

УДК 683.03

Н.Г. Ярушкина, И.Г. Перфильева, Т.В. Афанасьева

## ИНТЕГРАЛЬНЫЙ МЕТОД НЕЧЕТКОГО МОДЕЛИРОВАНИЯ И АНАЛИЗА НЕЧЕТКИХ ТЕНДЕНЦИЙ<sup>1</sup>

**Ярушкина Надежда Глебовна**, проректор по научной работе, доктор технических наук, профессор, заведующая кафедрой «Информационные системы» Ульяновского государственного технического университета. Имеет статьи, монографии в области интеллектуального анализа данных. [e-mail: jng@ulstu.ru].

**Перфильева Ирина Григорьевна**, кандидат физико-математических наук, профессор Остравского университета, Чехия. Имеет статьи, монографию в области интеллектуального анализа временных рядов. [e-mail: Irina.Perfilieva@osu.cz].

**Афанасьева Татьяна Васильевна**, кандидат технических наук, докторант, доцент кафедры «Прикладная математика и информатика» Ульяновского государственного технического университета. Имеет статьи, монографию в области интеллектуального анализа данных. [e-mail: tv.afanaseva@mail.ru].

### Аннотация

Рассматривается новый метод моделирования временных рядов (ВР), интегрирующий интеллектуальные методы решения задач извлечения знаний из временных рядов не только в числовой форме, но и в форме лингвистического описания уровней и элементарных тенденций.

Ключевые слова: нечеткая модель, временной ряд, нечеткая тенденция, извлечение знаний, прогнозирование.

### Abstract

The article deals with a new method of modeling of time series, which integrates intellectual methods of task solution concerning knowledge extraction from time series not only in numerical form but also in the form of linguistic description of levels and elementary tendencies.

Key words: fuzzy model, time series, fuzzy tendency, knowledge extraction, forecast.

В связи с бурным развитием информационных технологий практически во всех организациях и предприятиях активно используют хранение информации в базах данных (БД), в том числе и в виде временных рядов. Внедрение технологий баз данных привело к расширению круга пользователей БД, к которым относятся специалисты различного профиля: менеджеры среднего звена, аудиторы, специалисты в области контроля качества, экономисты, руководители, маркетологи, аналитики, политологи и т. д.

В связи с этим значительно повышается роль человека как поставщика и как пользователя информации в БД, который вносит свою долю неопределенности в оценку данных при их восприятии и использовании в решении профессиональных, обычно, нечетко сформулированных задач. Эта неопределенность в данных относится к неопределенности второго типа. В отличие от «стохастической» неопределенности первого типа она обусловлена нечеткостью восприятий и рассуждений человека в процессах

экспертного оценивания данных и моделируется нечетким множеством.

К таким сложным системам относятся социальные и организационно-технические системы: производственные предприятия, организации сферы коммерческих, информационных, телекоммуникационных, транспортных, образовательных, медицинских услуг, организации сферы связи, финансов, в которых экспертная качественная оценка состояний, динамики и оперативный прогноз тенденций развития процессов функционирования определяет эффективность принятия профессиональных решений.

Вместе с тем выполнение указанных аналитических процедур специалистами на основе программных продуктов, реализующих методологию стохастического моделирования временных рядов, сопряжено с объективными трудностями и, в первую очередь, с невозможностью учета неопределенностей второго типа — нечеткости в данных. Нечеткое моделирование и представление ВР в виде нечеткого времен-

ного ряда (НВР) активно развивается в трудах иностранных ученых, таких как К. Сонг, К. Чен, В. Новак, И. Перфильева, И. Батыршин, а также в работах отечественных ученых – Н. Ярушкиной, С. Ковалева, К. Дегтярева.

Несомненными достоинствами нечетких моделей НВР являются извлечение нелинейных зависимостей, моделирование коротких ВР и ВР, для которых затруднительно или невозможно построить адекватные стохастические модели, а также легко интерпретируемые результаты в виде продукционных правил, высокая степень автоматизации построения моделей и отсутствие высоких квалификационных требований к пользователям в области анализа временных рядов.

