СПбГЭТУ (ЛЭТИ), 1999. Т.2.
17. Волгин Л.И., Климовский А.Б. Интеллектуализация информационно-диагностирующих систем на основе предикатной алгебры выбора и графа Паскаля // Приборы и системы: Сб. материалов Всероссийской НТК "Приборы и приборные системы". Тула: ТулГУ, 2001.
18. Курносов В.Е., Курносова Т.В., Наумова И.Ю. Построение моделей конструкций на основе предикатной алгебры выбора и вариационной формулы МКЭ // Надежность и качество: Докл. Международного симпозиума, посвященного 275-летию РАН. Пенза: ПГУ, 1999.
19. Курносов В.Е. Теория и методы оптимального проектирования устройств радиотехники и связи на основе эволюционных дискретных моделей: Автoref дис... д-ра техн. наук. Пенза: ПГУ, 1999.
20. Курносов В.Е. Синтез формы теплонагру-

- женных элементов конструкций РЭС. Пенза: ПГУ, 2001.
21. Колосков В.А., Медведева М.В., Медведев А.В. Предикатная алгебра выбора в задаче построения клеточной самоорганизующей оболочки // Логико-алгебраические методы в науке, технике и экономике: Труды международной конф. "Континуальные логико-алгебраические и нейросетевые методы в науке, технике и экономике". Ульяновск: УлГТУ, 2000. Т.1.
22. Медведев А.В. Континуально-логические алгоритмы и устройства клеточной самоорганизации мульти микроконтроллера с программируемым резервом: Автореф дис... канд.техн.наук / Научн. руководитель Ф.А.Старков. Курск: КГТУ, 2000.
23. Кувшинов А.А. Синтез однофазных конверторов в элементном базисе реляторов на основе логико-топологических моделей // Проблемы и решения современной технологии: Сб. научн. трудов Поволжского технологического института сервиса. Тольятти: ПТИС, 2000. Вып.8.
24. Абрамов Г.Н., Кувшинов А.А.. Силовой релятор на гибридных интеллектуальных модулях // Проблемы и решения современной технологии: Сб. научн. трудов Поволжского технологического института сервиса. Тольятти: ПТИС, 2000. Вып.9.
25. Кувшинов А.А., Абрамов Г.Н. Предикатная алгебра выбора в задачах силовой электроники // Международная конф. по мягким вычислениям и измерениям. СПб.: СПбГЭТУ (ЛЭТИ), 2001. Т.2.
26. Кувшинов А.А. Системный подход к синтезу логико-алгебраических моделей и структур в задачах силовой электроники // Научно-технический калейдоскоп. Серия "Приборостроение, радиотехника и информационные технологии" / Под ред. Л.И. Волгина, 2001. №2.
27. Дёмин А.А., Масленников В.В., Мефеденко М.В. Возможности реализации реляторных структур на новом аналоговом базовом матричном кристалле // Научно-технический калейдоскоп. Серия "Приборостроение, радиотехника и информационные технологии" / Под ред. Л.И. Волгина. 2001. №2.
28. Волгин Л.И. Реляторные процессоры на основе графа Паскаля для адресно-ранговой идентификации, селекции и ранжирования аналоговых сигналов. Тольятти: Поволжский технологический институт сервиса, 2000.
29. Волгин Л.И. Релятор и реляторная схемотехника: Логико-алгебраические основы и применения. Библиографический указатель и комментарии. Тольятти: ПТИС, 2002.

CATEGORIZATION OF RELATORS ON DOMINANT SIGNS

© 2002 L.I. Volgin¹, A.I. Zarukin², A.B. Climovsky¹

¹Ulyanovsk State Technical University

²Ulyanovsk Branch of Military Academy of the Rear Services and Transport

In the paper is brought the categorization of relators, which are the universal logical circuit elements, reproducing binary operations of the set of algebraic logics (the special algebras and mathematical logics) in continual, discrete (including two-valued boolean algebra of the logic) areas.

УДК 535.6 62.50

ПРИМЕНЕНИЕ ПРИНЦИПА СОГЛАСОВАННОСТИ ОЦЕНОК В ЗАДАЧЕ ИДЕНТИФИКАЦИИ МОДЕЛЕЙ ЦВЕТОВОСПРОИЗВЕДЕНИЯ

© 2002 А.В. Никоноров¹, С.Б. Попов², В.А. Фурсов²

¹ Самарский государственный аэрокосмический университет

² Институт систем обработки изображений РАН, г. Самара

Решается задача идентификации параметров модели цветовоспроизведения. Метод опирается на так называемый принцип согласованности оценок, позволяющий строить оценки в условиях априорной неопределенности, связанной с тем, что число наблюдений на исходных образцах изображений мало. Разрабатываемый метод может применяться как для формирования изображений наиболее точно воспроизводящих цвет при офсетной печати, так и для создания электронных коллекций изображений, интернет-ресурсов и др.

Постановка задачи

Рассматривается задача построения модели цветовоспроизведения, позволяющий на каждом этапе допечатной подготовки адекватно отображать цветовой состав изображения. В пространстве спектральных коэффициентов отражения печатных оттисков (в дальнейшем называемых спектрами) эта задача формулируется следующим образом. Пусть задана модель, характеризующая зависимость спектра смеси красок r от вектора искомых концентраций $c = (c_i)$, $i = \overline{1, n}$ красок, входящих в смесь:

$$\mathbf{r} = F(\alpha, \mathbf{r}_p, \mathbf{r}_{pi}), \quad (1)$$

где R_p – спектр отражения печатной основы, R_{pi} спектр стопроцентной концентрации i -той краски, входящей в смесь, лежащей поверх основы.

Введем в рассмотрение так называемое цветовое пространство Lab, координаты которого определяются соотношениями [1, 2]:

$$L = 116f(Y/Y_N) - 16, \quad (2)$$

$$a = 500[(f(X/X_N) - f(Y/Y_N)], \quad (3)$$

$$b = 200[f(Y/Y_N) - f(Z/Z_N)], \quad (4)$$

где $X_N = 96,422$; $Y_N = 100$; $Z_N = 82,521$,

$$f(t) = t^{1/3} \text{ при } t < 0,009 \text{ и}$$

$$f(t) = 7,7867t + 16/116 \text{ при } t \geq 0,009. \quad (5)$$

Координаты цвета в пространстве XYZ определяются через спектр отражения образца $R(\lambda)$ как:

$$\begin{aligned} X &= k \int_{\Lambda} S(\lambda)x(\lambda)R(\lambda)d\lambda, \\ Y &= k \int_{\Lambda} S(\lambda)y(\lambda)R(\lambda)d\lambda, \\ Z &= k \int_{\Lambda} S(\lambda)z(\lambda)R(\lambda)d\lambda \end{aligned} \quad (6)$$

где $k = 100 / \int_{\Lambda} S(\lambda)y(\lambda)d\lambda$, $S(\lambda)$ – спектральный состав излучения от источника освещения; из некоторого набора стандартизованных спектров для типичных источников, а $x(\lambda)$, $y(\lambda)$ и $z(\lambda)$ – так называемые кривые сложения, характеризующие чувствительность глазных рецепторов человека.

Количественной мерой различия между цветами является цветовой контраст. Он определяется как расстояние между точками, в цветовом пространстве Lab [1, 2]:

$$\Delta E = \sqrt{(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2}, \quad (7)$$

где $\Delta L = L_i - L_j$, $\Delta a = a_i - a_j$, $\Delta b = b_i - b_j$, а $L_i, a_i, b_i, L_j, a_j, b_j$ – координаты i -го и j -го цвета соответственно. Единичное расстояние в пространстве Lab совпадает с порогом цветоразличения человеческого зрения.

Качество модели (1) характеризуется величиной цветового контраста ΔE между реальным цветом красочной смеси и цветом, оцененным с использованием модели. Зада-

ча заключается в нахождении вида и параметров модели, обеспечивающих минимальную разницу в пространстве Lab между реальным и рассчитываемым значением цвета красочной смеси. Для этого решается задача минимизации функционала, полученного на основе соотношения (7):

$$\min_{\tilde{\theta}} \sqrt{((L - \tilde{L}(\tilde{\mathbf{n}}))^2 + (a - \tilde{a}(\tilde{\mathbf{n}}))^2 + (b - \tilde{b}(\tilde{\mathbf{n}}))^2}. \quad (8)$$

Здесь (L, a, b) – координаты цвета, полученные на обучающей выборке, $(\tilde{L}, \tilde{a}, \tilde{b})$ – рассчитанные по формулам (2-6) с использованием модели цветовоспроизведения (1).

Проблема заключается в том, что задача оценки векторного параметра \mathbf{c} должна решаться по малому числу обучающих образцов. Объективной причиной этого является не только высокая стоимость цветных тестовых изображений (хотя и это существенно), но, главным образом, то, что не для всех базовых красок существуют тестовые шкалы для измерения спектров. Традиционно задача оценивания параметров решается в рамках статистической теории. При малом числе наблюдений использование стандартных априорных вероятностных гипотез не вполне правомерно.

В настоящей работе рассматривается общая схема и пример решения сформулированной задачи на основе так называемого принципа согласованности оценок. Используемые в рамках этого принципа критерии не требуют задания априорных вероятностных моделей ошибок измерений.

Общая схема решения задачи на основе принципа согласованности

Для описания зависимости типа (1) широко используется линейная модель вида

$$\mathbf{r} = \mathbf{X}\mathbf{c} + \xi, \quad (9)$$

где $\mathbf{X} = [\mathbf{r}_{p_1}^T, \mathbf{r}_{p_2}^T, \mathbf{r}_{p_3}^T, \dots]^T$ – $N \times M$ матрица, компоненты которой суть измеренные спектры заданных концентраций базовых цветов, а $\mathbf{c} = [c_0, c_p, c_s, c_{12}, \dots]$ – искомый $N \times 1$ -вектор параметров модели.

Задача заключается в построения оцен-

ки $\hat{\mathbf{c}}$ вектора параметров \mathbf{c} уравнения (9) по доступной для непосредственного наблюдения $N \times M$ - матрице \mathbf{X} и $N \times 1$ - вектору \mathbf{r} ($N > M$), при неизвестном $N \times 1$ -векторе ошибок измерений ξ .

Предполагается, что соответствующая уравнению (10) точная модель оцениваемой системы:

$$\mathbf{r}^* = \mathbf{X}\hat{\mathbf{c}} \quad (10)$$

где $\mathbf{r}^* = \mathbf{r} - \xi$ существует. Ясно, что равенство (10) выполняется также для всех подсистем меньшей размерности, сформированных из строк системы (9). Опираясь на это свойство, простейшая схема отыскания оценки $\hat{\mathbf{c}}$ вектора параметров \mathbf{c} уравнения (9) на основе принципа согласованности строится следующим образом.

С использованием строк системы уравнений (9) строятся переопределенные системы меньшей, чем исходное матричное уравнение (10) размерности k . Максимальное число таких систем будет равно числу сочетаний из N по k , где $M < k < N$ – заданная (для простоты фиксированная) размерность формируемых вариантов систем. Далее на каждом из построенных таким образом вариантов формируется множество подсистем размерности $M \leq q < k$, наибольшее число которых также равно числу сочетаний из k по q .

Пусть $\hat{\mathbf{c}}_{k,q}$ – оценка, полученная для Q -й подсистемы K -го варианта, а $W_k[\hat{\mathbf{c}}_{k,q}]$ – функция, характеризующая взаимную близость решений $\hat{\mathbf{c}}_{k,q}$, полученных для различных подсистем на K -м варианте. Задается также критерий отбора наиболее подходящих вариантов по значениям полученных функций близости $W_k[\hat{\mathbf{c}}_{k,q}]$. Искомое решение задачи

является точечной оценкой $\hat{\mathbf{c}}$, вычисленной в соответствии с заданным критерием на вариантах, для которых функция взаимной близости удовлетворяет критерию отбора.

