

КЛАССИФИКАЦИЯ ПОТОКА СИСТЕМНЫХ ОШИБОК С ПОМОЩЬЮ ГИБРИДНОЙ МОДИФИКАЦИИ НЕЙРОННОЙ СЕТИ ВАНГА-МЕНДЕЛЯ

© 2014 О.П.Солдатова¹, Е.М.Пудикова²

¹ Самарский государственный аэрокосмический университет имени академика С.П. Королёва
(национальный исследовательский университет)

² Компания NetCracker Technology Corp, г. Самара

Поступила в редакцию 17.12.2013

В статье исследуется применение предложенной авторами гибридной модификации нейронной нечёткой продукционной сети Ванга-Менделя с несколькими выходами, для решения задачи классификации потока сообщений об ошибках программной системы. Описаны свойства потока сообщений на естественном языке, приведена методика формирования обучающего множества на основе реального потока сообщений об ошибках. Описаны алгоритмы обучения сети и исследована эффективность использования гибридной сети для решения задачи.

Ключевые слова: потоки сообщений на естественном языке, обучающее множество, нечёткая нейронная продукционная сеть, сеть Ванга-Менделя, самоорганизующийся слой, алгоритм обучения, классификация.

ВВЕДЕНИЕ

Современные промышленные программные комплексы это многомодульные системы, разрабатываемые в течение длительного времени, часто независимыми друг от друга командами разработчиков. Результатом такой работы должен быть единый программный комплекс, разделяющий общие машинные и программные ресурсы вычислительной системы. В таких условиях, в рамках жизненного цикла продукта, наблюдаются ситуации, при которых система генерирует сообщения о сбоях или ошибках, которые необходимо исправить.

Сообщения о подобных событиях формируют поток, требующий обработки в реальном времени. Ввиду этого необходимо разработать программный модуль, который может своевременно обрабатывать входной поток сообщений, разделяя его на классы, для принятия решения о возможности автоматической обработки события или о необходимости разрешения возникшей ситуации при помощи экспертов.

Особенностью обрабатываемого потока данных является то, что в его состав входят редко повторяющиеся сообщения или же группы сообщений, принадлежащих единому процессу. Считается, что после успешного разрешения ошибочной ситуации она больше не повторится [1]. Та-

ким образом, описанная проблема сводится к интеллектуальной потоковой обработке данных в реальном времени. Подобные задачи в настоящее время в той или иной форме возникают при обработке финансовых данных, визуальных потоковых данных, документальных данных, сенсорных данных. Во многих работах рассматриваются автоматические методы извлечения семантики при потоковой обработке неструктурированных и структурированных данных [2,3], в том числе и адаптивные методы, основанные на нейросетевых моделях. Также в ряде работ описаны методы классификации динамически изменяющихся текстовых данных, основанных на модели ассоциативной семантической сети [4-6], а также с использованием нейронных сетей [7-10]. Авторами статьи для решения поставленной задачи предлагается использовать гибридную модификацию нечёткой нейронной сети Ванга-Менделя, состоящую из двух модулей: модуля извлечения семантики и модуля классификации.

1. ВЫБОР МОДЕЛИ КЛАССИФИКАТОРА

Исследуемый поток сообщений обладает следующими свойствами:

1. Все сообщения написаны на естественном языке – английском, таким образом, чтобы эксперт мог выявить семантический смысл сообщения.
2. Все сообщения имеют фиксированную длину в 100 символов, в связи с чем, могут быть сокращены в произвольном месте.
3. Все сообщения генерируются автоматически и наполняются данными по разным принципам, в зависимости от того, где возникла ошибка.

Солдатова Ольга Петровна, кандидат технических наук, доцент, доцент кафедры информационных систем и технологий. E-mail: op-soldatova@yandex.ru

Пудикова Екатерина Михайловна, магистр, старший инженер-программист компании NetCracker Technology Corp. E-mail: katherinmail@gmail.ru

Поэтому сообщение может быть как обычным текстовым описанием, так и содержать информацию о контексте возникновения ошибки.

4. Элементы потока практически не повторяются и изменяются во времени, в некоторых случаях сообщения могут иметь зависимость от времени их генерации (в системе наступают события, возникающие при работе компонентов, функционирующих по расписанию).

5. Поток имеет пять классов сообщений, определенных экспертами: *NC Defect*, *Other Defect*, *Manual Task*, *Info Task*, *Auto Task*.