В этой статье предлагается интегральный метод нечеткого моделирования временных рядов, основанный на интеграции следующих моделей и методов:

- Нечеткое преобразование (F-преобразование) для выделения числовых компонент основной тенденции ВР [1].

- Нечеткая модель семантики части естественного языка (оценочных лингвистических выражений) для генерации лингвистических оценок уровней ВР и их изменений на основе ACL-шкалы [2].

- Модель нечеткого ВР [3].
- Модель нечеткой тенденции [4].

### 1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ F-ПРЕОБРАЗОВАНИЯ

Нечеткое сглаживание временных рядов на основе нечеткого преобразования (F-преобразования) – методика, разработанная И. Перфильевой [1], которая может быть отнесена к методикам нечеткого приближения. Нечеткое преобразование (F-преобразование) представлено для непрерывных функций и функций на ограниченном наборе точек.

Предположим, что функция  $f$  известна в точках  $p_1, \dots, p_N$  !  $w$ . Мы делим интервал  $w$  на множество равноудаленных узлов фиксированной длины  $x_k = v_L + h(k-1)$ ,  $k = 1, \dots, n$ ,

где  $N > n$ ,  $h = \frac{v_R - v_L}{n-1}$ .

F-преобразование имеет две фазы.

**Прямое F-преобразование.** Определим  $n$  базисных функций  $A_1, \dots, A_n$ , которые покрывают  $w$  и разделим их на  $n$  неясных областей. Базисная функция должна удовлетворять следующим условиям ( $k = 1, \dots, n$ ):

1.  $A_k : w \rightarrow [0, 1]$ ,  $A_k(x_k) = 1$ .

2.  $A_k(x) = 0$  if  $x \notin (x_{k-1}, x_{k+1})$ ,

где мы формально полагаем, что  $x_0 = x_1 = v_L$ ,  $x_{n+1} = x_n = v_R$ .

3.  $A_k(x)$  непрерывна.

4.  $A_k(x)$  монотонно возрастает на  $[x_{k-1}, x_k]$  и монотонно убывает на  $[x_k, x_{k+1}]$ .

5.  $\sum_{k=1}^n A_k(x) = 1$  для всех  $x \in w$ .

Используя базисные функции, мы преобразуем данную функцию  $f$  в кортеж из  $n$  действительных чисел  $[F_1, \dots, F_n]$ , определенных

$$F_k = \frac{\sum_{j=1}^N f(p_j) A_k(p_j)}{\sum_{j=1}^N A_k(p_j)}, \quad k = 1, \dots, n.$$

**Обратное F-преобразование.** Результатом прямого F-преобразования является вектор из чисел  $[F_1, \dots, F_n]$ . Это множество содержит информацию об оригинальной функции  $f$  и может быть использовано для получения функции

$$f_{F,n}(x) = \sum_{k=1}^n F_k \cdot A_k(x).$$

Можно доказать, что если  $n$  возрастает, тогда  $f_{F,n}(p_j)$  сходится к  $f(p_j)$ ,  $j = 1, \dots, N$ . Ясно, что функция  $f_{F,n}$  непрерывна.

F-преобразование имеет (кроме прочих) следующие свойства:

1. У него прекрасные фильтрующие свойства.
2. Его легко вычислять.

3. F-преобразование стабильно относительно выбора точек  $p_1, \dots, p_N$ . Это означает, что при выборе других точек  $p_k$  (и, возможно, изменяя их число  $N$ ) результирующая функция  $f_{F,n}$  значительно не меняется. Отметим, что это не верно для многих классических численных методов. Применение F-преобразования для прогнозирования ВР приведено в работе [5].

### 2 ACL-шкала для генерации лингвистических оценок уровней ВР и их изменений

Лингвистические оценки (ЛО) являются средством качественного оценивания и сравнения характеристик объектов и используются проектировщиками, менеджерами, лицами, принимающими решения (ЛПР), экспертами. Важным свойством лингвистических оценок является их широкое применение на практике в принятии решений для выражения знаний о степени соответствия наблюдаемого объекта или его характеристик некоторому объективному или субъективному критерию.