В рамках принципа согласованности могут быть заданы различные функции взаимной близости, критерии отбора вариантов

(по вычисленным значениям функций взаимной близости), а также критерии построения точечных оценок на отобранных вариантах. В частности, в работе [6] рассматривалась функция взаимной близости решений \mathbf{c} на множестве подсистем вида:

$$W_k[\hat{\mathbf{c}}_{k,q}] = \sum_{q=1}^Q \left| \hat{c}_{q,i} - \bar{c}_i \right|^2, \quad (11)$$

где $\bar{c}_i = \frac{1}{Q} \sum_{q=1}^Q \hat{c}_{q,i}$ (12)

- i -я компонента вектора $\bar{\mathbf{c}}$, вычисленного путем усреднения оценок $\{\hat{\mathbf{c}}_{k,q}\}$, вычисленных на вариантах, для которых значения функции взаимной близости $W_k[\hat{\mathbf{c}}_{k,q}]$ оказались менее заданного порога. а $\hat{c}_{k,q}$ - i -е компоненты вектора оценок $\hat{\mathbf{c}}$, полученные в результате решения q -й подсистемы k -го варианта. При этом искомая точечная оценка определяется путем усреднения оценок полученных на этом варианте.

Данный критерий интегральный, т.е. его значение в некоторой точке зависит от значения во всех остальных точках. Таким образом, при переборе, с появлением нового варианта должен происходить пересчет значений критерия для всех предыдущих вариантов. На практике это приведет к значительным затратам памяти и увеличению вычислительной сложности.

Точечные критерии вроде (8) лишены этого недостатка. Однако при использовании (8) вычисляется не близость значений оценок, а ошибка аппроксимации целевой функции. Расчет ошибки аппроксимации невозможен при экстраполяции, что является существенным недостатком критерия (8).

Критерий (11) представляет собой отклонение оценки $\hat{\mathbf{c}}_{k,q}$ от среднего значения, полученное по выборке. Оптимальное значение оценки по этому критерию можно заменить на точечную оценку выборочного среднего \bar{c}_i . Возможно рекуррентное вычисление выборочного среднего как:

$$\bar{c}_i = \frac{(i-1)\bar{c}_{i-1} + \hat{c}_{k,q}}{i}.$$

Рассчитанное таким образом для всех переопределенных систем среднее принимается за оценку параметров системы (10). В предположении непрерывности изменения ошибки аппроксимации такая оценка должна быть оптимальной по критерию (11).

Другой способ использования приближения критерия (11) это расчет среднего и пересчет в связи с этим значения $W_k[\hat{\mathbf{c}}_{k,q}]$ не для всех $\hat{\mathbf{c}}_{k,q}$, а для некоторого набора значений, на предыдущих итерациях показавших лучшее значение критерия. Такое приближение также основано на требовании непрерывности. Это приближение может использоваться при генетических алгоритмах поиска переопределенной системы дающей оптимальное значение оценки.

Пример решения задачи

В качестве примера решалась задача идентификации модели для двух базовых красок (голубой и пурпурной). Строилась линейная модель для зависимости спектра отражения этих двух базовых красок взятых в некотором процентном соотношении, от спектров печатной основы (бумаги) и спектров соответствующих концентраций базовых красок. В частности, исследовался случай решения для 30% голубой и 20% пурпурной краски.

Для этого случая модель (9) принимает вид:

$$\mathbf{r}_k = \mathbf{c}_0 \mathbf{r}_p(k) + \mathbf{c}_1 \mathbf{r}'_{p1}(k) + \mathbf{c}_2 \mathbf{r}'_{p2}(k) + \mathbf{c}_{12} \mathbf{r}_{p12}(k) + \xi, \quad (13)$$

где $k = 1, N$; \mathbf{r}_p - спектр бумаги; \mathbf{r}'_{p1} и \mathbf{r}'_{p2} - спектры соответственно голубой и пурпурной краски поверх бумаги. Для линейной модели (13) матрица \mathbf{X} формируется из заданных спектров \mathbf{r}_p , \mathbf{r}'_{p1} , \mathbf{r}'_{p2} и \mathbf{r}_{p12} , образующих вектор-столбцы. Причем \mathbf{r}'_{p1} и \mathbf{r}'_{p2} являются спектрами не стопроцентных концентраций базовых красок, а тех концентраций, которые входят в смесь. Это связано с тем, что на практике задача моделирования спектра произвольной красочной смеси представляется в

виде суперпозиции двух задач: получение спектра необходимой концентрации базовых красок, и нахождения спектра смеси. Вторая задача фактически является обобщением первой, поэтому предполагаем первую задачу решенной и спектры \mathbf{r}'_{p_1} и \mathbf{r}'_{p_2} известными.

Спектры отражения образцов измерялись в диапазоне 380-700 нм с шагом 10 нм, так, что $N = 32$, $M = 4$. Измерения проводились по шкале цветового охвата напечатанной на цветопробном принтере Scitex Iris 4Print с использованием спектрофотометра Techcon SP810, повторяемость прибора $0,02\Delta E$. Lab координаты цвета результирующей смеси следующие: (74, 1, 15). Результаты оценивания приведены в табл.1 и на рисунке.

Для МНК-оценок параметров модели ошибка аппроксимации составила $1,2844 \Delta E$. В рамках принципа согласованности, используя цветовой контраст (8) в качестве функции взаимной близости, удалось получить ошибку аппроксимации в $0,1132 \Delta E$. При этом перебирались все варианты формирования переопределенных систем меньшей размерности. Наибольшая точность была достигнута на варианте размерности 8, в котором переопределенная система была составлена из строк (2, 6, 8, 14, 17, 20, 24, 27) исходной матрицы. Для других размерностей ошибка увеличивается, что иллюстрируется на графике.

Для нескольких других спектров из шкалы цветового охвата для двух красок сравнительные данные точности МНК-оценок и оценок, полученных на основе принципа согласованности, приведены в таблице 1. В графике N_{opt} приведена размерность варианта переопределенной системы, для которой достигалась наименьшая ошибка аппроксима-

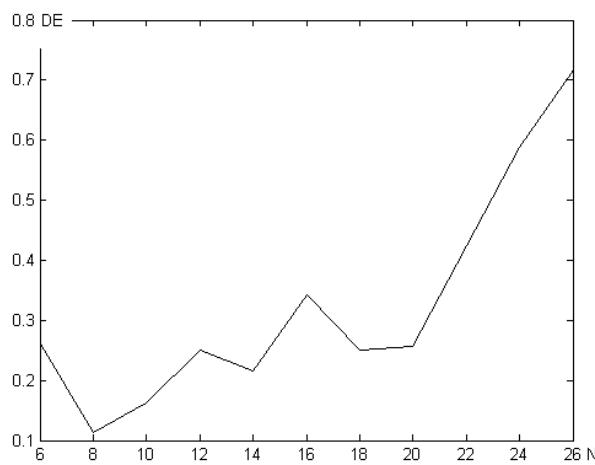


Рис. Зависимость ошибок аппроксимации от размерности варианта системы

ции согласно критерию минимума цветового контраста.

Вычислительные аспекты реализации

Рассмотрим вопрос реализации оценок с использованием принципа согласованности. Это задача полного перебора. Если необходимо решить задачу для полного набора подсистем (с размерностями от 1 по N), то вычислительная сложность зависит от N как $O(2^N)$. Однако сложность можно существенно сократить, если задан некоторый интервал изменения размерности подсистем. В частности для одной размерности $k < N$ необходимо перебрать $I = C_N^k$ вариантов.

Сформулируем задачу перебора. Переопределенные системы, размерности k , построенные на основании системы (9) можно представить в виде двоичного вектора \mathbf{b} размерности $I \times N$. Единица в l -том разряде вектора \mathbf{b} означает наличие l -той строки исходной системы в переопределенной системе, ноль – отсутствие. Норма Хэмминга вектора \mathbf{b} равна размерности переопределенной системы. Таким образом, перебирая вектора \mathbf{b}_i с нормой равной k мы перебираем все переопределенные системы размерности k .

Перебор \mathbf{b}_i в порядке возрастания в двоичном коде или коде Грэя не может обеспечить постоянство нормы Хэмминга для всех вариантов. Это приводит к появлению большого количества "холостых" итераций перебора.

Таблица 1. Ошибки аппроксимации

№	Конц. голубой краски	Конц. пурпурной краски	Ошибка (ΔE) МНК-оценок	Ошибка (ΔE) согласованных оценок	N_{opt}
1	30%	20%	1,2844	0,1132	8
2	60%	20%	1,1955	0,1228	11
3	70%	90%	4,4230	0,2988	9
4	10%	40%	0,4305	0,1266	10
5	50%	50%	2,3162	0,1495	11

Осуществить перебор \mathbf{b}_i , где i изменяется от 0 до $I = C_N^k - 1$, с сохранением нормы Хэмминга позволяет алгоритм, состоящий из следующих шагов.

1. Пусть стартовым значением процедуры перебора будет $\mathbf{b}_0 = \underbrace{0\dots 0}_{N-k} \underbrace{1\dots 1}_k$.

2. На i -ом шаге выполняется следующая процедура. Пусть l наименьший разряд, для которого $\mathbf{b}_{i-1}^{l-1} = 1$ и $\mathbf{b}_{i-1}^l = 0$. Тогда $\mathbf{b}_i^j = \mathbf{b}_{i-1}^j$ при $j > l$, $\mathbf{b}_i^l = 1$ и $\mathbf{b}_i^{l-1} = 0$. Если $\mathbf{b}_{i-1}^0 = 0$ и $l > 1$, тогда $\mathbf{b}_i^{l-1-j} = \mathbf{b}_{i-1}^j$ при $j < l - 1$.

Если необходимого l не найдено, значит, перебор завершен и получен вектор

$$\mathbf{b}_I = \underbrace{1\dots 1}_{k} \underbrace{0\dots 0}_{N-k}.$$

3. Если не получено \mathbf{b}_I переход на шаг 2. В табл.2 приведена последовательность \mathbf{b}_i для $N = 5$ и $k=3$, $I = C_6^3 - 1 = 19$.

Задача оценивания выполняется для каждой переопределенной системы, определяемой вектором \mathbf{b}_i . Оценивание для нескольких значений размерности переопределенных систем, можно выполнить для каждой размерности при помощи описанной процедуры. Назовем такой алгоритм поиска согласованных оценок алгоритмом с последовательным увеличением размерности.

Рассмотрим изменение вычислительной сложности с увеличением размерности на 1. Вектор оценок параметров модели (9) находится по МНК.

$$\hat{\mathbf{c}} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{r}. \quad (14)$$

Вычислительная сложность формирования информационной матрицы пропорциональна k^2 , сложность обращения матрицы методом Гаусса равна [8]

Таблица 2. Формирование двоичного представления переопределенных систем

i	b_i	i	b_i	i	b_i	i	b_i
0	000111	5	010101	10	100011	15	101100
1	001011	6	010110	11	100101	16	110001
2	001101	7	011001	12	100110	17	110010
3	001110	8	011010	13	101001	18	110100
4	010011	9	011100	14	101010	19	111000

$$\frac{2k(k+1)(k+2)}{3} + k(k-1).$$

Тогда вычислительную сложность операции (14) можно оценить как

$$G_k = \frac{2k(k+1)(k+2)}{3} + 3k^2.$$

При увеличении k на 1 получим

$$\Delta G = G_{k+1} - G_k = 2(k+3)^2 - 11. \quad (15)$$

Число операций (14) с увеличением k на 1 увеличивается в $\frac{N-k}{k+1}$ раз. С учетом (15) получим квадратичное увеличение сложности.

Параллельная обработка может выполняться так, что все процессоры решают задачи одной размерности, при этом вычислительные затраты на решение каждой задачи будут одинаковыми. Общее количество задач размерности k , равное числу сочетаний из N по k , можно разделить между всеми процессорами кластера поровну. Выдачу заданий каждому процессору в виде набора векторов \mathbf{b}_i должен проводить процессор-планировщик. При гетерогенной структуре кластера объем всех наборов одинаков. Иначе на некоторой размерности проводится калибровка скорости вычислений на каждом процессоре, в дальнейшем размер наборов должен быть скорректирован обратно пропорционально скорости вычислений.

Кроме вычислений на кластере описание переопределенных систем в виде двоичных векторов позволяет проводить численные эксперименты с использованием генетических алгоритмов перебора. Такой подход в комбинации с ограничением размерности позволяет получать неплохое приближение оптимального решения на

персональном компьютере.

Заключение

Применение принципа согласованности оценок позволяет уменьшить ошибку аппроксимации по сравнению с МНК-оценками более чем на порядок. Однако достигается это, кроме прочего, за счет существенного возрастания вычислительных затрат. Добиться приемлемого времени счета удается за счет применения высокопроизводительных системы обработки информации, например, кластеров или суперкомпьютеров.

Работа выполнена при частичной поддержке РФФИ (гранты № 01-01-00097, 00-01-05001).