6. Содержимое сообщений для различных классов может в значительной степени совпадать, но при этом, в них могут содержаться ключевые слова, указывающие на класс события.

Так как поток сообщений изменяется с течением времени, статический классификатор не справится с решением данной задачи. Поэтому классифицирующая система должна самостоятельно уметь накапливать опыт и на его основе выполнять классификацию. Подобным свойствам в настоящее время отвечают нейросетевые модели классификаторов.

При этом нейронная сеть должна обладать следующими свойствами:

1. Обеспечивать сравнительно высокую скорость обучения.

2. Обеспечивать высокую скорость отклика: длительность классификации должна быть менее 0,004 секунды.

3. Обеспечивать возможность классификации пересекающихся классов.

Перечисленным выше свойствам лучше всего соответствуют нечеткие нейронные сети, однако ни одна из известных на сегодняшний день моделей не может одновременно решать задачу извлечения семантики из текстовых данных и задачу классификации, поэтому авторами предложена новая гибридная модель нечеткой нейронной сети.

Нечеткие продукционные нейронные сети, основанные на моделях нечеткого вывода, имеют в своей структуре базу нечетких продукционных правил, которая может использоваться для решения задачи извлечения семантических знаний из входных данных. Кроме того, нечеткие продукционные нейронные сети хорошо справляются с задачей классификации нечетких множеств данных. В качестве базовой модели, была выбрана сеть Ванга-Менделя, так как из данного класса нейросетевых моделей она имеет наиболее простую структуру и способна решать не только задачу классификации, но и задачу кластеризации множества входных данных на основе алгоритма самоорганизации [11, 12].

2. ФОРМИРОВАНИЕ ОБУЧАЮЩЕГО МНОЖЕСТВА

В качестве примеров и соответствующих им семантических классов, определенных группой экспертов, были выбраны сообщения, сгенерированные реальным действующим промышленным программным комплексом. Поток сообщений составляет 1000 записей. Для решения задачи классификации с использованием машинного обучения из массива данных необходимо выделить множество сообщений, которые будут использоваться как обучающая выборка. Другие сообщения следует отнести к множеству тестовых данных, на которых будет проверяться работа модели.

Каждое сообщение предоставленного массива было обработано вручную с целью проверки корректности отнесения его к указанному классу и исключения противоречивых ситуаций классификации.

При формировании данного множества была выдвинута следующая идея формирования семантических правил вывода для классификации: обучаемая система сначала строит гипотезу или предположение о выводе, основываясь на статистических данных, а затем при опровержении выдвинутой гипотезы, правило вывода модифицируется так, что все предыдущие данные удовлетворяют новому правилу. Процесс модификации правила вывода будет происходить до тех пор, пока существуют сообщения, которые нельзя классифицировать данным правилом.

В процессе обработки был поставлен эксперимент, целью которого было определение возможности построения на основе данных правил вывода, необходимых для классификации сообщений. Идея формирования правил вывода была проверена экспериментально. В качестве обучающего множества были отобраны первые 200 элементов потока. При этом участвующие в эксперименте испытуемые должны были быть не знакомы с предметной областью данных, описанных в сообщениях, для того чтобы имитировать обучение системы с "пустого листа", ведь именно в этом состоянии находится любая обучаемая машина. Кроме того, в идеале, испытуемые должны быть плохо знакомы с английским языком, на котором написаны сообщения, чтобы они не могли извлечь дополнительный смысл, исходя из знаний языка. В эксперименте участвовало трое испытуемых, обладавших в той или иной мере данными качествами.

В результате эксперимента были получены первичные критерии полноты множества сообщений, достаточного для корректной классификации. Так же были исключены все противоречивые, а также неполные сообщения. Проведен-

ный в таких условиях эксперимент показал, что:

1. Для наискорейшего получения корректного правила классификации испытуемый должен сравнивать ответ с эталоном либо сразу же, либо через 2-3 итерации, большее число шагов осложняет процесс формирования правил.

2. Для увеличения скорости формирования корректных правил вывода испытуемому следует сначала подавать сообщения с наиболее четкими сигнатурами, которые легко классифицируются, и лишь потом подавать все более и более нечеткие примеры.

3. Испытуемый сначала строит статистически обусловленное правило классификации, на основе повторяющихся закономерностей, а затем формирует семантическое правило, на основе смысла, который извлекает из предъявленных сообщений. При этом если какой-либо класс представлен повторяющимися сообщениями одного типа, то семантическое и статистическое правила совпадают.