Ниже представлена специальная лингвистическая шкала в качестве инструмента как абсолютного, так и сравнительного нечеткого оценивания – ACL-шкала (Absolute & Comparative Linguistic) [2]. Абсолютные оценки, полученные по ACL-шкале, будут соответствовать нечетким меткам уровней ВР, а сравнительные оценки – нечетким тенденциям нечеткого ВР.

ACL-шкалу представим в виде алгебраической системы:

$$C = \{H, \Omega, \Psi\},$$

где  $H$  – множество объектов,

$$H = \{X, \tilde{X}, G, P, \tilde{V}, \tilde{A}\};$$

где  $X$  – универсальное множество объектов  $x$ , образующее область определения шкалы;

$\tilde{X}$  – базовое конечное терм-множество аб-

солютных нечетких оценок  $\tilde{x}_i \in \tilde{X}, i! [1, m]$ , например,  $\tilde{X} = \{\text{Плохой, Удовлетворительный, Хороший, Отличный и др.}\};$

$G$  – синтаксические правила вывода (порождения) цепочек оценочных высказываний (множеств производных термов  $\tilde{X}^*$ , не входящих в базовое терм-множество);

$P$  – семантические правила, определяющие функции принадлежности для каждого термина

$$\tilde{x}_i \in \tilde{X} \cup \tilde{X}^*, i! [1, m^*];$$

$\tilde{V}$  – нечеткая шкала для оценивания типов сравнительных оценок;

$\tilde{A}$  – нечеткая шкала для оценивания различия абсолютных оценок;

$\Omega$  – множество операций,  $\Omega = \{F_T, F_C, F_{Er}\}$ , на множестве  $H$ , включающее функционалы оценивания –  $F_T$ , функционалы вычисления оценок –  $F_C$  и функционалы вычисления погрешностей –  $F_{Er}$ ;

$\Psi$  – множество параметров,  $\Psi = \{E, d, MF, nmin, nmax\}$ ;

$E$  – тип нечеткой шкалы (номинальная, порядковая или «квазиинтервальная (равномерная/неравномерная)»);

$d$  – параметр, определяющий носитель типа изменения нечеткой градации «Стабильность», то есть длину интервала на  $X$ , все значения  $x$  в котором могут рассматриваться с позиции данной экспертной шкалы, как одинаковые, неразличимые;

$MF$  – тип функций принадлежности, модели-

рующих нечеткие оценки  $\tilde{x}_i \in \tilde{X}, i! [1, m]$ , например, треугольные;

$$nmin = \inf(X), nmax = \sup(X).$$

Использование единого базиса в виде ACL-шкалы для порождения абсолютных и соответствующих им сравнительных нечетких оценок позволит строить модели нечетких временных рядов и обнаруживать в них нечеткие тенденции [6].

### 3 Модель нечеткого временного ряда

Представление временных рядов в классе нечетких временных рядов основывается на предположении, что возможна семантически значи-

мая интерпретация значений временного ряда, основанная на понятии нечетких множеств.

**Определение 1** [6]. Нечетким временным рядом называют упорядоченную последовательность наблюдений над некоторым явлением, состояния которого изменяются во времени, если значение состояния в момент времени выражено с помощью нечеткой метки.

Нечеткая метка может быть сформирована непосредственно экспертом или получена на основе ACL-шкалы по исходному временному ряду. Во втором случае она связана с исходным значением временного ряда.

Пусть  $x_i \in R^l$  – универсум, на котором определены нечеткие множества  $y_i^j, i = 1, 2, \dots$ , и  $Y_t$  – коллекция  $y_i^j, i = 1, 2, \dots$ . Тогда  $Y_t, t = 1, 2, \dots$  называется нечетким ВР [3].

На практике в большинстве ВР последовательные наблюдения зависимы так, что:

$$R = \{(y_t, y_{t-1}), (y_{t-1}, y_{t-2}) \dots\} \subseteq Y_t \times Y_{t-1},$$

где  $Y_t, Y_{t-1}$  – переменные;

$y_t, y_{t-1}$  – наблюдаемые значения этих переменных.