СПИСОК ЛИТЕРАТУРЫ

1. Джадд Д., Вышецки Г. Цвет в науке и технике. М.: Мир, 1978.
2. Шашлов Б. А. Цвет и цветовоспроизведение. М.: Мир книги, 1995.
3. Berns S.R. The Spectral Modeling of Large-Format Ink-Jet Printers. Barselona: RIT Munsell Color Science Laboratory. 1996.
4. Stollnitz E. J. Reproducing Color Images Using Custom Inks. University of Washington.
5. Фурсов В.А. Идентификация моделей систем формирования изображений по малому числу наблюдений. Самара: СГАУ, 1998.
6. Фурсов В.А. Проблемы вычисления оценок по малому числу наблюдений // Труды молодежной школы "Математическое моделирование 2001". Самара, 2001.
7. Попов С.Б., Никоноров А.В. Сравнительный анализ моделей цветообразования при офсетной многокрасочной печати. Компьютерная оптика. 23. 2002.
8. Демидович Б.П., Марон И.А. Основы вычислительной математики. М.: Наука, 1979.

THE CALCULATING APPEARANCE OF COLOR REPRODUCTION MODEL IDENTIFICATION

© 2002 A.V. Nikonorov¹, S.B. Popov², V.A. Fursov²

¹Samara State Aerospace University

²Image Processing Systems Institute of Russian Academy of Sciences, Samara

Problem of color reproduction model parameters identification is resolved. The method is based on so-called conformity estimator principle by which estimations are made in conditions of expected uncertainty. This uncertainty arises from small number of original pattern observation. Developed method could be used for most accurately color matching during offset printing as well as creation of electronic picture collections, Internet resource and others.

УДК 543.226;621.391.7

ДВУХКАНАЛЬНАЯ СИСТЕМА РЕГИСТРАЦИИ И ОБРАБОТКИ ДАННЫХ ТЕРМОАНАЛИТИЧЕСКОГО ЭКСПЕРИМЕНТА

© 2002 Ю.В. Мошенский

Самарский государственный технический университет

Предложена информационно-измерительная система термического анализа, включающая в себя дифференциальный сканирующий микрокалориметр ДСК-500 и ЭВМ типа IBM PC. Данные от микрокалориметра к компьютеру передаются по двум информационным каналам – "температура" и "дифференциальная температура". В качестве устройства сопряжения применён интерфейс, использующий LPT порт. Обмен данными организован в режиме реального времени с возможностью считывания данных в "фоновом" режиме. Скорость обмена данными – до 25кб/сек.

Широкое распространение методов термического анализа (ТА), в частности, дифференциального термического анализа (ДТА) и дифференциальной сканирующей калориметрии (ДСК) в различных отраслях промышленности и многих отраслях естественнонаучных исследований обусловлено прежде всего информативностью методов ТА и их экспрессностью. Любые явления, которые сопровождаются изменением энталпии или теплоёмкости можно обнаружить методами ДТА или ДСК при условии достаточной чувствительности прибора, а указанные явления вызываются фундаментальными изменениями в физическом состоянии, химическом составе, реакционной способности молекул вещества и т.д. По положению пиков на шкале температур, их числу и форме производится качественная идентификация вещества, а площади пиков, пропорциональные энталпии реакции, используются для количественной оценки, например, прореагировавших веществ или для термохимических определений. Таким образом, термические методы анализа открывают новые возможности для решения существующих в химии проблем и позволяют ставить новые задачи, поэтому "современная химическая лаборатория немыслима без термовесов или прибора для дифференциального термического анализа" [1]. На химико-технологическом факультете СамГТУ разработан, изготовлен и испытан опытный образец системы регистрации и обработки данных термоаналитического эк-

сперимента. Функциональная схема системы приведена на рис.1. В её состав входит дифференциальный сканирующий микрокалориметр ДСК-500 с интерфейсом и ЭВМ типа IBM PC с соответствующим программным обеспечением [2]. Микрокалориметр обладает высокой (порог обнаружения по мощности 10 мкВт) чувствительностью и точностью, удовлетворяющими требованиям фармакопейного анализа и позволяет проводить термический анализ широкого класса веществ в диапазоне температур от -150 до 500°C [3]. Система призвана обеспечить пять основных типов задач, наиболее часто встречающихся в практике термического анализа [4].

1. Текущие вычисления и составление таблиц.

2. Численная обработка, включающая дифференцирование, интегрирование, сглаживание экспериментальных зависимостей.

3. Выбор подходящих функциональных соотношений, аппроксимирующих частично или полностью экспериментальную зависимость и оценка параметров этих соотношений. Сюда же относится решение обратной кинетической задачи, т.е. нахождение некоторой функции, аппроксимирующей экспериментальные данные и соответствующей механизму процесса.

4. Анализ экспериментальных данных для получения и интерпретации информации о состоянии объекта (нахождение сингулярных точек на ТА-кривой).

5. Статистический анализ результатов.

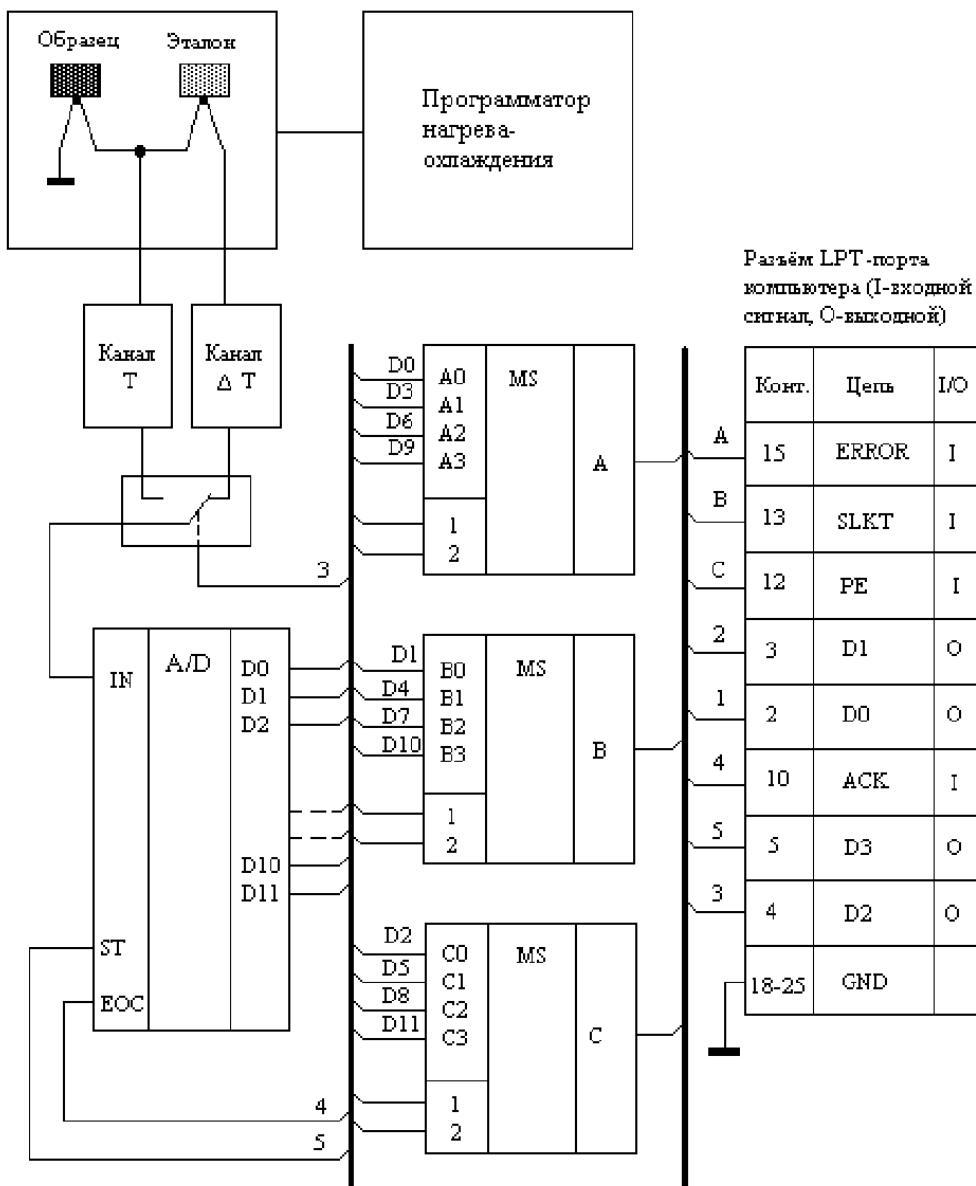


Рис.1. Функциональная схема системы

Для успешного решения этих задач было разработано специализированное устройство сопряжения, позволяющее организовать ввод данных от ДСК по двум информационным каналам – "температура" и "дифференциальная температура". Из входящих в базовую конфигурацию компьютера трёх типов стандартных внешних интерфейсов выбран наиболее простой по сопряжению и удобству программирования параллельный интерфейс Centronics, поскольку жёстких требований по скорости информационного обмена и длине линии связи не предъявляется. Интерфейс Centronics и, соответственно, LPT – порт персонального компьютера асимметричен, 12

линий (8-битная шина данных, 4-битная шина управляющих сигналов) работают на вывод, а 5 линий -на ввод, поэтому для ввода данных была использована пятиразрядная шина состояния. Длина одного передаваемого слова 12 бит (в ДСК применён 12-разрядный АЦП), поэтому слово разбивается на 4 части и принимается через 3 управляющие линии: -"ERROR", "SLCT", "PE" [5]. Процесс приёма слова данных происходит в следующем порядке:

- компьютер выставляет запрос на приём слова, тип данных (значение абсолютной температуры или дифференциальной температуры) и АЦП калориметра начинает цикл

преобразования;

- по завершении цикла преобразования на щине "ACK" появляется сигнал готовности прибора;
- начинается приём слова четырьмя пакетами по 3 бита. Номер пакета принудительно выставляется на шинах данных "D0-D1".

Скорость обмена данными по описанному алгоритму выбрана равной 25 кб/сек.

Для термоаналитического эксперимента большое значение имеет точность измерения площади пика дифференциальной кривой, поэтому необходима точная синхронизация по времени данных, поступающих в компьютер. Поскольку точность работы стандартного таймера Microsoft Windows невысока, для инициализации запросов использованы системные часы реального времени, доступ к которым предоставляет функция Windows API GET Tick Count, позволяющая вести отсчёт времени в миллисекундах с высокой точностью. Весь процесс опроса часов реального времени и считывания данных вынесен в отдельный поток операционной системы (thread), что позволило добиться удовлетворительной работы программы в "фоновом" режиме. Программа выполнена с использованием многодокументного интерфейса Windows, поэтому во время сканирования данных термоаналитического эксперимента (как правило – достаточно длительного) процессор может быть загружен дополнительно. Если в какой-то момент времени ресурсов процессора не хватит для обработки очередной порции данных, пропущенные данные будут либо интерполироваться на основе предыдущих данных по линейному алгоритму, либо заменяться предыдущей порцией данных.

Программа позволяет производить:

- одновременное считывание и отображение данных о тепловых превращениях веществ в двух системах координат: "время – дифференциальная температура" и "температура – дифференциальная температура";
- выбор единиц измерения для осей координат (отсчёты АЦП или градусы для шкал температур, отсчёты АЦП или секунды, минуты, часы для шкалы времени);

- сохранение результатов эксперимента во внутреннем формате программы, в стандартном формате файлов текстовых данных CSV или в графическом формате BMP;

- вычисление площадей термических пиков в абстрактных единицах площади и в Дюоулях;

- сглаживание кривых (цифровая фильтрация шумов);
- коррекция базовой линии;
- настройка базовой линии;
- калибровка прибора при помощи настраиваемых калибровочных таблиц.