4. Неполные семантические правила в процессе обучения претерпевали реструктуризацию или уточнение.

Результаты эксперимента послужили основой генерации модели нейронной сети и методики построения обучающей выборки.

3. ГИБРИДНАЯ МОДИФИКАЦИЯ НЕЙРОННОЙ СЕТИ ВАНГА-МЕНДЕЛЯ

Из результатов эксперимента следует, что в изменяющихся сообщениях должны содержаться семантические инварианты, постоянные для всего потока. Классы-инварианты необходимо выявить и сохранить в структуре нейронной сети. Инвариант есть не что иное, как семантика, заключенная в группе ключевых слов сообщения.

В процессе обучения сети необходимо сформировать и формализовать набор ключевых семантических групп, на основе которых можно будет выполнять классификацию сообщений. Исходя из этого предположения, задачу классификации можно интерпретировать как процесс извлечения семантики из указанного набора данных в терминах конечного набора классов. Например, для сообщения "[ISP_T] Exception while call ServiceOrderStatus Stub for order [id=]" семантическое значение, которое может быть извлечено в рамках текущей задачи, есть класс *NC Defect*. Это означает, что глубина семантического смысла всегда ограничена пространством контекста, в рамках которого работает обучаемая машина. Модификация архитектуры сети Ванга-Менделя, представленная на рис. 1, позволяет решить данную задачу.

Сеть состоит из двух модулей, представляющих собой сеть Ванга-Менделя с несколькими выходами. Особенность первого модуля состоит в том, что число выходных нейронов в слое дефuzziфикации в начальный момент времени не определено, и должно быть сформировано в процессе обучения. Поэтому первый модуль является модулем самоорганизации, выходные нейроны которого на основе предпосылок правил вывода слоя фuzziфикации, описывают входное пространство данных наименьшим числом выходных нейронов. Первый модуль извлекает последовательность семантических инвариантов, характеризующих входное множество данных.

Второй модуль содержит фиксированное множество заранее определенных выходных классов, он восстанавливает отношения между семантически значимыми элементами одного вектора, для обеспечения ассоциативной связи между различными семантически значимыми элементами,

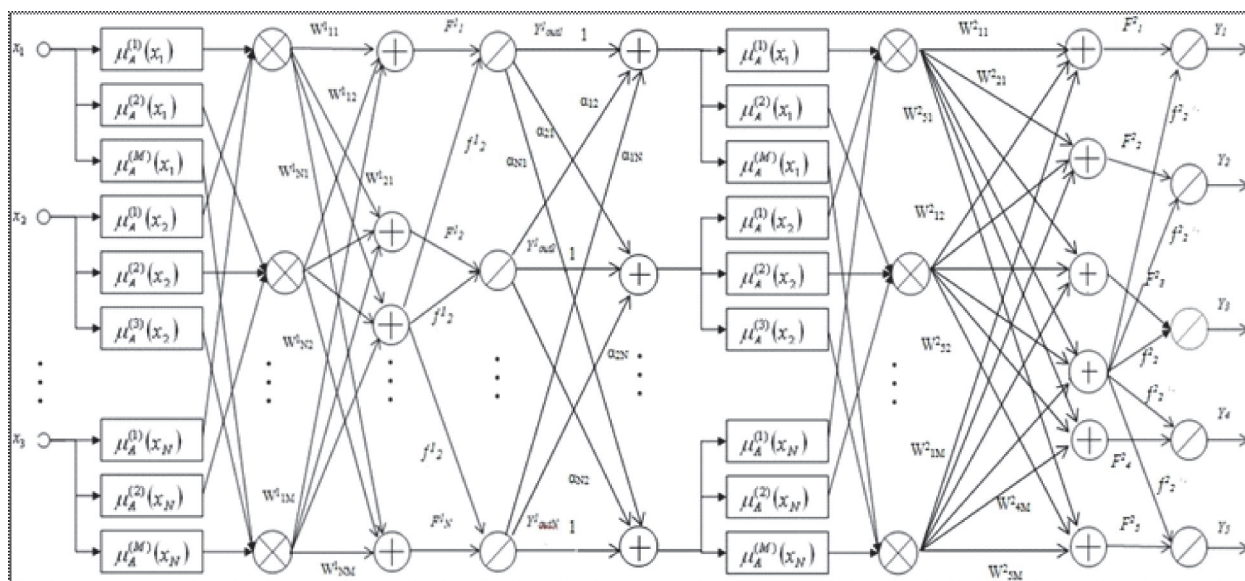


Рис. 1. Гибридная двухмодульная модификация сети Ванга-Менделя с ассоциативным слоем

поэтому в сети должен присутствовать слой создающий ассоциативные связи между модулями. Этот подход соответствует методу анализа текстовых данных, который основан на использовании ассоциативной модели построения текстов [4-6].