Наиболее частой моделью зависимости является явная функция отображения:

$f: Y_{t-1} \rightarrow Y_t$ , представленная линейной функцией (марковским процессом, модель AR):

$$y_t = f(y_{t-1}, \phi, \varepsilon) = \phi y_{t-1} + \varepsilon,$$

где  $\varepsilon$  – случайная ошибка, шум.

В случае нечеткого ВР в качестве модели авторегрессии используется нечеткое разностное уравнение:

$$y_t^j = y_{t-1}^i \circ R_{ij}(t, t-1),$$

$$y_t^j \in Y_t, y_{t-1}^i \in Y_{t-1},$$

$i \in I, j \in J, \circ$  – max min композиция,

$$R(t, t-1) = \bigcup_{i,j} R_{ij}(t, t-1) \text{ – система нечет-$$

ких отношений, которая символически может быть записана в виде  $Y_t \rightarrow Y_{t-1}$ . Систему отношений  $R$  в выражении  $Y_t = Y_{t-1} \circ R(t, t-1)$  называют моделью нечеткого ВР первого порядка, данная модель – важный частный случай общей модели порядка  $p$ :

$$Y_t = (Y_{t-1} \times Y_{t-2} \times \dots \times Y_{t-p}) \circ R(t, t-p);$$

$$R(t, t-p) = \max_p \left\{ \min_{j, i_1, i_2, \dots, i_p} \{y_t^j, y_{t-1}^{i_1}, \dots, y_{t-p}^{i_p}\} \right\}.$$

Построение нечеткого временного ряда  $\tilde{Y} = \{t_i, \tilde{x}_i\}$  по исходному временному ряду

$Y = \{t_i, x_i\}$  с помощью ACL-шкалы, позволяющей генерировать ВР и оценивать изменения

(различия) между любыми нечеткими метками уровней ВР, обеспечивает необходимый формальный базис для оценки поведения НВР в терминах нечетких тенденций.

#### 4 Модель нечеткой тенденции

Нечеткой тенденцией (НТ) нечеткого временного ряда будем называть нечеткую метку, выражающую характер изменения последовательности нечетких значений НВР в заданном интервале времени [6]. Нечеткая тенденция определяет природу НВР не в аналитическом, а в лингвистическом виде. Формальным аппаратом для извлечения знаний из нечетких временных рядов, позволяющих представлять его поведение в терминах лингвистических оценок тенденций, является ACL-шкала и ее функциональное наполнение, связанное с генерацией сравнительных нечетких оценок изменений. Рассмотрим нечеткий временной ряд  $\tilde{Y} = \{\tilde{t}_i, \tilde{x}_i\}$ ,  $i \in [1, n]$ ,  $n$  – количество членов нечеткого ВР.

**Утверждение 1.** Каждая нечеткая тенденция

$\tau \in \mathcal{S}$  нечеткого временного ряда  $\tilde{Y} = \{\tilde{t}_i, \tilde{x}_i\}$  может быть представлена структурной моделью в виде кортежа отношения, построенного на декартовом произведении свойств нечеткой тенденции  $\tilde{V} \times \tilde{A} \times \Delta \tilde{T} \rightarrow \mathcal{S} : \tau = \langle \tilde{v}, \tilde{a}, \Delta \tilde{t}, \mu \rangle$ , где  $\tau$  – наименование нечеткой тенденции (идентификатор) из множества  $\mathcal{S}$ ,  $\tau \in \mathcal{S}$ ;

$\tilde{v}$  – тип нечеткой тенденции (тип изменений),  $\tilde{v} \in \tilde{V}$ ;

$\tilde{a}$  – степень изменения, то есть интенсивность нечеткой тенденции,  $\tilde{a} \in \tilde{A}$ ;

$\Delta \tilde{t}$  – продолжительность данной нечеткой тенденции,  $\Delta \tilde{t} \in \Delta \tilde{T}$ ;

$\mu$  – функция принадлежности участка нечеткого временного ряда, длительностью  $\Delta \tilde{t}$ , ограниченного нечеткой тенденцией  $\tau$ .

