

УДК 004.942

Н.Г. Ярушкина, Т.В. Афанасьева, А.М. Наместников, Г.Ю. Гуськов

ИНТЕГРАЦИЯ НЕЧЕТКО-ГРАНУЛЯРНЫХ И ОНТОЛОГИЧЕСКИХ МЕТОДОВ В ЗАДАЧЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Ярушкина Надежда Глебовна, доктор технических наук, профессор, первый проректор - проректор по научной работе Ульяновского государственного технического университета. Окончила радиотехнический факультет Ульяновского государственного технического университета. Имеет более 250 работ в области мягких вычислений, нечеткой логики, гибридных систем. [e-mail:jng@ulstu.ru].

Афанасьева Татьяна Васильевна, доктор технических наук, доцент, заместитель заведующего кафедрой "Информационные системы" УлГТУ. Окончила радиотехнический факультет Ульяновского государственного технического университета. Имеет статьи и монографии в области интеллектуального анализа временных рядов. [e-mail:tv.afanaseva@ulstu.ru].

Наместников Алексей Михайлович, кандидат технических наук, доцент, доцент кафедры "Информационные системы" УлГТУ. Окончил радиотехнический факультет Ульяновского государственного технического университета. Имеет статьи и монографию в области интеллектуальных систем обработки и хранения информации. [e-mail:nam@ulstu.ru].

Гуськов Глеб Юрьевич, аспирант кафедры "Информационные системы" УлГТУ, окончил факультет информационных систем и технологий УлГТУ. Имеет работы в области интеллектуального анализа временных рядов. [e-mail:guskovgleb@gmail.com].

Аннотация

Для решения проблемы анализа нечетких тенденций временных рядов (ВР) были разработаны теоретические и методологические основы нечетко-гранулярного моделирования и предложена их практическая реализация в виде комплекса программ, что позволило получить решение ряда прикладных задач. Нечетко-гранулярное представление временного ряда включает в себя пять уровней грануляции: от гранулы числового значения до гранулы основной тенденции. Для интерпретации результатов анализа временных рядов предложено использовать нечеткие онтологии проблемной области. Базис онтологии основывается на модели RDF, предполагающей определение классов, экземпляров, онтологических отношений и ограничений. Логический вывод рекомендации осуществляется на основе взаимодействия нечеткой OWL-онтологии и системы продукционных SWRL-правил. В работе рассмотрены возможности интеграции нескольких методов прогнозирования временных рядов и соответствующие агрегирующие показатели.

Приведенная в данной статье интеграция методов анализа временных рядов и онтологического анализа показывает конкурентоспособность моделей нечетких тенденций, основанных на взаимном усилении онтологических и нечетко-гранулярных методов.

Ключевые слова: нечеткие временные ряды, нечеткие тенденции, онтология, нечеткая онтология.

INTEGRATION OF FUZZY GRANULAR AND ONTOLOGICAL METHODS FOR TIME SERIES ANALYSIS

Nadezhda Glebovna Yarushkina, Doctor of Engineering, Professor, First Vice-Rector – Vice-Rector for Science of Ulyanovsk State Technical University; graduated from the Faculty of Radioengineering of Ulyanovsk State Technical University; an author of more than 250 papers in the field of soft computing, fuzzy logic, and hybrid systems. e-mail: jng@ulstu.ru.

Tatiana Vasilevna Afanaseva, Doctor of Engineering; Associate Professor, Deputy Head of Information Systems Department at Ulyanovsk State Technical University; graduated from the Faculty of Radioengineering at Ulyanovsk State Technical University; an author of articles and monographs in the field of time series data mining. e-mail: tv.afanasjeva@gmail.com.

Aleksei Mikhailovich Namestnikov, Candidate of Engineering, Associate Professor, Associate Professor at the Department of Information Systems at Ulyanovsk State Technical University; graduated from the Faculty of Radioengineering of Ulyanovsk State Technical University; an author of articles and a monograph in the field of intelligent systems for storage and processing of information. e-mail: nam@ulstu.ru.

Gleb Iurevich Guskov, Post-Graduate Student of the Department of Information Systems at Ulyanovsk State Technical University; graduated from the Faculty of Information Systems and Technologies at Ulyanovsk State Technical University; an author of articles in the field of time series data mining. e-mail: guskovgleb@gmail.com.

Abstract

The theoretical and methodological foundations of fuzzy granular modeling were developed to solve the problem of fuzzy trend analysis of time series. Their practical implementation as a program complex allows to get the solution of some applied problems. Fuzzy granular presentation of time series includes five levels: from numeric value granulation to the main trend granulation. The authors propose to use fuzzy ontologies of problem area to interpret the results of time series analysis. The problem area fuzzy ontologies were offered to interpret the results of time series analysis. The ontology basis is based on the RDF model that defines classes, instances, ontological relationships, and contingencies. The logical interference of recommendation is realized on the basis of interaction between fuzzy OWL ontology and rule-oriented SWRL rules system. The article deals with the possibilities of integration of several methods of time series forecasting and corresponding aggregates.

