

ИНТЕГРАЦИЯ НЕЧЕТКИХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

© 2010 Н.Г.Ярушкина¹, И.Г.Перфильева², Т.В.Афанасьева¹

¹ Ульяновский государственный технический университет

² Университет г. Острава (Чешская республика)

Поступила в редакцию 14.05.2010

Работа посвящена актуальной задаче извлечения новых знаний о поведении временных рядов.
Ключевые слова: нечеткие модели, временные ряды.

Растущий интерес к нечеткому моделированию временных рядов обусловлен тем, что оно свободно от предположений, требуемых в статистическом подходе, позволяет строить нелинейные модели, обеспечивая снижение требований к математической подготовке персонала и делая современные технологии доступными для широкого круга пользователей. При этом существующие нечеткие модели и методы, так же как и количественные ставят своей единственной целью увеличение точности числового прогноза в максимальной степени. В тоже время в задачах управления и принятия решений большую значимость имеют знания экспертов о тенденциях развития процессов. Такие тенденции относятся к классу нечетких, так как выражаются оценочными суждениями эксперта, которые обладают внутренней нечеткостью. Данная работа посвящена методу нечеткого моделирования временных рядов, который позволяет извлекать из временных рядов контекстно-зависимые знания о нечетких тенденциях, выражать их в лингвистической форме и использовать для получения числового прогноза.

Системы, использующие нечеткие модели, базы знаний для хранения экспертных знаний о состоянии и функционировании процессов выраженных в виде оценочных и причинно-следственных лингвистических выражений, обеспечили новый виток в развитии экспертных систем и систем поддержки принятия решений. Однако, работа с экспертами по извлечению зна-

ний о тенденциях во временных рядах трудоемка и дорогостояща.

Поэтому при создании нечетких моделей ВР перспективным является применение интеллектуальных методов направления *Times-Series Data Mining*, в котором нечеткое моделирование ВР позволяет извлекать дополнительные знания в виде легко интерпретируемых продукционных правил. Нечеткое моделирование и представление ВР в виде нечеткого временного ряда (НВР) активно развивается в трудах иностранных ученых, таких как Х. Танака, Келминс и Диамонд, К. Сонг, К. Хирота, Я. Капржик, В. Педрич, В. Новак, И. Перфильева, И. Батыршин, а также в работах отечественных ученых – Н. Ярушкиной, С. Ковалева, К. Дегтярева. Несмотря на перечисленные работы, ряд научных проблем нечеткого моделирования ВР в новом направлении *Times-Series Data Mining* остаются нерешенными, в частности проблемы моделирования, анализа и прогноза развития такого объекта временного ряда, как нечеткая тенденция.

Здесь для постановки задачи исследований выбран временной ряд в виде последовательности упорядоченных в моменты времени пар $Y = \{t_i, x_i\}$, таких что $x_i \in B, t_i \in Bt, i \in [1, n]$. Требуется построить нечеткую модель временного ряда, позволяющую извлекать знания о тенденциях и правилах их следования, образующую базу знаний $Rule = \{R_k^p\}$, где k – количество нечетких продукционных правил в базе знаний, p – количество переменных (нечетких высказываний) в каждом правиле:

R_k^p : Если x_1 есть \tilde{x}_1 и x_2 есть \tilde{x}_2 и ... x_p есть \tilde{x}_p
то y есть \tilde{y}_i .

Метод нечеткого моделирования заключается в построении для исследуемого ВР контекстно-зависимой параметрической ACL-шкалы [1], генерации базы знаний нечеткой модели ВР, со-

Ярушкина Надежда Глебовна, доктор технических наук, профессор, проректор по научной работе, заведующая кафедрой "Информационные системы".

E-mail: jng@ulstu.ru

Перфильева Ирина Григорьевна, кандидат физико-математических наук, профессор.

E-mail: Irina.Perfilieva@osu.cz

Афанасьева Татьяна Васильевна, кандидат технических наук, доцент. E-mail: tv.afanaseva@mail.ru

держателю отношения между нечеткими элементарными тенденциями [2] временного ряда, и поиск такого параметра p базы правил $Rule = \{R_k^p\}$, для которого ошибка прогноза минимальна.