Система была применена для анализа состава лекарственных препаратов, который в настоящее время проводится в основном объёмными методами химического анализа. Из физико-химических методов наиболее широко используются хроматографические и спектрографические методы, обладающие известными недостатками. К достоинствам термоаналитического метода относится простота проведения эксперимента (не требуется приготовления никаких дополнительных стандартов, растворов сравнения и т.д.), возможность формализации процесса исследования (что снижает уровень требований к квалификации аналитика), а также существенное сокращение времени анализа. К недостаткам метода можно отнести снижение точности определений при содержании лекарственного вещества в твёрдой форме менее 0,01 грамма и применимость метода только к веществам, не разлагающимся до температуры плавления. С использованием системы проводились эксперименты по идентификации индивидуальных лекарственных веществ и следующих фармпрепаратов: кофеин, парацетамол, пирацетам, анестезин и лекарственные формы – баралгин. Оптимальные условия эксперимента для каждого стандартного образца определялись экспериментально и воспроизводились при контролльном исследовании соответствующего лекарственного средства. Была подтверждена высокая эффективность использования термоаналитических методов как для анализа состава лекарственных препаратов, так и для определения чистоты индивидуальных лекар-

ственных веществ по величине понижения температуры плавления с точностью до 0,1% при содержании основного вещества не менее 98% [6]. Таким образом, в результате проделанной работы по разработке и изготовлению отечественной информационно-измерительной системы термического анализа открываются широкие перспективы использования сканирующей калориметрии для целей стандартизации лекарственных препаратов.

Работа выполнена при поддержке гранта РФФИ №00-04-68607.

СПИСОК ЛИТЕРАТУРЫ

1. Уэндландт У. Термические методы анализа. М.: Мир, 1978.
2. Федотов С.В. Интерфейсное программное обеспечение DSC Tool. Руководство пользователя. Самара: Самарский государственный технический университет, 2001.
3. Мощенский Ю.В., Трунин А.С., Космынин А.С. Система термического анализа для калориметрических исследований. Самара: Самарский государственный технический университет, 1999.
4. Шестак Я. Теория термического анализа. М.:Мир, 1987.
5. Гук М. Интерфейсы ПК: справочник. СПб.: Питер, 1999.
6. Жнякина Л.Е., Ткаченко М.Л., Космынин А.С., Мощенский Ю.В. Кинетические особенности растворения в системе парацетамол – мочевина // Химико-фармацевтический журнал. 2001. Т.35. №12.

TWO-CHANNEL SYSTEM OF REGISTRATION AND DATA PROCESSING OF TERMAL ANALYSIS EXPERIMENT

© 2002 Yu.V. Moshchensky

Samara State Technical University

The informative measuring system of the thermal analysis including differential scanning microcalorimeter DSC-500 and the computer such as IBM PC is offered. The data from a microcalorimeter to the computer are transferred on two information channels - "temperature" and "differential temperature". As the device of interface the interface using LPT port is applied. Data exchange is organized in a mode of real time with an opportunity of data read-out in a "background" mode. Speed of data exchange - up to 25 kb/sek.

УДК 62-50

МЕТОДИКА КОНТРОЛЯ ТОЧНОСТИ И ДОСТОВЕРНОСТИ РАСЧЕТНЫХ ДАННЫХ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ МОНИТОРИНГА И АВТОМАТИЗИРОВАННОГО УПРАВЛЕНИЯ

© 2002 Л.А. Шерешевский, А.А. Сидоров

Научно-внедренческая фирма "Сенсоры, модули, системы", г. Самара

Представлена методика анализа точности и достоверности результатов вычислений в распределенных системах мониторинга и автоматизированного управления. В основу методики положена модель, базирующаяся на принципах потоковых вычислений и вычислений с автоВалидацией. Описана программная реализация методики и рассмотрены результаты ее практического применения.

Введение

Точность и достоверность данных в автоматизированных комплексах управления технологическими процессами являются критическими факторами, во многом определяющими безопасность и эффективность функционирования системы в целом. Наряду с измеряемыми, в автоматизированных системах управления технологическими процессами (АСУ ТП) все шире используются вычисляемые параметры, описывающие недоступные для прямых измерений характеристики состояния объекта автоматизации. В отличие от непосредственно измеряемых величин, погрешность которых однозначно определяется суммарной погрешностью измерительного тракта, определение точности расчетных параметров на основании имеющихся данных о точности используемых средств измерений представляет нетривиальную задачу.

Вплоть до настоящего времени типичной остается ситуация, когда входные данные тщательно метрологически описываются, приборы и устройства, в которых они генерируются, регулярно эталонируются, тестируются и поверяются, но после переноса этих данных в вычислительную систему вся метрологическая информация оказывается совершенно невостребованной и никак в дальнейшем не используется, а получаемые в результате обработки выходные данные либо имеют метрологическую характеристику, достоверность которой вызывает сомнения, либо не имеют ее вовсе [5].

На точность вычисляемых параметров

оказывают влияние три основных фактора:

- погрешности измерительных трактов (первичных измерительных преобразователей, устройств сопряжения, аналого-цифровых преобразователей, устройств telemetry и т.п.),
- вычислительные погрешности (погрешности представления исходных данных в ЭВМ, погрешности округления результатов математических операций),
- апертурные погрешности (обусловленные несинхронностью измерения различных параметров).

Отсутствие универсального метода оценки влияния перечисленных факторов не позволяет в общем случае определить точность расчетных параметров. Стандартные же методы, применяемые при поверке и аттестации измерительных каналов, неприменимы по причине принципиальной невозможности формирования эталонных величин. В связи с этим актуальной становится задача *разработки методики, позволяющей единообразно описывать и учитывать точность как непосредственно измеряемых, так и расчетных параметров*.

Влияние вычислительных и измерительных погрешностей можно учесть при помощи традиционного анализа погрешностей, но данный метод обладает двумя существенными недостатками. Во-первых, анализ погрешностей производится вручную и для достаточно сложных алгоритмов данных он становится недопустимо трудоемким [1]. Во-вторых, анализ погрешностей выполняется а

priori, на основании данных о максимальных границах погрешности отдельных параметров и диапазонах их изменения, а информация о текущих значениях параметров никак не используется.

Единственной на сегодня моделью вычислений, позволяющей учитывать одновременно вычислительные и измерительные погрешности, является аппроксиметика [5], в основу которой положено понятие аппроксимета - класса объектов, представляемых триадой $A = mSp$, где $m \in Z$ - мантисса аппроксимета, $S \in N$, $S > 1$ - основание аппроксимета, $p \in Z$ - степень аппроксимета. К сожалению, аппроксиметика не может выступать в роли универсального метода оценки точности и достоверности данных в АСУ ТП, т.к. она

- позволяет оценить лишь порядок точности результата,
- относительно неудобна в работе, в связи с чем не получила практического распространения,
- не дает возможности учитывать апертурные погрешности.

Задача учета апертурных погрешностей в АСУ ТП стала актуальной сравнительно недавно - с внедрением концепции распределенных систем. В централизованных системах предыдущего поколения информация с датчиков поступала непосредственно на вход ЭВМ. В настоящее время применяются более эффективные распределенные иерархические системы, в которых первичная обработка данных на нижнем уровне производится отдельными контроллерами в непосредственной близости от датчикового оборудования, а результаты в цифровом виде пересыпаются на верхний уровень для отображения, дальнейшей обработки и анализа. При такой архитектуре измерительные данные могут поступать асинхронно с заметными задержками. К примеру, в системах телеметрии интервал опроса может достигать нескольких десятков секунд. В связи с этим возникает необходимость ответа на вопрос о том, насколько точно о значении расчетного параметра $p(t) = f(x_1(t), \dots, x_n(t))$ можно судить по результатам вычисления $f(x_1(t_1), \dots, x_n(t_n))$, где t_i - реальные моменты измерений $x_i(t)$.

В результате поисков способа универсального описания и учета точности измеряемых и расчетных параметров была разработана специальная модель потоковых вычислений с автовалидацией (МПВА) и создана ее программная реализация. В основу МПВА положены две фундаментальные концепции организации вычислений: потоковые вычисления (англ. dataflow computations) и вычисления с автовалидацией (англ. self-validated computations).

Парадигма потоковых вычислений изначально создавалась для эффективной организации параллельной обработки данных, но позднее выяснилось, что она с успехом может быть использована при решении задач межпроцессной синхронизации по данным и управлению в распределенных вычислительных системах [3]. В потоковых моделях процесс вычислений задается потоковым графом, узлы которого являются пассивными вычислительными элементами (акторами), а дуги представляют каналы передачи данных (токенов). Акторы исполняются по мере прихода к ним всех необходимых токенов (операндов), которые, в свою очередь, возникают либо как исходные данные программы, либо как результат исполнения предыдущих команд [8]. После срабатывания актора токены-результаты пересыпаются по каналам в другие акторы для дальнейших вычислений.

В контексте задачи учета точности расчетных параметров концепцию потоковых вычислений оказалось целесообразно использовать в силу наглядности представления зависимостей параметров и удобства моделирования поведения распределенной АСУ ТП в реальном времени.

В свою очередь, модели вычислений с автовалидацией появились как отклик на проблему оценки точности и достоверности информации, используемой в расчетах. Эти модели позволяют производить проверку точности и надежности результатов производимых расчетов автоматически, в качестве составной части самого вычислительного процесса. Известно множество моделей вычислений с автовалидацией: алгебра многозначных величин (R.C. Young, 1931), интер-

вальная арифметика (R. Moore, 1966), обобщенная интервальная арифметика (E. Hansen, 1975), арифметика направленных интервалов (S.M. Markov, 1992), ступенчатая интервальная арифметика (R.J. Lohner, 1993), аффинная арифметика (J.L.D. Comba, J. Stolfi, 1993), модифицированная арифметика гистограмм (D. Berleant, 1993), эллипсоидный анализ (Черноушко, Куржанский, Овсеевич, 1994).

В МПВА преимущественно используются идеи и методы интервальной арифметики [7], получившей в практических приложениях более широкое по сравнению с другими моделями распространение в силу своей простоты и универсальности. В данной модели вычислений данные представляются в виде замкнутых интервалов $X = [\underline{x}, \bar{x}] = \{x \in \mathbb{R} | \underline{x} \leq x \leq \bar{x}\}$, а операции над ними производятся согласно правилу

$$X_A \otimes X_B = \{x \in \mathbb{R} | x = x_1 \otimes x_2, x_1 \in X_A, x_2 \in X_B\}, \\ \otimes \in \{+, -, \cdot, /\}.$$

Арифметика токенов

Основной единицей информационного взаимодействия в МПВА является токен. При этом токен играет двойную роль: с одной стороны он определяет структуру представления результатов измерений и вычислений, а с другой является неделимой порцией данных, управляющих функционированием потоковой модели. Токен состоит из четырех компонентов:

- интервал значения $X = [\underline{x}, \bar{x}]$,
- метка времени $T = \{t \in \mathbb{N} | \underline{t} \leq t \leq \bar{t}\}$,
- динамическая характеристика $k \in \mathbb{R}$,
- уровень достоверности $r \in [0, 1]$.

Первый компонент токена представляет собой традиционную интервальную оценку измеряемой или расчетной величины, описываемой токеном. Помимо оценки значения параметра в системах автоматизации также важно знать, какому моменту времени это значение соответствует. Эта информация содержится в метке времени токена. В силу возможности неточного определения момента измерения и необходимости корректного описания результатов вычислений над дан-

ными несинхронными измерениями в МПВА используется интервальная метка времени.

Третий компонент токена - динамическая характеристика - используется для описания максимальной скорости изменения величины, описываемой токеном за период, соответствующий метке времени данного токена. Динамическая характеристика необходима для оценки влияния отдельных операндов на апертурную погрешность результата вычислений. Для этой же цели используется уровень достоверности, задающий степень надежности интервальной оценки. Конкретные механизмы определения точности и надежности результатов вычислений и измерений будут продемонстрированы ниже при описании арифметики токенов и алгоритмов работы потоковых элементов МПВА.

Рассмотрим правила выполнения арифметических операций над токенами. Множество всех токенов обозначим символом τ . Пусть A и B - токены с совпадающими метками времени:

$$A \in \tau, B \in \tau, T_A = T_B = T_0.$$

Результатом выполнения арифметической операции над токенами A и B будет также токен, причем

$$A+B = (X_A + X_B, T_0, k_A + k_B, r_{A0B}), \quad (1)$$

$$A-B = (X_A - X_B, T_0, k_A + k_B, r_{A0B}), \quad (2)$$

$$A \cdot B = (X_A \cdot X_B, T_0, k_A \cdot |X_B| + k_B \cdot |X_A|, r_{A0B}), \quad (3)$$

$$A/B = (X_A/X_B, T_0, (k_A \cdot |X_B| + k_B \cdot |X_A|)/\rho^2(X_B, 0), r_{A0B}), \\ \text{если } 0 \notin X_B, \quad (4)$$

где $r_{A0B} = \min\{r_A, r_B\}$, амплитуда интервала $|X| = \max\{|\underline{x}|, |\bar{x}|\}$, а расстояние между интервалами $\rho(X_A, X_B) = \max\{|\underline{x}_A - \underline{x}_B|, |\bar{x}_A - \bar{x}_B|\}$.