The integration of techniques for time series analysis and ontological analysis demonstrates the competitiveness of fuzzy trend models based on the mutual reinforcement of ontological and fuzzy granular methods.

Key words: fuzzy time series, fuzzy tendencies, ontology, fuzzy ontology.

ВВЕДЕНИЕ

Идея нечеткого моделирования поведения сложных систем по ВР основана на базовой модели нечеткого процесса, получившей название «нечеткого временного ряда» (НВР) [1, 4]. В НВР состояниям динамического процесса сопоставлены нечеткие значения, моделируемые параметрическими функциями принадлежности, зависимости между ними строятся по наблюдениям и представляются базой нечетких правил «Если-То», численно выражаемых матрицей нечеткого отношения. Однако в традиционных моделях НВР отсутствует возможность анализа его нечетких тенденций и содержательной интерпретации результатов анализа.

Для решения проблемы анализа нечетких тенденций НВР были разработаны теоретические и методологические основы нечетко-гранулярного моделирования ВР и предложена их практическая реализация в виде комплексов программ, что позволило получить решение ряда прикладных задач [3]. Для интерпретации результатов анализа временных рядов предложено использовать нечеткие онтологии проблемной области. Интеграция методов анализа временных рядов и онтологического анализа основана на взаимном усилении онтологических и нечетко-гранулярных методов [12, 13].

1 ВИДЫ НЕЧЕТКИХ ПРОЦЕССОВ И НЕЧЕТКО-ГРАНУЛЯРНОЕ МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

Представление временных рядов в классе нечетких процессов основывается на базовой модели нечеткого временного ряда [1, 4], конкретизация которой в зависимости от предположений и класса используемых переменных приводит к различным моделям процессов. В отличие от традиционного ВР значениями нечеткого ВР являются нечеткие множества, а не действительные числа. В 1993 году Сонг и Чиссом [4] предложили модели нечетких временных рядов первого порядка.

Определение 1. Пусть $X_t, (t = 1, 2 \dots) \subset R^1$ – универсальное множество, на котором определены нечеткие множества $\tilde{x}_t^i, (i = 1, 2 \dots)$, и \tilde{X}_t – коллекция $\tilde{x}_t^i, (i = 1, 2 \dots)$. Тогда \tilde{X}_t – НВР.

Предполагается, что поведение НВР может быть описано нечетким процессом, в котором существует зависимость текущего состояния от предыдущих $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots, \tilde{X}_{t-p} \rightarrow \tilde{X}_t$, представляемая в виде нечетко-разностного уравнения [1, 4]:

$$\tilde{X}_t = (\tilde{X}_{t-1} \times \tilde{X}_{t-2} \times \dots \times \tilde{X}_{t-p}) \circ R(t, t-p),$$

где $\tilde{X}_t, \tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots, \tilde{X}_{t-p}$ – состояния нечеткого процесса, кодируемые нечеткими множествами (лингвистическими терминами);

$R(t, t-p)$ – нечеткое отношение, определяющее модель нечеткого процесса p -го порядка в терминах нечетких значений \tilde{X}_t , которая может быть выражена совокупностью нечетких «уравнений» в виде правил «Если-То»;

\circ – знак композиции;

p – порядок модели.

В дальнейшем модель нечеткого процесса для НВР сокращенно будем обозначать S -моделью. Модель НВР лингвистически представима базой нечетких правил «Если-То» и численно может быть реализуема на основе алгоритма нечеткого логического вывода Мамдани [1]. Применение модели для прогнозирования первых разностей числового временного ряда рассмотрено в работах [3, 5].

В развитии теории нечетких временных рядов рассматриваются изменения в нечетком временном ряду как результат совместного влияния неизвестных факторов, которые могут быть значимы и незначимы. Значимые влияния приводят к качественному изменению значений и должны быть включены в модель, а незначимыми изменениями можно пренебречь. Такие качественные изменения (нечеткие приращения) в нечетких временных рядах получили название «нечеткие тенденции» [1].

Пусть задан дискретный временной ряд $Y = \{t_i, x_i\}$, $i \in [1, n]$. Согласно базовым положениям теории нечетких временных рядов, разработанным Заде, Сонгом и Чиссомом, любой конечный дискретный временной ряд – числовой, нечисловой и смешанный – может быть преобразован в нечеткий ВР $\tilde{Y} = \{t_i, \tilde{x}_i\}$, $i \in [1, n]$ при условии, что его множество значений $X = \{x_i\}$ будет покрыто специальными функциями (нечеткими множествами) $\tilde{x}_i \in \tilde{X}$.