В основу моделирования поведения ВР в терминах нечетких тенденций положим предположение о развитии поведения ВР как результата зависимости нечеткой тенденции от значений нечетких тенденций в предыдущие моменты времени.

Предлагаются три нечеткие модели временного ряда, выражающие эту зависимость через компоненты модели элементарной нечеткой тенденции. Результатами прогноза ВР являются параметры нечеткой элементарной тенденции τ , выражающие тип $\tilde{\nu}$, интенсивность $\tilde{\alpha}$, степень принадлежности μ будущих изменений и числовая оценка уровней ВР, полученные по наилучшей модели.

В основе нечеткой модели временного ряда F2S лежит нечеткая модель элементарной тенденции $\tau = (\mu, \tilde{\nu}, \tilde{\alpha})$. Тогда нечеткая модель ВР порядка p , выраженная через компоненты модели элементарной тенденции представима следующим образом:

$$\begin{aligned} \tilde{\nu}(t-p) \circ R_{\tilde{\nu}}(t, t-p) &\rightarrow \tilde{\nu}(t), \\ \tilde{\alpha}(t-p) \circ R_{\tilde{\alpha}}(t, t-p) &\rightarrow \tilde{\alpha}(t). \end{aligned}$$

Результатом моделирования ВР является база правил $\{R(t, t-p)\}$ нечеткой системы логического вывода, выявляющая нечеткую зависимость следования отдельных компонент модели элементарной тенденции, извлеченная из нечеткого ВР, пример которой с частотной характеристикой каждого правила для $p=1$ представлен в табл. 1.

Данная модель временного ряда F1N основана на представлении временного ряда в виде числового ВР дефазифицированных интенсивностей нечеткой элементарной тенденции:

$$\tau = (\mu, n, \alpha), \quad n = DeFuzzy(\tilde{\nu}), \quad \alpha = DeFuzzy(\tilde{\alpha}).$$

Для прогнозирования значений этого ряда удобно использовать трехслойную нейросетевую модель, с количеством входов, равным порядку модели p и одним выходом:

$$\alpha' = F\left(\sum_{i=1}^p w_i \alpha_i\right),$$

где $(\alpha_1, \alpha_2, \dots, \alpha_p)^T$ – вектор входных интенсивностей, (w_1, w_2, \dots, w_p) – вектор весовых коэффици-

циентов, F – оператор нелинейного преобразования, неявно выражающий зависимость выхода от входа.

Модель F3N1S использует представление нечеткой элементарной тенденции в виде $\tau = (\tilde{\alpha}, \mu_1, \mu_2, \mu_3)$, где μ_1, μ_2, μ_3 – числовые временные ряды степеней принадлежности (функции принадлежности) типов элементарных тенденций:

$$\tilde{\alpha}(t-p) \circ R_{\tilde{\alpha}}(t, t-p) \rightarrow \tilde{\alpha}(t)$$

$$\mu'_1 = F_1\left(\sum_{i=1}^p w_i \mu_{1i}\right), \quad \mu'_2 = F_2\left(\sum_{i=1}^p w_i \mu_{2i}\right),$$

$$\mu'_3 = F_3\left(\sum_{i=1}^p w_i \mu_{3i}\right), \quad r = \arg(\max(\mu'_1, \mu'_2, \mu'_3)),$$

где r – номер типа тенденции.

Далее проводится вычислительный эксперимент, целью которого является исследование продуктивности метода нечеткого моделирования ВР для генерации авторских моделей временных рядов, позволяющих извлекать знания о нечетких тенденциях в форме продукционных правил “IF-THEN” и использовать их для прогнозирования ВР.

Для выбора наилучшей из трех предлагаемых нечетких моделей ВР в терминах нечетких тенденций используются следующие критерии качества:

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t^0 - y_t}{y_t} \right| \cdot 100\%,$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^0)^2,$$

$$T Tend = (ErrorCount / TotalCount) * 100\%.$$

$MAPE$ – средняя абсолютная процентная ошибка. Данный критерий удобно использовать для сравнения продуктивности метода при его применении на множестве временных рядов.

MSE – средняя квадратичная ошибка, традиционно используется для выбора наилучшей модели временных рядов из множества конкурирующих. Иногда применяют корень квадратный из MSE . Такой показатель качества обозначают $RMSE$.

