

## КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ГИБРИДНОЙ СИСТЕМЫ АНАЛИЗА НЕЧЕТКИХ ВРЕМЕННЫХ РЯДОВ

Т.В. Афанасьева, С.Г. Валеев, Н.Г. Ярушкина

*ГОУ ВПО «Ульяновский государственный технический университет», г. Ульяновск*

*Рецензент Н.П. Пучков*

**Ключевые слова и фразы:** задачи Data Mining Time Series; нейронные сети; нечеткие временные ряды; нечеткие тенденции; обобщенный алгоритм; регрессионное моделирование; структурно-лингвистическая модель.

**Аннотация:** Обоснованы актуальность и проблема анализа нечетких временных рядов. Предложена формальная концептуальная модель программной гибридной системы анализа нечетких временных рядов, основанная на интеллектуальных технологиях и технологиях регрессионного моделирования.

**Введение.** В связи с ростом хранимых, упорядоченных во времени, данных о характеристиках объектов, процессов и систем в промышленности, экономике, медицине, образовании, социологии расширяются возможности анализа и прогнозирования количественных и качественных изменений этих характеристик. Систематическое и комплексное исследование тенденций развития объектов, процессов, систем на основе анализа временных рядов (**ВР**), извлеченных из хранилищ и баз данных, является сферой профессиональной деятельности лиц, принимающих решение (**ЛПР**) в задачах управления, планирования, проектирования, аудита, экспертизы. В процессе решения указанных задач ЛПР оперирует в большей степени нечеткими, зависящими от контекста терминами, выраженными лингвистическими оценками, позиционирование которых в фиксированные моменты времени образует временные ряды. Под контекстом будем понимать семантику, зависящую от пространственных, временных, проблемных и других параметров. Такие временные ряды, получившие название нечетких временных рядов (**НВР**), обладают специфическим свойст-

---

Афанасьева Т.В. – кандидат технических наук, доцент кафедры «Прикладная математика и информатика», докторант УлГТУ; Валеев С.Г. – доктор физико-математических наук, профессор, заведующий кафедрой «Прикладная математика и информатика» УлГТУ; Ярушкина Н.Г. – доктор технических наук, профессор кафедры «Информационные системы», проректор по научной деятельности УлГТУ, г. Ульяновск.

вом: в значениях уровней НВР отображена контекстно-зависимая лингвистическая трихотомия «ЗНАК-ЗНАЧЕНИЕ-ОБОЗНАЧЕНИЕ» в отличие от традиционной лингвистической дихотомии «ЗНАК-ЗНАЧЕНИЕ». Лингвистическая трихотомия расширяет семантическое пространство ЗНАКА за счет возможности использования четких и нечетких значений. Согласно концепции точности/неточности понятий Заде [1], каждое понятие, знак имеет содержание (*value*), которое может быть задано точно или неточно (*v-precise, v-imprecise*).

Термин «ЗНАЧЕНИЕ» применительно к НВР будем трактовать как точную числовую оценку уровня временного ряда (*v-precise*). Для указанного аспекта ЗНАКА разработаны эффективные методы анализа ВР. ОБОЗНАЧЕНИЕ – это лингвистическая оценка, выражающая уровень ВР в контекстно-зависимых терминах ЛПР (*v-imprecise*). Наиболее удобными математическими средствами моделирования и описания лингвистических оценок могут служить средства нечетких множеств и нечетких шкал.

Следует отметить ощутимый разрыв между возможностями, предоставляемыми традиционными инструментальными системами поддержки задач анализа ВР, и практическими потребностями ЛПР в автоматизации анализа тенденций нечетких временных рядов. Этот разрыв обусловлен, с одной стороны, ориентацией традиционных систем на использовании лингвистической дихотомии, а с другой, – недостаточным развитием формального описания моделей и методов анализа нечетких временных рядов.

Одним из способов преодоления указанной проблемы является разработка и развитие гибридных систем анализа НВР, в которых исходные данные, выходные результаты определены на понятном ЛПР языке, выражающем лингвистическую трихотомию в семантике состояний и изменений объектов предметной области, а модели и методы анализа и прогноза основываются на технологиях быстро развивающегося направления Time Series, Data Mining.