Алгоритм обучения данной сети значительно отличается от алгоритмов, обычно используемых для обучения подобных сетей.

4. АЛГОРИТМ ОБУЧЕНИЯ ГИБРИДНОЙ МОДИФИКАЦИИ СЕТИ ВАНГА-МЕНДЕЛЯ

Описываемая модель нейронной сети работает с числовыми входными векторами. Так как в данной задаче целевыми данными являются текстовые сообщения, то отображение их в числовой вектор можно реализовать, поставив в соответствие каждому слову сообщения его числовое значение. Таким образом, задача кодирования сводится к построению хэш-функции, которая будет осуществлять такое отображение. В данной работе была разработана хэш-функция, которая учитывает родственность слов по морфологическим признакам и отвечает следующим требованиям:

1. Каждое слово представлено числом от 0 до 1.
2. Хэш-функция не порождает коллизий.
3. Отображение в числовое пространство однокоренных или морфологически сходных слов генерирует близкие числовые значения.

Алгоритм вычисления хэш-функции, состоит из следующих шагов:

1. Из всего множества слов сообщений выбирается самое длинное и сохраняется в качестве эталона. Это слово кодируется в числовое представление с помощью кодировки UTF32.
2. Кодированное слово сообщения также получает числовое значение с помощью кодировки UTF32.
3. Код слова сравнивается с выбранным эталоном. Мерой близости закодированных слов является евклидово расстояние.
4. Полученное евклидово расстояние принимается в качестве кода слова.
5. Выбирается максимальное значение кода и все полученные коды нормируются на его величину.

Особенность данного типа отображения заключается в том, что все коды сосредоточены в числовом пространстве $M\{0..1\}$. Коды определяются в основном числом букв, входящих в слово. Чем больше их в последовательности, тем меньше расстояние между текущим словом и “словом – центром”. Однокоренные слова находятся на числовой плоскости рядом, но расстояния между словами одинаковой длины меньше, чем расстояния между однокоренными словами, имеющими разную длину.

Алгоритм обучения сети является результатом синтеза метода обратного распространения ошибки и принципа WTA (“победитель получает всё”), использующегося для поощрения нейронов-победителей при обучении на основе самоорганизации [8,12].

Алгоритм состоит из следующей последовательности действий:

1. Создается первый нейрон слоя дефuzziфикации и его веса W_{1j} и веса α_{1l} , обозначающие степень связанности данного нейрона с его соседями, инициализируются произвольными значениями в диапазоне $\{0;0,05\}$.

2. На вход слоя фуззификации первого модуля, последовательно подаются входные векторы, преобразованные функциями формирования предпосылок правил вывода. Выходные сигналы генерируются выходными нейронами первого модуля в соответствии с (1):

$$y_t = f(x) = \frac{\sum_{i=1}^M w_{ti} [\prod_{j=1}^N \exp[-(\frac{x_j - c_j^{(i)}}{\sigma_j})^{2b_j^{(i)}}]]}{\sum_{i=1}^M [\prod_{j=1}^N \exp[-(\frac{x_j - c_j^{(i)}}{\sigma_j})^{2b_j^{(i)}}]]}, \quad (1)$$

3. Рассчитывается результирующий сигнал каждого из связанных ассоциативных нейронов на основе выражения (2):

$$y_{out\ t} = y_{def\ t} + \beta \times \frac{\sum_{l=t}^N \alpha_{tl} \times y_{def\ l}}{\sum_{l=t}^N y_{def\ l}}, \quad (2)$$

где $y_{def\ t}$ – сигнал, сгенерированный текущим нейроном, β – это степень доверия текущего нейрона слоя дефuzziфикации к выходным сигналам соседей. Этот параметр определяет, какую максимальную долю от выходного сигнала может составлять суммарный выходной сигнал соседей. Параметр α_{tl} – это элементы матрицы весов связей выходных нейронов. В данном алгоритме самоорганизации считается, что текущий нейрон может быть связан с половиной всех выходных нейронов, формирующихся в процессе самоорганизации, которые являются ближайшими топологическими соседями текущего нейрона, $y_{def\ l}$ – это выходные сигналы нейронов-соседей, $y_{out\ t}$ – всегда принимает значения от $\{0;1\}$.