$T Tend$ – коэффициент ошибок в обнаружении типов нечетких элементарных тенденций вре-

Таблица 1. Таблица извлеченных правил для типов тенденций

Правило	Частота
if ((Input0 is Падение)) then (Output is Рост)	3
if ((Input0 is Рост)) then (Output is Рост)	1
if ((Input0 is Рост)) then (Output is Падение)	2
if ((Input0 is Падение)) then (Output is Падение)	2

менного ряда в процентах, где *ErrorCount* – количество ошибок в моделировании типов нечетких элементарных тенденций, *TotalCount* – общее количество нечетких элементарных тенденции временного ряда.

Вычислительный эксперимент на нестационарных временных рядах показал, что авторские нечеткие модели ВР в терминах нечетких тенденций имеют недостаточно хорошие показатели качества для среднесрочного и долгосрочного прогноза за счет обработки “сырых данных” без выделения тренда. В то же время для краткосрочного прогноза авторские модели показали неплохие результаты, как по критериям качества моделирования и оценки прогноза нечетких тенденций, так и для оценки числовых уровней ВР.

В эксперименте были использованы авторские нечеткие модели тенденций *F2S(X, Y, Z)*, *F1N(X, Y, Z)*, *F3N1S(X, Y, Z)* и для сравнения нечеткие модели *S(X, Y, Z)* [3], *D(X, Y, Z)* [4]. В этих моделях *X* обозначает мощность генерируемой АСЛ-шкалы, *Y* – порядок модели, *Z* – глубина

прогноза (количество интервалов для прогноза). В некоторых моделях будет использоваться авторский алгоритм отбора нечетких правил, где к модели добавляется текст “+ отбор”.

Метод нечеткого моделирования и анализа тенденций временных рядов применяется для моделирования и прогнозирования временного ряда статьи 211 по данным бухгалтерской отчетности “Доходы и расходы” бюджетной организации. Данные, образующий ВР, характеризуют месячные расходы на заработную плату за четыре с половиной года (рис. 1). Для этого нестационарного ВР проводился прогноз на один интервал сначала (а) для усеченного ВР на 3 точки, (б) затем – на две точки, (в) затем на одну точку, в заключении (с) исследован весь временной ряд. Целью такого исследования одного и того же временного ряда – проанализировать устойчивость краткосрочного прогноза авторских нечетких моделей по сравнению с известными S- и D-моделями. Результаты лучших моделей приведены в табл. 2.