**Концептуальная модель гибридной системы анализа нечетких временных рядов.** Можно обозначить следующие особенности гибридной системы анализа (ГСА) нечетких временных рядов.

1. Ориентация в представлении входных и выходных данных на лингвистическую трихотомию «ЗНАК-ЗНАЧЕНИЕ-ОБОЗНАЧЕНИЕ» [2]: например, ЗНАК – это уровень временного ряда, ЗНАЧЕНИЕ – числовое значение уровня конкретного временного ряда, ОБОЗНАЧЕНИЕ – это нечеткая метка (нечеткое множество), соответствующая числовому уровню временного ряда. Для перехода от ЗНАКА к ЗНАЧЕНИЮ необходимы процедуры измерения, вычисления или извлечения, от ЗНАКА к ОБОЗНАЧЕНИЮ – процедуры построения нечеткой шкалы, от ЗНАЧЕНИЯ к ОБОЗНАЧЕНИЮ – процедуры оценки по нечеткой шкале (фаззификации).

2. Применение множеств нечетких лингвистических оценок, задающих семантику предметной области и ориентированных на ЛПР.

3. Использование нечетких лингвистических шкал для генерации нечеткой лингвистической входной и выходной оценок.

4. Вычисления для анализа и прогноза на основе интеллектуальных технологий и статистического (регрессионного) моделирования.

Указанные особенности определяются тем фактом, что целью создания гибридных систем анализа НВР является обработка и управление новыми типами данных, образованными сложными слабоформализованными структурами представления семантических понятий, для извлечения новых знаний о состоянии и тенденциях развития объектов предметной области.

Исходя из них, проектировщику ГСА НВР необходимо определить в первую очередь семантику системы, выраженную следующими стратами знаний.

1. Где-знания. Область применения системы, пространственные ограничения, условия, соотношение с другими системами.

2. Кто-знания. Пользователи системы, компетенции и требования пользователей.

3. Зачем-знания. Цель применения системы и совокупность решаемых задач.

4. Что-знания. Входные и выходные данные системы. Их типы, структуры, модели.

5. Как-знания. Функции, операции, алгоритмы, определяющие функционирование системы.

6. Почему-знания. Значимость, актуальность применения системы.

7. Когда-знания. Временные ограничения.

8. Сколько-знания. Количественные показатели, постоянные, затраты, показатели эффективности.

9. Какие-знания. Качественные оценки системы.

Заметим, что страты знаний могут быть укрупнены в такие объекты системы как вход–выход с возможными ограничениями (страты 4, 1, 2, 7), функции (страты 3, 5), критерии (страты 6–9). Эти укрупненные страты знаний могут быть взяты за основу при концептуальном проектировании ГСА НВР.

Рассмотрим в рамках концептуального моделирования абстрактной гибридной системы анализа нечетких временных рядов  $S$  определение ее модели, ориентированной на введенную выше лингвистическую трихотомию.

Обобщенная компонентная модель системы  $S$ , включающая вход  $Y$ , выход  $\hat{Y}$ , множество моделей преобразования входа в выход  $F$ , множество критериев  $K$ , используемых для идентификации модели, оценки ее параметров и критериев, определяющих качество моделей  $F$ , представима в виде  $Cs = \langle Y, \hat{Y}, F, K \rangle$ . Эта традиционная модель системы может иметь, по крайней мере, двухуровневое представление для задач в конкретной предметной области: на одном уровне – в терминах ЛПР, на другом – в терминах математических моделей и методов.

Каждая компонента модели  $Cs$  характеризуется своей семантикой, имеющей внутреннюю и внешнюю интерпретации. Внешняя интерпретация семантики задает интенсификацию модели  $S$  в форме лингвистического представления каждой компоненты, выраженной в терминах ЛПР, а внутренняя – определяет экстенсификацию модели  $S$  в виде правил выполнения допустимых операций для каждой компоненты, совокупности отношений и

математических зависимостей между компонентами модели. Таким образом, внутренняя семантика определяется моделями решения задач анализа НВР и используется для генерации «ЗНАЧЕНИЙ», а внешняя семантика задает субъектно-объектное отображение предметной области и объектов известных исследователю формальных систем в компоненты модели  $S$  и определяет «ОБОЗНАЧЕНИЕ».