4. Среди всех сгенерированных модулем сигналов выбирается максимальный, и сравнивается с пороговым значением “узнавания”. В данной работе величина порога узнавания равна 0,2.

5. Если ни один нейрон не смог узнать входной вектор, то создается новый выходной нейрон, который должен научиться распознавать представителей нового класса. Для этого нейрон обучают с помощью метода обратного распростра-

нения ошибки до тех пор, пока не сможет распознавать данный вектор, то есть сгенерировать сигнал должен быть выше порога узнавания. Модификация параметров сети производится согласно выражениям (3) и (4):

$$\Delta_{w_{tk}}(n) = \eta_w (1 - y_{out\ t}(n)) \frac{\prod_{k=1}^{N_1} \mu_k^i(n)}{\sum_{i=1}^{M_1} \prod_{k=1}^{N_1} \mu_k^i(n)}, \quad (3)$$

$$\Delta_{\alpha_{ti}}(n) = \eta_{\alpha} (1 - y_{out\ t}(n)) \beta \frac{y_{defl\ i}(n)}{\sum_{i \neq t} y_{defl\ i}(n)}, \quad (4)$$

где η_{α} и η_w – коэффициенты обучения; $y_{out\ t}(n)$ – итоговые выходные сигналы нейронов слоя дефuzziфикации на n -ой итерации; $\mu_k^i(n)$ – функция принадлежности, использованная в слое фuzziфикации; β – это степень доверия текущего нейрона слоя дефuzziфикации к выходным сигналам соседей; N_1 – число элементов входного вектора; M_1 – число функций фuzziфикации в слое; y_{defl} – выходные сигналы нейронов-соседей. В процессе обучения формируется предположаемое правило вывода, которое может, как сохраниться, так и модифицироваться в процессе обучения. Эта фаза соответствует фазе формирования гипотез, которая наблюдалась в эксперименте, описанном выше.

6. Если порог узнавания преодолен, то считается, что нейрон-победитель распознал входной вектор и тогда только данный нейрон получает право на модификацию своих параметров в соответствии с выражениями (3) и (4).

Данный алгоритм позволяет сформировать столько независимых классов, что каждый из них отвечает своему набору правил вывода.

5. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Авторами статьи были проведены экспериментальные исследования, связанные с подбором параметров разработанной модели нейронной сети и изучением её свойств.

Одним из основных вопросов исследования является исследование эффективности разработанного алгоритма самоорганизации, использованного для обучения первого модуля. В резуль-

тате работы фазы самоорганизации, должны быть подобраны параметры сети для обеспечения максимальной разделимости классов минимальным числом нейронов. Функция ошибки рассчитывалась в соответствии с выражением (5):

$$E(n) = 1/2 \times \sum_{p=1}^{M_2} ((d_p(n) - y_p(n))^2). \quad (5)$$

Результаты исследования разделимости классов при различных значениях радиуса кластеризации первого модуля приведены в табл. 1.

На основании данных таблицы 1 можно сделать вывод, что, лучший радиус кластеризации, подходящий для решения поставленной задачи, это минимальное из приведённых в таблице значений. Однако, слишком малое значение радиуса кластеризации первого модуля снижает обобщающую способность сети. Так же замечено, что скорость уменьшения значения функции ошибки снижается после прохождения значения $\{r_1 = 0,055\}$, а значит, качественно иного результата распознавания будет сложно достичь. Исходя из этих наблюдений, для описанной сети было выбрано значение $\{r_1 = 0,055\}$.

Для второго модуля выбор значения радиуса определяется на основе предположения, что все отклики, генерируемые первым модулем, удалены друг от друга не более чем на $\sqrt{2} + \Delta$, так как генерируемые значения должны стремиться к формированию отклика вида: $(0,0,...,1,...,0)$. В данной работе значение радиуса второго модуля было выбрано равным $r_2 = 0,8$, как отвечающее степени перекрытия выходных сигналов первого модуля.

Были проведены исследования зависимости результатов классификации сообщений в процессе обучения. Результаты классификации представлены в таблице 2 для сети с параметрами $\{r_1 = 0,055; r_2 = 0,8; \eta_i = 0,1\}$.