В формулах (1)-(4) интервал значения токена вычисляется согласно стандартным правилам интервальной математики. Метка времени результата по очевидным соображениям устанавливается такая же, как и у обоих операндов. Правила определения динамической характеристики продиктованы физическим смыслом данного компонента токена: представляя верхнюю оценку абсолютной

величины производной искомого параметра, динамическая характеристика вычисляется по правилам, похожим на обычные правила нахождения производных. Учитывая интервальную природу интервалов значений токенов в формулах (30, (4), $|X|$ и $\rho(X, 0)$ используются таким образом, чтобы в итоге получить верхнюю оценку скорости изменения результата. В свою очередь, последний компонент токена формируется в соответствии с естественным предположением о том, что достоверность результата любого действия над двумя операндами не может быть выше минимальной из достоверностей этих операндов.

Несложно показать, что операции сложения и умножения для токенов с общей временной меткой обладают свойствами коммутативности и ассоциативности. Как и для интервалов, для токенов не выполняется закон дистрибутивности, однако, если операцию включения определить как

$A \subset B \Leftrightarrow X_A \subset X_B \wedge T_A \subset T_B \wedge k_A \leq k_B \wedge r_A \geq r_B$, то для токенов с общей временной меткой имеет место закон субдистрибутивности, хорошо известный в интервальной арифметике. Для таких токенов также имеет место свойство монотонности по включению: если $A, B, C, D \in \tau$, $T_A = T_B = T_C = T_D$ и $A \subset B, C \subset D$, то $A \otimes C \subset B \otimes D \forall \otimes \in \{+, -, \cdot, /\}$.

Для токенов несложно определить основные элементарные функции, например,

$$\exp(A) = (\exp(X_A), T_A, k_A \cdot \exp(\bar{x}_A), r_A),$$

$$\ln(A) = (\ln(X_A), T_A, k_A / \underline{x}_A, r_A), \text{ при } X_A > 0.$$

Одна из основных особенностей МПВА - это автоматическая оценка влияния несинхронности операндов на достоверность результата операции. Для этого в модели введена операция приведения токенов к общей временной метке, заключающаяся в снижении уровня достоверности токена при расширении его метки времени. При этом учитывается три фактора: динамическая характеристика токена, ширина интервала значения и разность ширины исходной и расширенной временных меток. Пусть $A \in \tau$ - произвольный токен, T_1 - некоторая метка времени, причем

$\underline{t}_1 \leq \underline{t}_A, \bar{t}_A \leq \bar{t}_1$, т.е. $T_A \subset T_1$. Результатом приведения токена A к временной метке T_1 будем называть токен

$$A|_{T_1} = (X_A, T_A, k_A, r'_A),$$

где

$$r'_A = r_A \cdot \frac{\omega(X_A)}{\omega(X_A) + k_A \cdot (\omega(T_1) - \omega(T_A))}, \\ \omega(X) = \bar{x} - \underline{x}, \omega(T) = \bar{t} - \underline{t}. \quad (5)$$

Общей временной меткой для токенов A_1, \dots, A_n будем называть метку $T(A_1, \dots, A_n)$ вида

$$T(A_1, \dots, A_n) = \\ = \{t \in N \mid \min_{i=1, n} \{\underline{t}_{A_i}\} \leq t \leq \max_{i=1, n} \{\bar{t}_{A_i}\}\}.$$

Операция приведения токенов к общей временной метке позволяет распространить арифметические операции (1)-(4) на случай произвольных токенов. Для этого достаточно потребовать предварительного приведения к общей временной метке всех токенов, участвующих в вычислении какого-либо расчетного параметра.

При проведении вычислений с токенами следует учесть эффект зависимости результата от порядка проведения вычислений. Так, например, результат вычисления выражения $F = A \cdot (B+C)$ может отличаться от результата цепочки вычислений $D = B+C$, $F = A \cdot D$. Однако подобный эффект не является особенностью, характерной только для токенов, он имеет место быть и для интервалов (см., например, [6]). Кроме того, это свойство арифметики токенов никак не отражается на строгости получаемых интервальных оценок - варианты реализации могут влиять лишь на их точность.

Принципы работы МПВА

Описанная выше арифметика токенов определяет лишь правила выполнения операций, но не описывает логику организации вычислительного процесса. Рассмотрим принципы функционирования модели в целом.

Вычислительный процесс в МПВА описывается схемой взаимодействия потоковых элементов - связным ориентированным графом. Данная схема формируется из потоковых элементов, объединенных однонаправленными каналами передачи данных. Потоковые элементы представляют собой автономные функциональные объекты, выполняющие в модели генерацию, преобразование, проверку и уничтожение данных. Потоковые элементы оснащаются портами ввода и портами вывода, к которым подключаются каналы передачи данных. Номенклатура потоковых элементов, использующихся в МПВА, гораздо уже, чем в большинстве потоковых моделей: почти любой алгоритм, использующийся в реальных АСУ ТП, несложно реализовать при помощи всего 5 функциональных элементов - генератора, актора, валидатора, терминатора и канала передачи данных.

Генератор и терминатор являются интерфейсными элементами модели. Каждый генератор осуществляет опрос одного параметра и порождает токены, описывающие его, в соответствии с заданными настройками. Терминатор, наоборот, служит для экспорта результатов работы модели во внешнюю вычислительную среду. Вычислительные элементы МПВА - актор и валидатор - реализуют преобразования токенов согласно заданному алгоритму и проверку соответствия результатов наложенным семантическим ограничениям. В свою очередь каналы, моделирующие реальные сетевые соединения, выполняют функции пересылки токенов между потоковыми элементами с промежуточным их хранением в очереди ограниченной длины. Все потоковые элементы модели функционируют автономно и асинхронно.

Важной особенностью МПВА является правило автоматического "срабатывания" элементов при поступлении всех необходимых данных. Оно позволяет адекватно моделировать асинхронные по своей природе вычисления распределенных АСУ ТП. В отличие от МПВА на сегодняшний день в большинстве реальных систем используется принцип циклического срабатывания, при котором все

расчетные параметры вычисляются вне зависимости от темпа поступления измерительных данных.

Генератор, выполняющий импорт данных из внешней вычислительной среды, позволяет абстрагировать МПВА от особенностей реализации взаимодействия с конкретными системами АСУ ТП. В общем виде алгоритм функционирования генератора выглядит следующим образом. В начале цикла j -го генерации запоминается значение наблюдаемого параметра $x_{i,0}$, после чего он опрашивается с фиксированным, заранее заданным периодом Δt . Результат каждого очередного опроса $x_{j,i}$ сравнивается с предыдущим опросом $x_{j,i-1}$ и вычисляется разность $\Delta x_{j,i} = x_{j,i} - x_{j,i-1}$. В зависимости от режима работы, генерация токена производится либо после K_0 циклов опроса, либо при выходе значения измеряемого параметра за заданную апертуру (т.е. когда $|x_{j,i} - x_{j,0}| > D$), после $K < K_0$ циклов.

Интервал значения X j -го токена определяется диапазоном значений параметра, зарегистрированных за время генерации токена K с учетом всех возможных неточностей, задаваемых параметром *Error*:

$$\underline{x} = \min_{i=1,K} \{x_{j,i}\} - \text{Error}, \quad \bar{x} = \max_{i=1,K} \{x_{j,i}\} + \text{Error}.$$

Временная метка T формируется на основе информации о длительности периода генерации токена с поправкой $\text{Delay} = [\text{Delay}_{\min}, \text{Delay}_{\max}]$ на задержку между моментом получения результата опроса параметра генератором и временем, регистрации соответствующей физической величины в объекте контроля и управления:

$$\underline{t} = t_j - \text{Delay}_{\max}, \quad \bar{t} = t_j - \text{Delay}_{\min}.$$

Динамическая характеристика k , описывающая максимальную скорость изменения регистрируемого параметра за период генерации токена, определяется в соответствии со следующим выражением:

$$k = \frac{\max_{i=1,K} \{|\Delta x_{j,i}|\}}{\Delta t} \cdot \text{TVRatio}.$$

Здесь *TVRatio* представляет размерностный коэффициент соотношения использую-

щихся единиц измерения времени и регистрируемого параметра.

Иллюстрация принципа формирования токенов генератором приведена на рис.1 и 2.

Актёр представляет собой самостоятельный фрагмент вычислительной схемы Канторовича (см. [2]). У актёра имеется некоторое количество конфигурируемых входных ($N_I > 0$) и выходных ($N_O > 0$) портов, которые служат интерфейсом с каналами передачи данных. Программа работы актёра представляется набором унарных и бинарных операций над входными токенами. При получении всех необходимых данных актёр поглощает токены из входных портов, выполняет над ними заданные операции, а результаты также в форме токенов помещает в порты вывода.

Валидатор представляет собой автономный элемент, реализующий функции семантического контроля данных. У валидатора имеется некоторое число ($N_{IO} > 0$) конфигурируемых входных и такое же число выходных портов, которые служат интерфейсом с каналами передачи данных. Как и актёр, валидатор "срабатывает" при получении токенов во все порты ввода и производит последовательную проверку всех семантических ограничений. Семантические ограничения валидатора задаются в виде упорядоченного набора пар вида "отношение - коэффициент достоверности": $(\psi_i, c\psi_i)$, где $c\psi_i \in [0,1]$, $i = \overline{1, M}$. Коэффициент достоверности отра-

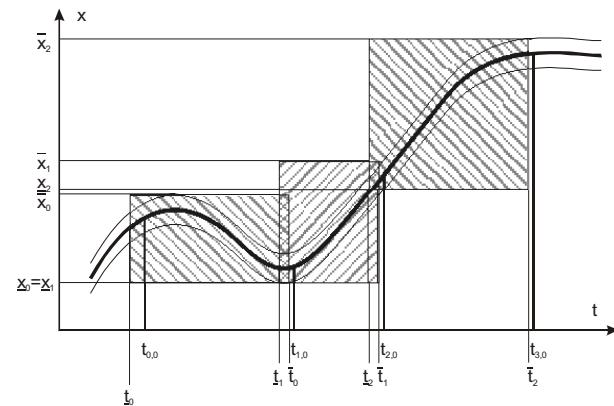


Рис.2. Токены, формируемые генератором

жает степень уверенности составителя модели в выполнимости соответствующего отношения в различных режимах функционирования системы.

Отношения, которые после подстановки в них соответствующих значений токенов не выполняются, считаются нарушенными, они индицируют присутствие в полученных токенах "плохих" данных. Уровни достоверности токенов, входящих в состав нарушенного отношения ψ_j , умножаются на коэффициент достоверности этого отношения $c\psi_j$. После проверки всех отношений токены с уровнями достоверности не превышающими критического порога r_{min} уничтожаются, а остальные токены переносятся из портов ввода в порты вывода.

Терминатор осуществляет уничтожение токенов и экспорт из МПВА представленных ими результатов вычислений. Он как и генератор позволяет абстрагировать модель от деталей взаимодействия с внешними вычислительными системами.

Канал в МПВА осуществляет передачу данных между потоковыми элементами и является единственным средством взаимодействия компонентов модели. Канал функционирует независимо от других элементов СВПЭ и параллельно с ними. Каналы осуществляют передачу токенов из выходного порта элемента-источника во входной порт элемента-приемника с промежуточным хранением пересылаемых данных в ограниченной очереди с настраиваемой дисциплиной обслуживания.

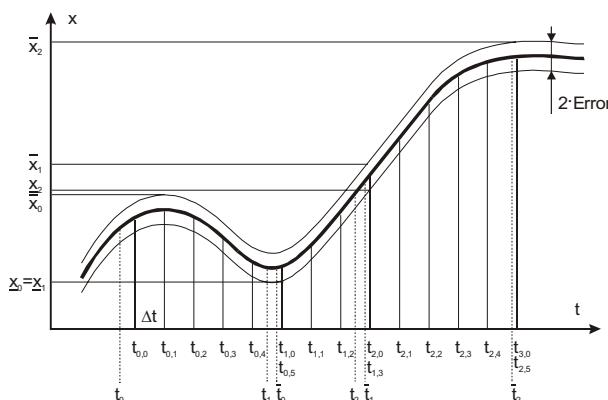


Рис.1. Определение компонентов токенов, формируемых генератором

Программная реализация МПВА

С целью практического тестирования разработанной модели была создана ее программная реализация - диагностическое приложение Dataflow SVC Simulator. Данное приложение предназначено для использования в составе АСУ ТП для решения задач оценки точности и достоверности расчетных данных. Оно позволяет сформировать модель, по которой осуществляются вычисления, и исследовать ее поведение на реальных данных без внесения изменений в программную и аппаратную архитектуру системы.