Определим для модели  $Cs$  системы  $S$  функционал  $R_s = R(Y, \hat{Y}, F, K)$ , задающий внутреннюю семантику, и функционал  $P_s = P(Y, \hat{Y}, F, K)$  – для представления внешней семантики.

Тогда концептуальная модель абстрактной гибридной системы анализа НВР  $S$  может быть определена в виде  $M_s = \langle Cs, R_s, P_s \rangle$ , и в этой модели представлена лингвистическая трихотомия «ЗНАК-ЗНАЧЕНИЕ-ОБОЗНАЧЕНИЕ». Сопоставление компонент и семантических моделей системы  $S$  при анализе конкретного класса НВР обычно осуществляет проектировщик системы, принимающий наиболее адекватное решение по контексту среды и своего опыта.

Рассмотрим подход к формализации функционала  $P_s$ , представляющего внешнюю семантику системы  $S$ , так как внутренняя семантика компонент системы  $S$  определяется внешней семантикой.

Временной ряд  $Y$  зададим бинарным отношением, являющимся подмножеством декартового произведения, и определим структурную модель объекта. Временной ряд  $Y = \langle T, X \rangle$ , где  $T$  – атрибут (домен) времени, задаваемый упорядоченными по возрастанию моментами времени (ключевой атрибут);  $X$  – атрибут (домен), хранящий уровни ВР.

Нечеткий временной ряд  $\tilde{Y}_x$  по атрибуту  $X$  есть результат операции лингвистического оценивания  $\tilde{x}_i = Fuzzy(x_j)$ ,  $x_j \in X$ ,  $\tilde{x}_i \in \tilde{X}$  уровней  $X$  носителя – временного ряда  $Y$ . Структурно-лингвистическая модель уровней нечеткого временного ряда  $\tilde{Y}_x$  есть расширение отношения  $Y$  следующего вида

$$\tilde{Y}_x = \langle X, \tilde{X}, \mu_{\tilde{X}}(X) \rangle,$$

где  $\tilde{X}$  – задает абсолютную лингвистическую оценку;  $\mu_{\tilde{X}}(X)$  – степень принадлежности  $X$  лингвистической метке  $\tilde{X}$ . Тогда внешняя семантика входных данных ГСА НВР  $\tilde{Y} = P_s(Y)$  есть структурно-лингвистическая модель ВР  $\tilde{Y} = \langle T, \tilde{X}, \mu(X) \rangle$ .

Для определения модели выходных данных  $\hat{Y}$  каждому моменту времени в структурно-лингвистической модели НВР сопоставим нечеткую тенденцию, представленную моделью  $\tau = \langle v, \mu, \alpha, \Delta t \rangle$ , где  $\tau$  – наименование нечеткой тенденции;  $v$  – тип нечеткой тенденции, из базового множества  $Ture = \{ \langle \text{Рост} \rangle, \langle \text{Падение} \rangle, \langle \text{Стабильность} \rangle \}$ . Базовое множество может быть расширено за счет производных типов тенденций. Последова-

тельность типов нечетких тенденций моделирует структуру изменений НВР;  $\mu$  – функция принадлежности нечеткой тенденции, которая формируется на основе образующих ее нечетких значений ВР;  $\alpha$  – функция интенсивности каждого типа нечеткой тенденции (контекстное расширение тенденции);  $\Delta t$  – длительность каждого типа нечеткой тенденции.

Введем *FT-расширение* структурно-лингвистической модели НВР в базе нечетких тенденций. *FT-расширение* есть расширенная структурно-лингвистическая модель НВР  $\tilde{Y}' = \langle T, X, \tilde{X}, \mu(X), \tau \rangle$ .

Обозначим в виде функционала  $P_s = P(Y, \hat{Y}, F, K)$  совокупность внешних моделей  $F$ , представляющих собой образованные на структурно-лингвистической модели НВР проекционные модели.

Проекция расширенной модели  $\tilde{Y}'$  по моментам времени и степени нечеткости НТ при фиксированном  $v$  ( $v$ -фильтр) есть отношение  $Pr_2 = Proj_{T,\alpha v}(\tilde{Y}')$ , моделирующее в виде временного ряда функцию принадлежности тенденций НВР. Проекция расширенной модели  $\tilde{Y}'$  по моментам времени и степени интенсивности НТ при фиксированном  $v$  ( $v$ -фильтр) есть отношение  $Pr_3 = Proj_{T,\alpha v}(\tilde{Y}')$ , моделирующее в виде временного ряда функцию выраженности соответствующего типа тенденции  $v$ .