В табл. 2 отражены несколько примеров того, как менялись результаты классификации сообщений (тексты сообщений приведены не полностью, но все тексты различны) после обучения первого модуля (фаза самоорганизации) и после полного обучения первого и второго модуля, а также соответствие классов сообщений эталонным классам экспертов. Среди приведённых примеров есть

Таблица 1. Зависимость величины ошибки и числа выходных нейронов первого слоя от выбора радиуса кластеризации первого модуля

Радиус	Число выходных нейронов первого модуля	Значение функции ошибки
0,1	37	2,09409460060899
0,08	46	1,72017292519478
0,06	68	0,933313440278266
0,055	77	0,108200676965054
0,03	129	0,0863023255360541

Таблица 2. Примеры результатов классификации сообщений кластерами первого модуля и после полного обучения

Номер кластера	Номер распознанного примера после фазы самоорганизации	Номер распознанного примера после полного обучения	Класс эксперта	Содержание сообщения
23	33	33	Manual Task	[BGC_B] Reject reason: Invalid Data – 'Horizontal
		51	Manual Task	Perform data VC provisioning on ATM backbone
		52	Manual Task	Perform voice VC provisioning on ATM backbone
		128	Manual Task	[BGC_B] Reject reason: Action type incorrect :
42	87	87	Other Defect	[BSS_T] Error invoking web service: http://miami
	199	199	Other Defect	[BGC_T] Error invoking web service: http://miami
28	51		Manual Task	Perform data VC provisioning on ATM backbone
	52		Manual Task	Perform voice VC provisioning on ATM backbone
	62	62	Manual Task	Contract the customer about PBX maintiner
		74	Manual Task	[EMPTY DESCR] User Action to do the certain
	79	79	Manual Task	Create and send "Provide" request to BGC
	80	80	Manual Task	[EMPTY DESCR] Trigger FNP Change In Mobific
	91	91	Manual Task	[BGC_B] VP and VCI for BGC Validate and BGC
	93	93	Other Defect	[BGC_B] BGC request for order:hasn't been sent.
	97	97	Manual Task	[BGC_B] Reject reason: VC Number not found
		101	Manual Task	[BGC_B] Review a new BGC Due Date and submit
		105	Manual Task	[BGC_B] Reject reason: SNA detected while order
		130	Manual Task	[BGC_B] Reject reason: Another request is pending

класс с номером 42, примеры которого не менялись в процессе обучения, а также классы, которые "перетягивали" сообщения. Так в начале процесса обучения класс 28 распознавал сообщения 51 и 52, но под конец обучения данные сообщения стал распознавать класс 23. Так же в приведенных данных видно, что кластеры стремятся "притянуть" сообщения с наиболее похожей структурой и содержанием, что и является целью построения первого модуля самоорганизации.

Также в работе исследовалась зависимость скорости уменьшения значения СКО от коэффициента обучения сети η_l . СКО рассчитывалось по формуле (6).

$$E(n) = 1/2 \times \sum_{p=1}^{M_2} ((d_p(n) - y_p(n))^2). \quad (6)$$

В результате исследований было определено, что наиболее оптимальное значение коэффициента обучения $\eta_l = 0,1$. Уменьшение значения не приводило к увеличению точности, но сильно замедляло процесс обучения сети, а увеличение значения приводило к увеличению точности, однако приводило к значительным осцилляциям в конце процесса обучения.

Процесс обучения приведён на графике зависимости значения СКО от итерации обучения с параметрами конфигурации: $\{r_1 = 0,055; r_2 = 0,8; \eta_l = 0,1\}$, представленном на рис. 2.