Приложение Dataflow SVC Simulator работает в среде операционных систем Windows NT 4.0 и Windows 2000 и в качестве основного источника данных использует SCADA-систему WinCC. Разработка данного приложения была осуществлена на высоконивневом языке программирования общего назначения VisualC++ v.6.0. Реализация МПВА в рамках специализированных систем моделирования была признана нецелесообразной в силу

сложности применения пользовательских типов данных и принципиальной невозможности или чрезмерной трудоемкости организации динамического взаимодействия со SCADA-системой. Кроме того, предпочтение было отдано языку VisualC++ v.6.0 в силу наличия в нем гибких и эффективных механизмов многопоточной обработки, позволяющих моделировать автономное функционирование компонентов распределенной вычислительной среды АСУ ТП в реальном масштабе времени.

На рис.3 показан внешний вид приложения и приведен пример потоковой схемы вычислений.

Практическое применение модели и ее программной реализации

Апробирование предложенной модели потокоориентированных вычислений с авторализацией и диагностического пакета Dataflow SVC Simulator было проведено в рамках проекта по разработке эксперименталь-

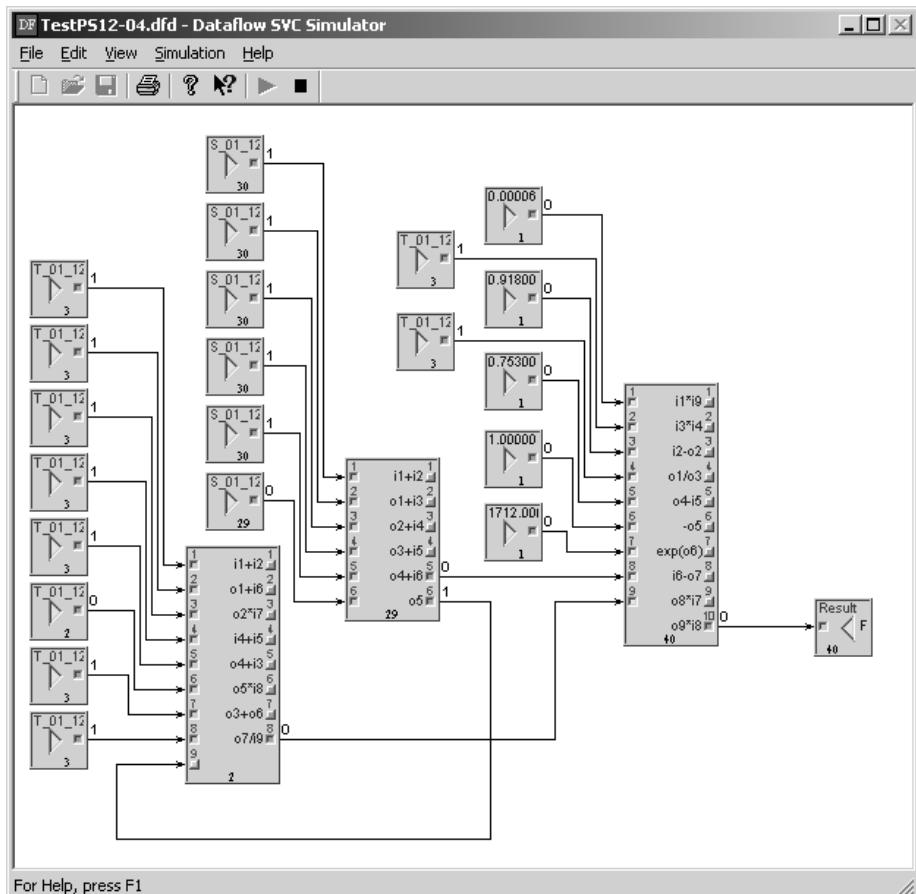


Рис.3. Пример потоковой схемы в Dataflow SVC Simulator

ной системы отображения информации для диспетчерского щита Самарских тепловых сетей (СТС). Данная система представляет собой типичный пример распределенной АСУ ТП: данные в ней собираются с множества удаленных объектов по каналам телеметрии, а их централизованная обработка и отображение в реальном масштабе времени производятся на АРМе диспетчера.

Одним из важных расчетных параметров, использующихся при оценке состояния оборудования отдельных насосных и целостности коммуникаций, является расход теплоносителя. Для вычисления расхода теплоносителя Q на насосных установках СТС используется следующая формула [4]:

$$Q = 1712 \times \left[1 - \exp \left(\frac{0,057855 \cdot \sum_{k=1}^6 (U_k \cdot I_k) / \sum_{k=1}^6 a_k}{P_{out} - 0,918 \cdot P_{in}} - 0,753 \right) \right] \times \sum_{i=1}^6 a_i, \quad (6)$$

где Q - расход теплоносителя [$\text{м}^3/\text{ч}$], U_k - линейное напряжение, действующее в цепи питания электродвигателя № k [кВ], I_k - ток, потребляемый электродвигателем насоса № k [А], a_k - двоичный признак включения насоса № k [0, 1], P_{in} - давление на приеме насосной установки [$\text{кгс}/\text{см}^2$], P_{out} - давление на выкиде насосной установки [$\text{кгс}/\text{см}^2$], а постоянные величины рассчитаны в соответствии с конкретными особенностями используемых насосных установок СЭ-1250-70.

Для формулы (6) была составлена потоковая модель (см. рис.3), все параметры компонентов которой были заданы в соответствии с паспортными характеристиками измерительного и коммуникационного оборудования. В результате проведенной серии тестов были изучены характеристики расчетных данных, получить которые без Dataflow SVC Simulator не удавалось. Во-первых, было установлено, что точность результат расчета расхода теплоносителя крайне низка и использовать получаемые значения можно лишь как грубую аппроксимацию: отклонение расчетного значения от истинного может до-

стигать 15%. Во-вторых, было исследовано влияние задержек измерений отдельных параметров на достоверность результата в целом. Полученные данные могут быть использованы при планировании периодичности проведения телеметрии в системе с целью повышения достоверности расчета расхода теплоносителя. Кроме того, Dataflow SVC Simulator дает возможность автоматического обнаружения отключения отдельных каналов и использования в системе признака недостоверности параметра.

Заключение

В настоящее время проверка точности и достоверности расчетных данных АСУ ТП в большинстве случаев не производится в силу отсутствия простых и эффективных методов учета измерительных, вычислительных и апертурных погрешностей. Для решения данной задачи была разработана модель потоковых вычислений с автовалидацией, позволяющая оценивать точность и надежность получаемых результатов автоматически в ходе вычислительного процесса.

На основе предложенной модели вычислений было создано диагностическое приложение Dataflow SVC Simulator, позволяющее отслеживать упомянутые характеристики расчетных параметров в реальных АСУ ТП без внесения изменений в программную и аппаратную архитектуру системы. В ходе апробации модели и ее программной реализации в составе экспериментальной системы отображения информации для диспетчерского щита СТС были получены результаты, подтверждающие практическую полезность предложенных методов.

СПИСОК ЛИТЕРАТУРЫ

1. Житников В.П., Шерыхалина Н.М. Оценка достоверности численных результатов при наличии нескольких методов решения задачи // Вычислительные технологии. 1999. Т.4. №6.
2. Канторович Л.В. О математической символике, удобной для выполнения машинных вычислений // ДАН СССР. 1957. №113.

-
3. Колосовский А.П. Организация вычислений, управляемых потоками данных, в распределенных системах автоматизации эксперимента: Дисс. на соиск. уч. ст. к.т.н. Куйбышев: Куйбышевский филиал Института машиноведения им. А.А. Благонравова, 1990.
4. Патент РФ №1783869.
5. Юровицкий В.М. Аппроксиметика. М.: ИАА ФинИст, 1995.
6. Hyvonen E. Evaluation of Cascaded Interval Function Constraints. Proceedings of the International Workshop on Constraint-Based Reasoning CONSTRAINT-95, Melbourne Beach, USA. 1995.
7. Moore R.E. The automatic analysis and control of error in digital computing based on the use of interval numbers// Error in Digital Computation, L.B. Rall, Ed., John Wiley and Sons, Inc., New York, 1965. Vol.1.
8. Silc J., Robic B., Ungerer T. Asynchrony in Parallel Computing: From Dataflow to Multithreading. Parallel and Distributed Computing Practices, 1998. Vol.1. №1.

AN APPROACH TO EVALUATION OF ACCURACY AND RELIABILITY OF COMPUTATIONAL DATA IN DISTRIBUTED MONITORING AND AUTOMATION SYSTEMS

© 2002 L.A. Shereshevsky, A.A. Sidorov

Sensors, Modules, Systems, Ltd, Samara

An approach to evaluation of accuracy and reliability of computational data in distributed monitoring and automation systems is presented. At the core of this approach is a model based on both dataflow computations and self-validated computations. A software implementation of the model is described and the results of its practical application are considered.

ОЦЕНКА ДОСТОВЕРНОСТИ ПАРАМЕТРОВ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ АНАЛИЗА НЕЧЕТКИХ ВЕЛИЧИН

© 2002 А.Г. Шопин

Научно-внедренческая фирма "Сенсоры, модули, системы", г. Самара

Представлен математический аппарат для оценки достоверности значений параметров технологических процессов. В его основу положено представление значения параметра как нечеткой величины. Рассмотрен процесс устаревания информации. Выделены классы нечетких величин, инвариантных к устареванию. Для данных классов определены вероятностные характеристики.

Введение

Современные системы, контролирующие ход технологических процессов, характеризуются большим количеством контролируемых параметров, значения которых получаются из разных источников. Разные источники поставляют информацию асинхронно, независимо друг от друга, с разной периодичностью. Полученная информация обрабатывается не по мере поступления, а в произвольные моменты времени. Поэтому в таких системах недостаточно получать значения параметров, а необходимо знать, можно ли использовать в расчетах данные значения через некоторый промежуток времени, или они будут устаревшими, и результат вычислений будет недостоверным. В данной статье делается попытка оценки достоверности полученных данных для некоторого момента времени, в общем случае не совпадающего с моментом времени, в который было зафиксировано значение параметра. Кроме того, в работе рассматривается достоверность результата некоторых алгебраических операций.

Описание значения параметра как нечеткой величины

В ходе наблюдения за технологическими параметрами мы получаем некоторые значения этих параметров в некоторые моменты времени. При этом в данных значениях присутствует суммарная погрешность, состоящая из следующих частей:

1. Погрешность, присущая методу измерений параметра (*неопределенность значения*).

2. Погрешность, возникающая при представлении информации в компьютере (*погрешность хранения*). Она складывается из *погрешности округления значений параметра* и *погрешности округления времени получения параметра*.

3. Погрешность, обусловленная тем, что нам не известен момент времени, в который был зафиксирован параметр (*неопределенность времени*). Этот тип погрешности имеет ту же природу, что и погрешность округления времени.

В результате мы не знаем точного значения параметра, но мы можем попытаться оценить область, в которой с некоторой достоверностью находится параметр. Фактически эта область и характеризует параметр. Для того чтобы описать эту область воспользуемся теорией возможности[1].

Рассмотрим непустое подмножество X из универсального множества U . Оно однозначно определяется характеристическим функционалом [2]

$$I_X(u) = \begin{cases} 1 & \text{если } u \in X \\ 0 & \text{если } u \notin X \end{cases}, \quad (0)$$

т.е. подмножество X определяется как совокупность объектов, имеющих некоторое общее свойство, наличие или отсутствие которого у любого элемента u задается характеристическим функционалом. Причем относительно природы объекта не делается никаких предположений.

Для нечеткого подмножества, являющегося расширением понятия множества в классическом смысле, на пространстве объектов

$U=\{u\}$ вводится уже не функционал вида (0), а характеристическая функция, задающая для всех элементов степень наличия у них некоторого свойства, по которому они относятся к подмножеству X . Эта характеристическая функция для нечеткого множества носит название *функции возможности*.

Нечеткое подмножество X множества U характеризуется функцией возможности $\mu_x : U \rightarrow [0,1]$, которая ставит в соответствие каждому элементу $u \in U$ число $\mu_x(u)$ из интервала $[0, 1]$, характеризующее степень принадлежности элемента u подмножеству X . Причем 0 и 1 представляют собой соответственно низшую и высшую степень принадлежности элемента к определенному подмножеству.