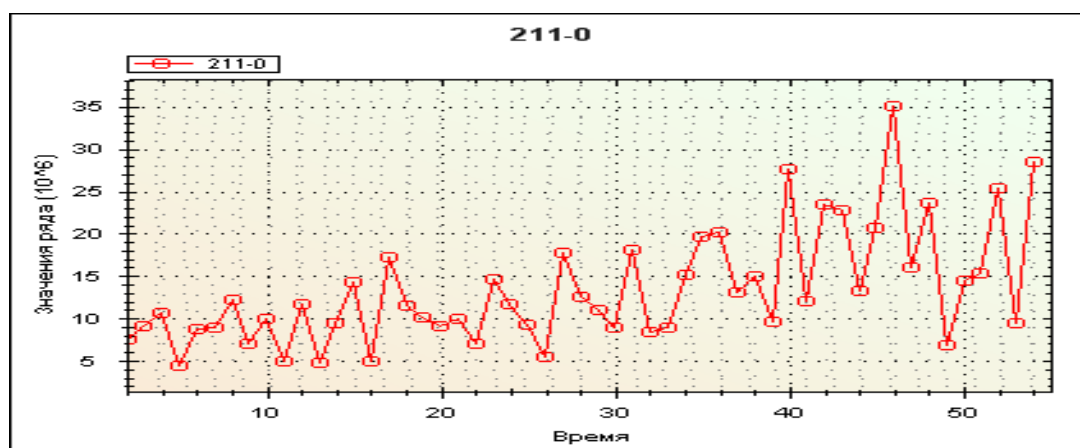


Рис. 1. Временной ряд статьи 211

Таблица 2. Показатели внешних мер качества. Прогноз на один интервал

	Название модели	MAPE	TTend	MSE
(а)	F2S(10,1,1)	5,4	0	0,69*10 ¹²
	S(10,1,1)	0,6	50	0,8*10 ¹⁰
	D(10,3,1)	9,4	0	0,2*10 ¹³
(б)	F2S(10,1,1)	8,2	0	0,43*10 ¹³
	S(10,3,1)	22	0	0,3*10 ¹⁴
	D(10,1,1)	42	50	0,1*10 ¹⁵
(в)	FIN(10,1,1)	5,9	0	0,1*10 ¹³
	S(10,4,1)	17	0	0,28*10 ¹³
	D(10,1,1)	52	0	0,24*10 ¹⁵
(с)	FIN(10,2,1)	4	0	0,1*10 ¹³
	S(10,4,1)	30	0	0,78*10 ¹⁴
	D(10,1,1)	29	0	0,7*10 ¹⁴

Таблица 3. Показатели внешних мер качества. Прогноз на три интервала

Название модели	MAPE	RMSE	TTend
F1N(3,3,3)+отбор	28	$70 \cdot 10^5$	16,6
F3N1S(3,3,3) + отбор	36	$122 \cdot 10^5$	16,6
F2S(3,4,3) + отбор	42	$80 \cdot 10^5$	16,6
SARIMA(1,1,1)x(1,0,0)	49	$88 \cdot 10^5$	50
F3N1S(3,4,3) + отбор	49	$101 \cdot 10^5$	0
F2S(3,1,3) + отбор	55	$88 \cdot 10^5$	0

Анализ внешних мер качества показывает, что авторские модели при прогнозе на один интервал характеризуются стабильными показателями качества (MAPE < 10%, TTend=0), что нельзя сказать о нечетких S- и D- моделях.

Проведенное статистическое моделирование и прогнозирование данного ВР на три интервала в пакете “Статистика” показало, что наименьшее MSE соответствует модели SARIMA(1,1,1)x(1,0,0), для которой MAPE=49%, RMSE= $88 \cdot 10^5$. При нечетком моделировании предлагаемым методом были получены следующие результаты, приведенные в табл. 3.

Таким образом, исследование предлагаемого метода нечеткого моделирования и анализа тенденций временных рядов позволяет сделать вывод о его полезности при решении задач извлечения знаний о правилах следования нечетких

тенденций и краткосрочного прогноза временных рядов.

СПИСОК ЛИТЕРАТУРЫ

1. *Афанасьева Т.В.* Модель ACL-шкалы для генерации лингвистических оценок в принятии решений // Вопросы современной науки и практики. Университет им. В.И.Вернадского. Т.2. Серия “Технические науки”. 2008. № 4 (14). С. 91-97.
2. Интегральный метод принятия решений и анализа нечетких временных рядов/*Новак В., Перфильева И.Г., Ярушклина Н.Г., Афанасьева Т.В.* // Программные продукты и системы. 2008. № 4. С. 65-68.
3. *Song Q.* Fuzzy time series and its models/Song, Q. Chissom B. // Fuzzy Sets and Systems, 54 (1993). P. 269–277.
4. *Дегтярев К.Ю.* Применение специализированных компьютерных программ и методов, основанных на нечетких временных рядах для краткосрочного прогнозирования USB/RUB котировок [Электронный ресурс] URL: <http://Exponenta.ru> (дата обращения 16.02.2010)

FUZZY MODELS' INTEGRATION IN TIME SERIES ANALYSIS

© 2010 N.G. Yarushkina¹, I.G. Perfilieva², T.V. Afanaseva¹

¹Ulyanovsk State Technical University

²University of Ostrava, Czech Republic

In this article we described the fuzzy models for Time Series Data Mining.

Key words: fuzzy models, time series.

Nadezhda Yarushkina, Doctor of Technics, Professor, Vice-Rector of Scientific Work, Head at the Information Systems Department. E-mail: jng@ulstu.ru.

Irina Perfilieva, Candidate of Physics and Mathematics, Professor, E-mail: Irina.Perfilieva@osu.cz.

Tatjana Afanasieva, Candidate of Technics, Associate Professor, E-mail: tv.afanaseva@mail.ru