В том случае, если модель нечеткой тенденции определяется для двух соседних уровней НВР, то такую нечеткую тенденцию  $\tau$  отнесем к классу элементарных нечетких тенденций  $\tau$ . Содержательно, элементарная нечеткая тенденция моделирует изменения между двумя соседними значениями нечеткого временного ряда  $\tilde{x}_i, \tilde{x}_{i+1}$  и может быть сопоставлена с моментом времени  $t_i$  НВР.

Исходя из вышесказанного, при построении модели прогноза ВР в терминах нечетких тенденций интерес представляет получение прогнозных значений свойств нечетких элементарных тенденций.

В таком случае, уравнение модели нечеткой

элементарной тенденции в общем виде может быть записано по аналогии с моделью НВР:

$$\tau(t-p) \circ R(t, t-p) \rightarrow \tau(t)$$

или в декомпозированном виде по свойствам элементарной нечеткой тенденции:

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

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

#### 5 ОПИСАНИЕ ИНТЕГРАЛЬНОГО МЕТОДА НЕЧЕТКОГО МОДЕЛИРОВАНИЯ

Введем в рассмотрение функционал  $F\_trans\_1(Y)$ , где  $Y$  – исходный временной ряд, математическую основу которого представляет прямое F-преобразование, и функционал  $F\_trans\_2(F)$ , где  $F$  – компоненты дискретного тренда, полученные применением  $F\_trans\_1(Y)$ . Математическую основу функционала  $F\_trans\_2(F)$  представляет обратное F-преобразование.

Обозначим функционал  $ACL(Y)$ , реализующий построение ACL-шкалы по исходному ряду. Используем функционал  $Fuzzy(Y)$  для преобразования исходного временного ряда в нечеткий временной ряд путем оценивания исходного ВР по ACL-шкале, функционал  $ETend(\tilde{x}_i)$  – для преобразования нечеткого временного ряда во временной ряд элементарных тенденций и покомпонентного разложения временного ряда элементарных тенденций.

Обобщенный алгоритм интегрального метода нечеткого моделирования представим в виде последовательности функционалов:

1. Применение к исходному ВР F-преобразования для выделения основной тенденции и получения остатков  $R$ :

$$F = F\_trans\_1(Y), YI = F\_trans\_2(F), \\ R = Y - YI.$$

2. Фаззификация временных рядов  $YI$  с помощью ACL-шкалы:

$$\tilde{X} = ACL(Y), \tilde{x}_i = Fuzzy(YI), \tilde{x}_i \in \tilde{X}.$$

3. Построение временного ряда элементарных тенденций на основе НВР и ACL-шкалы:

$$\tau_i = ETend(\tilde{x}_i), \text{ где } \tau = \langle \tilde{v}, \tilde{a}, \Delta \tilde{t}, \mu \rangle.$$

4. Построение модели ВР для генерации числовых оценок прогноза  $F$ ,  $R$  и  $\hat{Y}$  по формуле линейной комбинации:

$$F_{k+1} = \alpha F_k + \beta F_{k-p};$$

$$R_{k+1} = \alpha R_k + \beta R_{k-p};$$

$$\hat{Y} = F_{k+p} + R_{k+p}.$$

5. Построение модели для прогноза основной тенденции и остатков в терминах элементарных

тенденций. При этом автоматически выполняется лингвистическая интерпретация

$$\begin{aligned} \tau &= (\mu, \tilde{v}, \tilde{\alpha}); \\ \tau(t-p) \circ R(t, t-p) &\rightarrow \tau(t); \\ \tilde{v}(t-p) \circ R_{\tilde{v}}(t, t-p) &\rightarrow \tilde{v}(t); \\ \tilde{\alpha}(t-p) \circ R_{\tilde{\alpha}}(t, t-p) &\rightarrow \tilde{\alpha}(t). \end{aligned}$$

6. Построение модели нечеткого временного ряда:

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

7. Представление результирующей многоуровневой модели временного ряда, полученной применением интегрального метода нечеткого моделирования:  $\langle \hat{Y}, \hat{v}, \hat{\alpha}, \hat{\mu}, \hat{x} \rangle$ ,

где  $\hat{Y}, \hat{v}, \hat{\alpha}, \hat{\mu}, \hat{x}$  — прогнозные значения числового уровня ВР, типа нечеткой тенденции, интенсивности нечеткой тенденции, степени принадлежности и нечеткого уровня ВР соответственно.