Нечеткая переменная характеризуется тройкой $(N, U, R(N,u))$, где N - название переменной; U - универсальное множество с базовой переменной u ; $R(N,u)$ - нечеткое подмножество множества U , представляющее собой нечеткое ограничение на значения переменной u , обусловленное N [2].

Будем считать, что универсальное множество – ось действительных чисел. Тогда нечеткая переменная X описывается своей функцией возможности $\mu(x)$.

Необходимо найти функцию возможности параметра в момент его измерения и рассмотреть, как изменится эта функция через некоторое время. Назовем такое изменение функции возможности *устареванием информации*.

Итак, рассмотренные выше неопределенность значения и погрешность округления описываются с помощью аппарата функций возможности, а погрешности округления времени и неопределенности времени описываются с помощью устаревания информации.

Необходимость учитывать устаревание информации возникает в следующих случаях:

- При сборе значений параметров посыпается запрос и ожидается ответ. При отсутствии датирования значений мы не можем определить, какому времени из интервала от запроса до ответа сопоставить полученное значение параметра.

- При сборе данных от некоторого устройства получаются ответы, но отсутствует информация о канале связи с датчиком. В результате может возникнуть ситуация, когда фиксируется некоторые устаревшие значения параметров.

- При необходимости произвести некоторые алгебраические операции над величинами, измеренными в разное время, используется приведение параметров к одному временному срезу. Кроме того, приведение к одному временному срезу используется при хранении групп параметров с единой временной меткой.

В данной статье будет рассмотрено, как меняются нечеткие величины (их функции возможностей) при алгебраических операциях над ними и в результате устаревания информации.

Кроме того, необходимо рассмотреть, как использовать полученные функции возможности для контроля выхода за уставки и прочих методик контроля.

Устаревание информации

Рассмотрим процесс изменения параметра X_t . Сначала предположим, что значение параметра в момент времени t является известным (идеальный случай – область, в которой лежит значение параметра, состоит из одной точки, нечеткая величина вырождена в четкую величину). Необходимо узнать значение параметра в момент времени $t + \tau$. Ниже будет рассмотрен общий случай для нечеткого значения X_t .

Предположим, что процесс изменения параметра винеровский (данное предположение подтверждается экспериментами по анализу рядов значений технологических параметров). Тогда изменения параметра будут удовлетворять нормальному закону

$$L(X_{t+\tau} - X_t) = N(0, \sigma^2 \tau).$$

Воспользовавшись вероятностным подходом построения функции возможности [2], по известному значению параметра X_t построим функцию возможности $\mu_{X_{t+\tau}}(x)$. Она

характеризует возможность того, что через некоторый промежуток времени τ значение параметра будет равно x , при условии, что в момент времени t значение было X_t .

$$\mu_\tau(x) = e^{-\frac{(x-X_t)^2}{2\sigma^2\tau}}. \quad (1)$$

Назовем ее *функцией устаревания информации*. Ее график приведен на рис.1.

Теперь рассмотрим общий случай, когда X_t можно считать нечеткой величиной с функцией возможности $\mu_{X_t}(x)$, вид которой можно определить исходя из дополнительной информации о методе и средствах измерения параметра, а также об известных погрешностях. По ней и по известной функции устаревания информации найдем нечеткую величину $X_{t+\tau}$. Воспользуемся принципом обобщения [1]:

$$\mu_{X_{t+\tau}}(x) = \sup_y \mu_{X_t}(y) e^{-\frac{(x-y)^2}{2\sigma^2\tau}}. \quad (2)$$

В частном случае, когда величина X_t является четкой, мы получаем формулу (1).

На рис.2 и 3 приведены примеры функций возможности нечетких величин и вычисленные функции возможности этих величин через некоторый промежуток времени.

Отметим, что для функции устаревания информации должно выполняться следующее интуитивно понятное *свойство* – результат устаревания нечеткой величины для произвольного промежутка времени должен

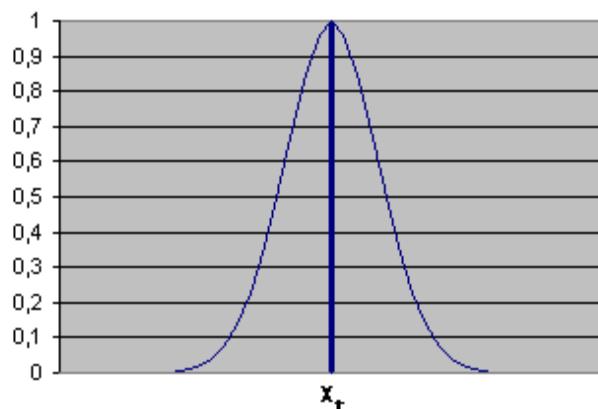


Рис.1. Функция устаревания информации

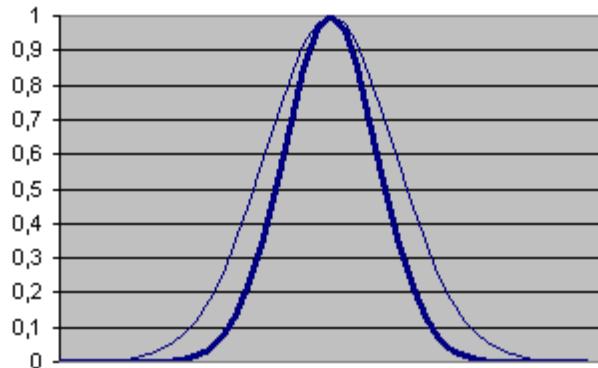


Рис.2. Устаревание гауссовой нечеткой величины

быть эквивалентен результату последовательного устаревания для разбиения этого промежутка на произвольные части:

$$\begin{aligned} \forall \tau_1, \forall \tau_2 \quad X_{t+\tau_1+\tau_2} &= h(X_t, \sigma^2, \tau_1 + \tau_2) = \\ &= h(X_{t+\tau_1}, \sigma^2, \tau_2) = h(h(X_t, \sigma^2, \tau_1), \sigma^2, \tau_2). \end{aligned} \quad (3)$$

Другими словами, процесс устаревания должен описываться таким образом, чтобы результат не зависел от способа расчета нечеткой величины в течение времени устаревания.

Докажем справедливость формулы (3).
Перепишем (3) используя (2)

$$\begin{aligned} \mu_{X_{t+\tau_1+\tau_2}}(x) &= \sup_y \mu_{X_{t+\tau_1}}(y) e^{-\frac{(x-y)^2}{2\sigma^2\tau_2}} = \\ &= \sup_y \left(\sup_z \mu_{X_t}(z) e^{-\frac{(z-y)^2}{2\sigma^2\tau_1}} \right) e^{-\frac{(x-z)^2}{2\sigma^2\tau_2}}. \end{aligned} \quad (3')$$

Необходимо доказать, что (3') окажется равным

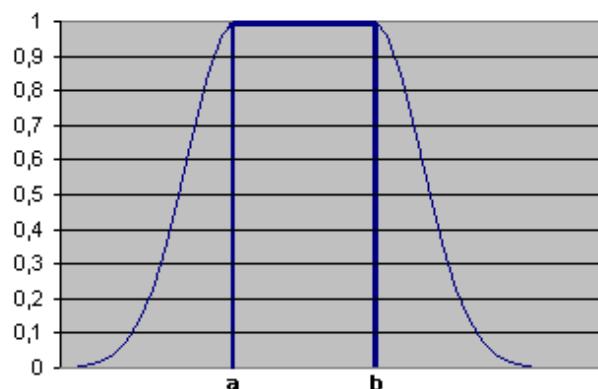


Рис.3. Устаревание кусочно-постоянной нечеткой величины

$$\mu_{X_{t+\tau_1+\tau_2}}(x) = \sup_z \mu_{X_t}(z) e^{-\frac{(x-z)^2}{2\sigma^2(\tau_1+\tau_2)}}, \quad (4)$$

$$\begin{aligned} \mu_{X_{t+\tau_1+\tau_2}}(x) &= \sup_y \left(\sup_z \mu_{X_t}(z) e^{-\frac{(z-y)^2}{2\sigma^2\tau_1}} \right) e^{-\frac{(x-y)^2}{2\sigma^2\tau_2}} = \\ &= \sup_y \sup_z \left(\mu_{X_t}(z) e^{-\frac{(z-y)^2}{2\sigma^2\tau_1}} e^{-\frac{(x-y)^2}{2\sigma^2\tau_2}} \right) = \\ &= \sup_x \mu_{X_t}(z) \sup_y \left(e^{-\frac{(z-y)^2}{2\sigma^2\tau_1}} e^{-\frac{(x-y)^2}{2\sigma^2\tau_2}} \right) \end{aligned} \quad . \quad (5)$$

Докажем, что

$$\sup_y \left(e^{-\frac{(z-y)^2}{2\sigma^2\tau_1}} e^{-\frac{(x-y)^2}{2\sigma^2\tau_2}} \right) = e^{-\frac{(x-z)^2}{2\sigma^2(\tau_1+\tau_2)}}.$$

Введем замены

$$\begin{aligned} z - y &= a \\ x - z &= b \end{aligned}$$

$$\sup_y \left(e^{-\frac{(z-y)^2}{2\sigma^2\tau_1}} e^{-\frac{(x-y)^2}{2\sigma^2\tau_2}} \right) = \sup_a \left(e^{-\frac{a^2}{2\sigma^2\tau_1}} e^{-\frac{(b-a)^2}{2\sigma^2\tau_2}} \right),$$

\sup_a достигается в точке определяемой из условия

$$\left(\frac{a^2}{2\sigma^2\tau_1} + \frac{(b-a)^2}{2\sigma^2\tau_2} \right)'_a = 0,$$

$$\frac{a}{\sigma^2\tau_1} + \frac{a-b}{\sigma^2\tau_2} = 0 \Rightarrow a = \frac{b\sigma^2\tau_1}{\sigma^2\tau_1 + \sigma^2\tau_2} = \frac{b\tau_1}{\tau_1 + \tau_2},$$

$$\begin{aligned} \sup_a \left(e^{-\frac{a^2}{2\sigma^2\tau_1}} e^{-\frac{(b-a)^2}{2\sigma^2\tau_2}} \right) &= e^{-\frac{\left(\frac{b\tau_1}{\tau_1 + \tau_2}\right)^2}{2\sigma^2\tau_1}} e^{-\frac{\left(b - \frac{b\tau_1}{\tau_1 + \tau_2}\right)^2}{2\sigma^2\tau_2}} \\ &= e^{-\frac{1}{2\sigma^2} \left(\frac{\left(\frac{b\tau_1}{\tau_1 + \tau_2}\right)^2}{\tau_1} + \frac{\left(\frac{b\tau_2}{\tau_1 + \tau_2}\right)^2}{\tau_2} \right)} = e^{-\frac{b^2}{2\sigma^2} \left(\frac{\tau_1 + \tau_2}{(\tau_1 + \tau_2)^2} \right)} \\ &= e^{-\frac{(x-z)^2}{2\sigma^2(\tau_1 + \tau_2)}} \end{aligned}$$

Результат подставляем в (5), получаем (4).

Устаревание нечетких переменных основных классов

Введем несколько классов нечетких величин и рассмотрим их устаревание.

1. Устаревание четкой величины (см. рис.1)

$$X_t = a \Rightarrow \mu_{X_t}(x) = \begin{cases} 1; & x = a \\ 0; & \text{иначе} \end{cases}$$

В результате устаревания получается нечеткая величина с функцией возможности

$$\mu_{X_{t+\tau}}(x) = e^{-\frac{(x-a)^2}{2\sigma^2\tau}}.$$

2. Устаревание одноточечной нечеткой величины $X_t = a$ с возможностью

$$\mu_0 \Rightarrow \mu_{X_t}(x) = \begin{cases} \mu_0; & x = a \\ 0; & \text{иначе} \end{cases}.$$

В результате устаревания получается нечеткая величина с функцией возможности

$$\mu_{X_{t+\tau}}(x) = \mu_0 e^{-\frac{(x-a)^2}{2\sigma^2\tau}}.$$

Введем обозначение: будем называть нечеткую величину с функцией распределения

возможности $\mu_0 e^{-\frac{(x-a)^2}{2\sigma^2}}$ гауссовой и будем ее обозначать $\Omega(\mu_0, a, \sigma^2)$.