Предлагаемая внешняя семантика системы  $S$ , выраженная на основе уточнения функционала  $P_s = P(Y, \hat{Y}, F, K)$ , определяет внутреннюю семантику системы  $S$ , то есть конкретизирует структуру функционала  $R_s = R(Y, \hat{Y}, F, K)$ .

Так как проекционные модели образуют новые числовые временные ряды, которые представляют для функционала  $R_s$  вход  $Y$  и выход  $\hat{Y}$ , то в качестве моделей и методов  $F$  решения задач Data Mining Time Series целесообразно использовать как методы регрессионного анализа, так и методы, основанные на интеллектуальных технологиях. Из методов регрессионного анализа наиболее точными являются методы, базирующиеся на методологии динамического регрессионного моделирования (ДРМ-подходе) [4]. В качестве конкурирующего подхода представляет интерес использование методов нейросетевых технологий [3].

**Обобщенный алгоритм функционирования.** Предлагаемая концептуальная модель гибридной системы анализа нечетких временных рядов в виде комплекса моделей порождает следующий обобщенный алгоритм ее функционирования:

- 1) анализ источника данных временных рядов;
- 2) преобразование данных временных рядов в нечеткие временные ряды и построение структурно-лингвистической модели НВР;
- 3) *FT-расширение* структурно-лингвистической модели НВР в базе нечетких тенденций;

- 4) построение проекционных моделей расширенной структурно-лингвистической модели НВР;
- 5) прогнозирование в пространстве проекционных моделей;
- 6) композиция результатов прогноза.

**Заключение.** В настоящей работе рассмотрена актуальность обработки нечетких временных рядов. Предлагаемая концептуальная модель системы обработки НВР базируется на структурно-семантическом анализе, позволяющем на основе аппарата нечетких множеств преобразовать исходный НВР в совокупность числовых временных рядов. Анализ и прогнозирование порожденных числовых рядов могут быть решены на основе нейросетевых технологий и статистического временного моделирования. Такая глубинная гибридизация обработки, как представляется, может служить эффективным инструментом при решении задач анализа и прогнозирования НВР.

#### *Список литературы*

1. Zadeh L.A. From Search Engines to Question-Answering System – The Problem of World Knowledge, Relevance, Deduction and Precision / IRI' 05, Las Vegas. URL. – Режим доступа : [http://www.cis.fiu.edu/conferences/IRI05/html/IRI05keynotes/IRI05\\_Keynote\\_Zadeh.pdf](http://www.cis.fiu.edu/conferences/IRI05/html/IRI05keynotes/IRI05_Keynote_Zadeh.pdf), свободный.
2. Магнитов, С.Н. От дихотомии к трихотомии, от трихотомии к триединству слова [Электронный ресурс] / С.Н. Магнитов. – Режим доступа : <http://www.trinitas.ru/rus/doc/0216/004a/02160008.htm>, свободный.
3. Ярушкина, Н.Г. Основы теории нечетких и гибридных систем : учеб. пособие / Н.Г. Ярушкина. – М. : Финансы и статистика, 2004. – 320 с.
4. Валеев, С.Г. Регрессионное моделирование при обработке наблюдений / С.Г. Валеев. – М. : Наука, 1991. – 272 с.

---

### **Conceptual Model of Hybrid System for Fuzzy Time Series Analysis**

**T.V. Afanasyeva, S.G. Valeev, N.G. Yarushkina**

*Ulyanovsk State Technical University, Ulyanovsk*

**Key words and phrases:** Data Mining Time Series tasks; neuron networks; fuzzy time series; fuzzy tendencies; general algorithm; regression modeling; structure-linguistic model.

**Abstract:** The paper studies the problem of fuzzy time series analysis. The formal conceptual model of hybrid system for fuzzy time series analysis based on intelligence and regression technologies is proposed.

---

© Т.В. Афанасьева, С.Г. Валеев, Н.Г. Ярушкина, 2008