#### 6 Оценка эффективности интегрального метода нечеткого моделирования и анализа нечетких тенденций ВР

Характеристики информационной базы исследуемых временных рядов:

- |   |             |
|---|-------------|
| 1. Общее количество                                     | — 22.       |
| 2. Количество искусственных ВР                          | — 8 (36%).  |
| 3. Количество ВР экономических показателей              | — 14 (64%). |
| 4. Количество стационарных ВР                           | — 5 (22%).  |
| 5. Количество нестационарных ВР                         | — 17 (78%). |
| 6. Количество нестационарных ВР класса TS               | — 7 (32%).  |
| 7. Количество нестационарных ВР класса DS               | — 17 (46%). |
| 8. Количество ВР малой длины (от 7 до 21 значения)      | — 11 (50%). |
| 9. Количество ВР средней длины (от 22 до 43 значений)   | — 5 (22%).  |
| 10. Количество ВР большой длины (от 44 до 150 значений) | — 6 (28%).  |

Так, внутренние ошибки интегрального метода нечеткого моделирования ВР в распознавании нечетких тенденций в среднем составили 3,15%, что в 3 раза лучше признанных нечетких моделей и в 7 раз лучше нейросетевой модели.

Ошибка внешних лингвистических оценок типов нечетких тенденций при краткосрочном прогнозе в среднем составила 0%, что лучше признанных нечетких моделей на 33,3% и нейросетевой модели на 34%.

Ошибка числовых оценок уровней ВР интегрального метода нечеткого моделирования ВР в среднем по внутреннему критерию MAPE составила 1,1% от нейросетевых моделей и лучше

признанных нечетких моделей на 20%.

Ошибка числовых оценок уровней ВР интегрального метода нечеткого моделирования ВР при краткосрочном прогнозе по внешнему критерию MAPE составила 14,6%, что превосходит признанные нечеткие модели в 3 раза и нейросетевую модель — в 3,7 раза.

#### ЗАКЛЮЧЕНИЕ

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.

Разработанный интегральный метод применяется для решения задачи экспресс-анализа временных рядов экономических показателей. Он обеспечивает повышение информативности результатов экспресс-анализа за счет возможности генерации как числовых, так и лингвистических оценок тренда, а также лингвистических оценок динамики в целом временных рядов.

Однако метод, являясь универсальным методом прогнозистики, может быть применен к анализу состояния любого объекта, если оно может быть представлено временными рядами.

#### СПИСОК ЛИТЕРАТУРЫ

1. Перфильева И. Нечеткое преобразование / И. Перфильева // Нечеткая логика. — Амстердам, 2003. — С. 275–300.
2. Афанасьева Т. В. Модель ACL-шкалы для генерации лингвистических оценок в принятии решений / Т. В. Афанасьева // Вопросы современной науки и практики. Университет им. В. И. Вернадского. — 2008. — № 4 (2). — С. 91–96.
3. Ярушкина Н. Г. Современный интеллектуальный анализ нечетких временных рядов / Н. Г. Ярушкина // Труды V-й Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). В 2 т. Т. 1. — М. : Физматлит, 2009. — С. 19–29.
4. Афанасьева Т. В. F-преобразование в прогнозировании временных рядов / Т. В. Афанасьева, А. А. Ивахина, И. Г. Перфильева // Труды Международной конференции «ИННОВАТИКА 2009» (ДГУ, Махачкала 25-29 мая 2009 г.).
5. Афанасьева Т. В. Моделирование лингвистических оценок на основе ACL-шкалы / Т. В. Афанасьева, А. О. Козлов, А. А. Ивахина // Труды V-й Международной научно-практической конференции «Интегрированные модели и мягкие вычисления» (Коломна, 20-30 мая 2009 г.). В 2 т. Т.1. — М. : Физматлит, 2009. — С. 298–304.
6. Ярушкина Н. Г. Основы теории нечетких и гибридных систем / Н. Г. Ярушкина. — М. : Финансы и статистика, 2004.