3. Гауссова нечеткая величина (см. рис.2)

$$X_t = \Omega(\mu_0, a, \sigma_0^2) \quad \mu_{X_t}(x) = \mu_0 e^{-\frac{(x-a)^2}{2\sigma_0^2}}.$$

В результате устаревания получается нечеткая величина с функцией возможности

$$\mu_{X_{t+\tau}}(x) = \mu_0 e^{-\frac{(x-a)^2}{2(\sigma^2\tau + \sigma_0^2)}}, \text{ т.е. гауссова нечеткая величина } X_{t+\tau} = \Omega(\mu_0, a, \sigma^2\tau + \sigma_0^2).$$

Из высказанного видно, что нечеткая гауссова величина в результате устаревания не выходит за пределы данного класса.

Множество гауссовых нечетких величин включает в себя множество четких величин и множество одноточечных нечетких величин.

4. Кусочно-постоянная нечеткая величина (рис.3)

$$\mu_{X_t}(x) = \begin{cases} \mu_0; & x \in [a, b] \\ 0; & \text{иначе} \end{cases}. \quad (6)$$

Введем обозначение: обозначим нечеткую величину с функцией возможности (6) $L(\mu_0, a, b)$. Множество нечетких величин класса $L(\mu_0, a, b)$ включает в себя множество четких величин и множество одноточечных нечетких величин.

В результате устаревания получается нечеткая величина с функцией возможности:

$$\mu_{X_{t+\tau}}(x) = \begin{cases} \mu_0 e^{-\frac{(x-a)^2}{2\sigma^2\tau}}; & x < a \\ \mu_0; & x \in [a, b] \\ \mu_0 e^{-\frac{(x-b)^2}{2\sigma^2\tau}}; & x > b \end{cases}. \quad (7)$$

Введем обозначение: обозначим нечеткую величину с функцией возможности (7) $M(\mu_0, a, b, \sigma^2)$. Множество нечетких величин класса $M(\mu_0, a, b, \sigma^2)$ включает в себя множества нечетких величин классов $L(\mu_0, a, b)$ и $\Omega(\mu_0, a, \sigma^2)$, и, следовательно, множество четких величин и множество одноточечных нечетких величин.

Нечеткую величину класса $M(\mu_0, a, b, \sigma^2)$, можно представить в виде объединения нечетких величин классов $L(\mu_0, a, b)$ и $\Omega(\mu_0, a, \sigma^2)$:

$X_{t+\tau} = \Omega(\mu_0, a, \sigma^2\tau) \cup L(\mu_0, a, b) \cup \Omega(\mu_0, b, \sigma^2\tau)$
где операция объединения определена как операция взятия максимума[1].

Нечеткая величина класса $M(\mu_0, a, b, \sigma^2)$ в результате устаревания не выходит за пределы данного класса.

Сложение нечетких переменных основных классов

1. Гауссова нечеткая величина

$$X_1 = \Omega(\mu_1, a_1, \sigma_1^2),$$

$$X_2 = \Omega(\mu_2, a_2, \sigma_2^2).$$

Рассмотрим частный случай – равенство максимальных возможностей:

$$X_1 = \Omega(\mu_0, a_1, \sigma_1^2),$$

$$X_2 = \Omega(\mu_0, a_2, \sigma_2^2),$$

тогда

$$X_1 + X_2 = \Omega(\mu_0, a_1 + a_2, (\sigma_1 + \sigma_2)^2).$$

Замечание: В отличие от теории вероятности сумма двух гауссовых величин имеет большую дисперсию $(\sigma_1 + \sigma_2)^2 > \sigma_1^2 + \sigma_2^2$

2. Величина класса $M(\mu_0, a, b, \sigma^2)$

$$X_1 = M(\mu_1, a_1, b_1, \sigma_1^2),$$

$$X_2 = M(\mu_2, a_2, b_2, \sigma_2^2).$$

Рассмотрим частный случай – равенство максимальных возможностей

$$X_1 = M(\mu_0, a_1, b_1, \sigma_1^2),$$

$$X_2 = M(\mu_0, a_2, b_2, \sigma_2^2),$$

тогда $X_1 + X_2 = M \times$

$$\times (\mu_0, a_1 + a_2, b_1 + b_2, (\sigma_1 + \sigma_2)^2)$$

Приложение – контроль уставок

При анализе хода технологических процессов используется контроль над превышением или принижением значениями параметров некоторых величин, называемых *уставками*. При превышении/принижении этих величин возникает необходимость в каких-либо действиях, которые могут исполняться либо автоматическими системами, либо человеком. По серьезности возможных последствий входа за уставки выделяются предупредительные и аварийные уставки.

Для нечетких величин нельзя однозначно сказать превышено ли значение уставки, поэтому возникает задача определения возможности выхода значением параметра за уставку.

Пусть нами была найдена нечеткая переменная X_t с функцией возможности $\mu_{X_t}(x)$.

Пусть у нас есть набор уставок, например, пара предупредительных и пара аварийных уставок ($\underline{u}_{\text{авар}}, \overline{u}_{\text{пред}}, \underline{u}_{\text{пред}}, \overline{u}_{\text{авар}}$).

Построим функцию возможности выхода за аварийные ($\mu_a(x)$) и предупредительные ($\mu_w(x)$) уставки:

$$\mu_a(x) = \begin{cases} 0, & \text{при } x \in [\underline{u}_{\text{авар}}; \overline{u}_{\text{авар}}] \\ 1, & \text{при } x > \overline{u}_{\text{авар}} \text{ или } x < \underline{u}_{\text{авар}} \end{cases}$$

$$\mu_w(x) = \begin{cases} 0, & \text{при } x \in [\underline{u}_{\text{пред}}; \overline{u}_{\text{пред}}] \\ 1, & \text{при } x > \overline{u}_{\text{пред}} \text{ или } x < \underline{u}_{\text{пред}} \end{cases}$$

Фактически мы определили две нечеткие величины

- 1) A с функцией возможности $\mu_a(x)$
- 2) W с функцией возможности $\mu_w(x)$.

На рис.4 представлен график этих функций.

Для определения возможности выхода за аварийные уставки необходимо найти нечеткую величину $X_t \cap A$ (для предупредительных - $X_t \cap W$) с функцией возможности $\min(\mu_a(x), \mu_{X_t}(x))$ (соответственно, $\min(\mu_w(x), \mu_{X_t}(x))$) и определить ее высоту (верхнюю грань ее функции возможности).

Соответственно достоверности выхода за уставки равны:

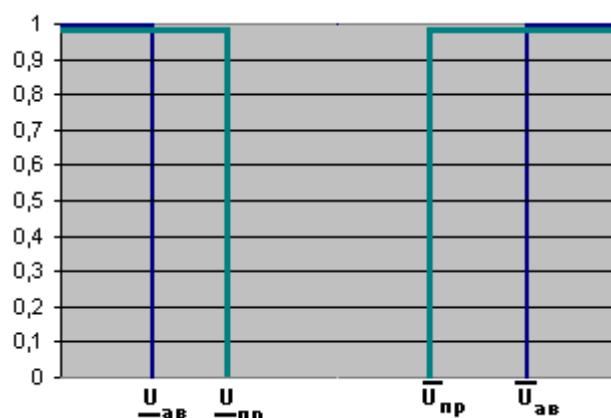


Рис.4. Функция возможности выхода за уставки

$N_a = \sup_x (\min(\mu_a(x), \mu_{X_t}(x)))$, для аварийных уставок;

$N_w = \sup_x (\min(\mu_w(x), \mu_{X_t}(x)))$, для аварийных уставок.

Интерпретация полученных результатов: в результате вычислений была получена возможность выхода за уставки N_a и N_w . Определим пороговое значение α из интервала $(0, 1)$, сравним полученные достоверности с ним и для данного α получим, превышена уставка или нет. При отсутствии дополнительной информации о контролируемых параметрах можно принять $\alpha = \frac{1}{2}$.

Стоит отметить, что для четких величин достоверности N_a и N_w могут принимать значения только 0 или 1, и независимо от порогового значения мы можем однозначно сказать превышена уставка или нет.

Приложение – определение вероятностных характеристик нечеткой величины

Пусть у нас есть нечеткая величина X с функцией возможности $\mu(x)$. Эту нечеткую величину можно рассматривать как случайную. Для нее строится вероятностная мера. В результате имеем случайную величину \mathbf{x} с плотностью распределения вероятности

$$f(x) = \frac{\mu(x)}{S}, \quad (6)$$

$$\text{где } S = \int_{-\infty}^{\infty} \mu(x) dx. \quad (7)$$

Найдем математическое ожидание и дисперсию случайной величины \mathbf{x} .

Пусть $X = M(\mu, a, b, \sigma^2)$, тогда из (7) получаем

$$\begin{aligned} S &= \sqrt{2\pi} \mu_0 \sigma + \mu_0 (b - a) = \\ &= \mu_0 (\sqrt{2\pi} \sigma + (b - a)) \end{aligned}, \quad (8)$$

$$M\mathbf{x} = \int_{-\infty}^{\infty} x \frac{\mu(x)}{S} dx = \frac{a + b}{2}, \quad (9)$$

$$\begin{aligned} D\mathbf{x} &= \int_{-\infty}^{\infty} x^2 \frac{\mu(x)}{s} dx - (M\mathbf{x})^2 = \\ &= \sigma^2 + \frac{(b-a)^2}{12} + \\ &+ (b-a)\sigma \frac{6\sigma + \sqrt{2\pi}(b-a)}{6\sqrt{2\pi}\sigma + 6(b-a)} . \quad (10) \end{aligned}$$

Для предельных случаев, когда нечеткая величина типа $M(\mu, a, b, \sigma^2)$ вырождается в гауссоиду или в прямоугольник получаем:

1. Для гауссовой величины

$$\Omega(\mu, a, \sigma^2) = M(\mu, a, a, \sigma^2)$$

дисперсия равна $D\mathbf{x} = \sigma^2$.

2. Для кусочно-постоянной величины $L(\mu_0, a, b) = M(\mu, a, b, 0)$ дисперсия равна

$$D\mathbf{x} = \frac{(b-a)^2}{12}.$$

Таким образом, по известной функции возможности можно оценить математическое ожидание и дисперсию нечеткой величины и по ним оценить погрешность полученного значения параметра.

Заключение

Представление значения параметра как

нечеткой величины дает возможность контролировать достоверность получаемых значений. Рассмотрение механизма устаревания позволило найти два класса нечетких величин: $\Omega(\mu, a, \sigma^2)$ и $M(\mu, a, b, \sigma^2)$, представители которых не выходят за рамки этих классов, как в результате устаревания, так и в результате операции сложения. Для устаревания доказана инвариантность к разбиению промежутка времени на интервалы.

Данный подход позволяет описывать неопределенность значений параметров, предсказывать поведение измеряемых (и зависимых от них вычисляемых) параметров с ходом времени и оценивать возможность различного рода критических ситуаций.

СПИСОК ЛИТЕРАТУРЫ

1. Дюбуа Д., Прад А. Теория возможностей. Приложения к представлению знаний в информатике. М.: Радио и связь, 1990.
2. Алтунин А.Е., Семухин М.В. Модели и алгоритмы принятия решений в нечетких условиях. Тюмень: Тюменский государственный университет, 2000.

THE EVALUATING OF RELIABILITY FOR ENGINEERING PROCESS PARAMETERS ON BASIS OF FUZZY VALUES ANALYSIS

© 2002 A.G. Shopin

Sensors, Modules, Systems, Ltd, Samara

Mathematical apparatus for evaluating of reliability for engineering process parameters are present. The basis is representation the parameter value as fuzzy value. The process of information obsolescence is considered. The obsolescence invariant classes are evolved. For this classes probabilistic characteristics are founded.

Сдано в набор 26.06.2002 г.
Офсетная печать Усл. печ. л. 21.6

Подписано к печати 18.10.2002 г.
Усл. кр.-отт 4.6 тыс. Уч.-изд. л 18.6

Формат бумаги 60×88^{1/8}
Тираж 200 экз. Зак. 80

Лицензия ЛР №040910 от 10.08.98 г. Государственного комитета Российской Федерации по печати
Учредители: Самарский научный центр Российской академии наук, Президиум СНЦ РАН

Адрес издателя: 443001, Самара, Студенческий пер., 3а
Отпечатано в типографии СНЦ РАН, 443001, Самара, Студенческий пер., 3а