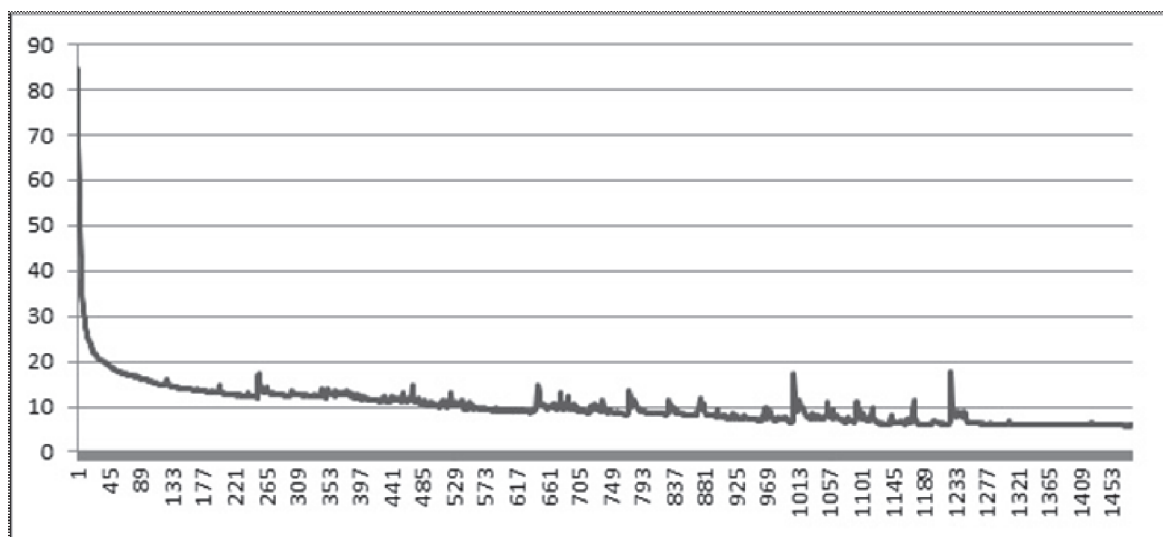


Рис. 2. Зависимость значения СКО от номера итерации

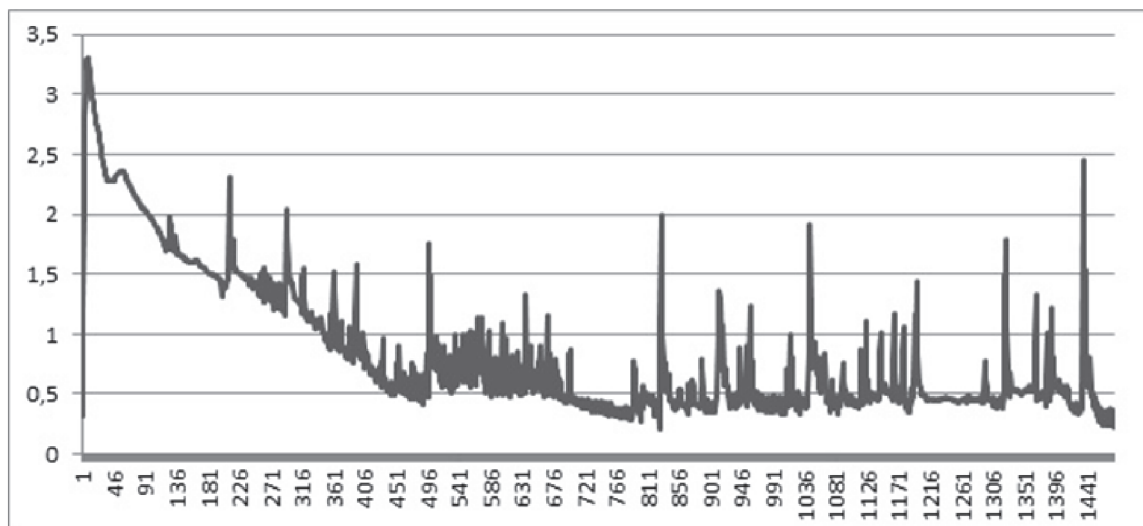


Рис. 3. Зависимость величины ошибки классификации в зависимости от номера итерации

Главной целью обучения является минимизация непосредственно ошибки классификации - отклонения фактического сигнала от эталонного отклика сети, рассчитываемой согласно выражению (5). График зависимости ошибки от номера итерации для сети с параметрами $\{r_1 = 0,055; r_2 = 0,8; \eta_i = 0,1\}$ представлен на рисунке 3.

Особенностью процесса обучения оказался факт, что существует несколько сообщений в обучающей выборке, которые получают некорректные значения классов, даже после полного обучения. Этот факт объясняет, почему процесс обучения не может достичь ожидаемой точности значения для функции ошибки и не приближается к значению 0, а задерживался на уровне 0,1 при конфигурации сети $\{r_1 = 0,055; r_2 = 0,8; \eta_i = 0,1\}$. При ближайшем рассмотрении обучающей выборки оказалось, что она содержит сообщения, которые получают другое значение класса, при сохранении

своей структуры. Использованный алгоритм обучения не поддерживает обработки такого типа действий, и для него это выглядит как шум в обучающих данных.

ЗАКЛЮЧЕНИЕ

После окончания процесса обучения сеть была протестирована на тестовом множестве данных, которое состояло из 202 сообщений, включающих сообщения, не входившие в обучающую выборку, и составляющие новый класс *RC Unknown*.

В результате процесса тестирования сделано заключение, что разработанная модель нейронной сети позволяет классифицировать данные тестового множества с точностью 71%. Скорость формирования отклика сети составила в среднем 0,002 секунды (вычисления выполнялись на одном процессорном ядре с частотой 2 ГГц), что отвечает требованиям по производительности, указанным в постановке задачи.

Одной из причин достижения полученного уровня точности классификации стало то, что в рассмотренной реализации модели не применялся алгоритм дообучения сети, который позволил бы адаптировать систему к распознаванию новых сообщений входного потока. Еще одна причина – зашумленность выборки, вследствие чего сообщения одного и того же типа получают различные значения классов. Данная ситуация свидетельствует о неполноте данных в обучающей и тестовой выборках.

СПИСОК ЛИТЕРАТУРЫ

1. The 8 Requirements of Real-Time Stream Processing / *M. Stonebraker, U. Zetintemel, S. Zdonik* // SIGMOD Record, Vol. 34, No. 4, Dec. 2005. P.42-47.
2. Технологии извлечения знаний из текста / *В. Рябышкин, С. Танков, С. Киселев, Н. Ильин* // Открытые системы. 2006. № 6. С.77-86.
3. Как работают новые Яндекс.Новости [Электронный ресурс] / *И. Сегалович, М. Маслов, Д. Нагорнов*. URL: http://vizaviz.blogspot.ru/2004/10/blog-post_26.html (дата обращения: 01.05.2013).
4. *Ермаков А.Е., Плешко В.В.* Ассоциативная модель порождения текста в задаче классификации // Информационные технологии. 2000. №12. С. 30-35.
5. Технология обработки текстовой информации с опорой на семантическое представление на основе иерархических структур из динамических нейронных сетей, управляемых механизмом внимания / *А.А. Харламов, А.Е. Ермаков, Д.М. Кузнецов* // Информационные технологии. 1998. №2. С. 26-32.
6. *Ермаков А.Е.* Компьютерная лингвистика и интеллектуальные технологии // Труды международного семинара Диалог'2002. Т.2. "Прикладные проблемы". М.: Наука, 2002. С. 180-185.
7. *Харламов А.А.* Автоматический структурный анализ текстов // Открытые системы. 2002. № 10. С. 62-65.
8. *Хайкин С.* Нейронные сети: полный курс. 2-е издание [пер с англ.]. М.: Издательский дом "Вильямс", 2006. 1104 с.
9. *Поляков П.Ю., Плешко В.В.* RCO (Russian Context Optimizer) на РОМИП 2008 // Российский семинар по оценке методов информационного поиска. Труды РОМИП 2007-2008. Санкт-Петербург: НУ ЦСИ, 2008. С.96-107.
10. *Поляков П.Ю., В.В. Плешко, А.Е. Ермаков.* RCO (Russian Context Optimizer) на РОМИП 2009 // Российский семинар по оценке методов информационного поиска. Труды РОМИП 2009. Санкт-Петербург: НУ ЦСИ, 2009. С.122-134.
11. Нечеткие модели и сети / *В.В. Борисов, В.В. Круглов, А.С. Федюлов*. М.: Горячая линия – Телеком, 2007. 284 с.
12. *Осовский С.* Нейронные сети для обработки информации [пер. с польского И.Д. Рудинского.]. М.: Финансы и статистика, 2002. 344 с.

CLASSIFICATION FLOW SYSTEM ERRORS WITH A HYBRID MODIFICATION OF NEURAL NETWORK WANG-MENDEL

© 2014 O.P. Soldatova¹, E.M. Pudikova²

¹ Samara State Aerospace University named after Academician S.P. Korolev
(National Research University)

² NetCracker Technology Corp, Samara

The article explores the application of the proposed authors hybrid modification of neural fuzzy multiple outputs Wang-Mendel network to solve a problem of error messages flow classification in the software system. Describes the properties of the message flow represented in natural language, methodology of training set formation, which based on real error messages stream. Describes the network learning algorithms and the efficiency of hybrid network for solving the problem.

Keywords: message flows in natural language training set, fuzzy neural production network, Wang-Mendel, self-organizing layer, learning algorithm classification.

Soldatova Olga Petrovna, Candidate of Technics, Associate Professor, Associate Professor of Information Systems and Technologies Department. E-mail: op-soldatova@yandex.ru
Pudikova Ekaterina Mikhailovna, Master, Senior Software Engineer at NetCracker Technology Corp. E-mail: katherinmail@gmail.ru