Утверждение 1. Каждая нечеткая тенденция $\tau \in \mathfrak{S}$ нечеткого временного ряда $\tilde{Y} = \{t_i, \tilde{x}_i\}$, $i \in [1, n]$ может быть представлена структурной моделью в виде кортежа отношения, построенного на декартовом произведении свойств нечеткой тенденции $\tilde{V} \times \tilde{A} \times \Delta T \rightarrow \mathfrak{S}$:

$$\tau = \langle \tilde{v}, \tilde{a}, \Delta t, \mu \rangle,$$

где τ – наименование нечеткой тенденции из множества \mathfrak{S} , $\tau \in \mathfrak{S}$;

\tilde{v} – тип нечеткой тенденции (тип изменений), $\tilde{v} \in \tilde{V}$ выражает основные качественные зависимости временных рядов. Введем инвариантное по отношению к предметным областям и временным рядам множество лингвистических термов типов нечетких тенденций: $\tilde{V} = \tilde{V}1 \cup \tilde{V}2$, где $\tilde{V}1$ определяет множество базовых типов тенденций НВР, которое конечно и ограничено тремя типами $\tilde{V}1 = \{\tilde{v}1, \tilde{v}2, \tilde{v}3\}$, $\tilde{v}1 = \langle \text{«Стабилизация»}$, $\tilde{v}2 = \langle \text{«Падение»}$, $\tilde{v}3 = \langle \text{«Рост»}$, $\tilde{V}2$ – множество производных типов тенденций НВР, которое также конечно и ограничено $\tilde{V}2 = \{\tilde{v}4, \tilde{v}5\}$, $\tilde{v}4 = \langle \text{«Колебания»}$, $\tilde{v}5 = \langle \text{«Хаос»}$;

\tilde{a} – интенсивность нечеткой тенденции, $\tilde{a} \in \tilde{A}$, которое может быть выражено лингвистически, например, значениями из множества $\{\text{«Большое»}$, «Среднее» , «Малое» , «Отсутствует» ;

Δt – продолжительность нечеткой тенденции, $\Delta t \in \Delta T$;

μ – функция принадлежности участка нечеткого временного ряда, ограниченного Δt нечеткой тенденции τ .

Классифицируем нечеткие тенденции НВР в зависимости от продолжительности на элементарную $\tau \in \mathfrak{S}$ ($\Delta t = 1$), локальную $N\tau \in N\mathfrak{S}$ ($1 < \Delta t < n - 1$) и общую $G\tau \in G\mathfrak{S}$ ($\Delta t = n - 1$).

Определение 2. Нечеткая элементарная тенденция – это нечеткая тенденция НВР $\tau_i \in \mathfrak{S}$, $\tau_i = \langle \tilde{v}_i, \tilde{a}_i, \mu_i \rangle$, выражающая качественный характер изменения на участке НВР между двумя соседними нечеткими метками НВР $\tilde{x}_i, \tilde{x}_{i+1}$.

Утверждение 2. Исходный нечеткий временной ряд $\tilde{Y} = \{t_i, \tilde{x}_i\}$, $i \in [1, n]$ может быть представлен в виде нечеткого временного ряда нечетких элементарных тенденций $\tau = \{t_i, \tau_i\}$, $i \in [1, n - 1]$ и нечеткого временного ряда локальных тенденций $N\tau = \{t_j, N\tau_j\}$ $j \in [1, n1]$, $n1 < n$.

Из утверждения 2 и теории нечетких временных рядов вытекает следующее утверждение.

Утверждение 3. Любой конечный дискретный временной ряд – числовой, нечисловой и смешанный – может быть преобразован во временной ряд нечетких элементарных тенденций.

2 НЕЧЕТКО-ГРАНУЛЯРНОЕ ПРЕДСТАВЛЕНИЕ ВР

Рассмотрим ВР на пяти уровнях абстракции, на каждом из которых его поведение может быть представлено нечетко-гранулярными моделями фиксированной семантики и порождаться процессами различных классов [2]. Для каждого уровня абстракции определим информационную гранулу, кодирующую состояние процесса, в виде

$$x \text{ is } \beta^r,$$

где r – уровень грануляции ВР;

$r = 0$: $\beta^0 = x$ – гранула числового значения исходного ВР (вырожденное нечеткое значение);

$$r = 1$$
: $\beta^1 = \tilde{x}$ – гранула нечеткого значения НВР;

$r = 2$: $\beta^2 = \tau$ – гранула нечеткой элементарной тенденции, идентифицируемая между двумя соседними нечеткими значениями НВР;

$r = 3$: $\beta^3 = n\tau$ – гранула нечеткой локальной тенденции, идентифицируемая между двумя выбранными нечеткими значениями НВР;

$r = 4$: $\beta^4 = g\tau$ – гранула основной нечеткой тенденции, идентифицируемая между начальным и конечным нечеткими значениями НВР.

Гранулярное представление системы объектов и модели анализа ВР будем рассматривать в виде тройки

$$Granule = \{\beta^r, P^r, M^r\}, r = 0, 1, \dots, 4,$$

где $\beta^r = \{\beta_s^r\}$, ($s = 1, 2, \dots$) – совокупность анализируемых объектов ВР на уровне грануляции r ;

$P^r = \{P_i^r\}$, ($i = 0, \dots, 4$), $i \neq r$ – совокупность математических соотношений, порождающих гранулы разного уровня; они задают гранулярные преобразования объектов ВР;

$M^r = \{M_k^r\}_2$ ($k = 1, 2, \dots$) – совокупность математических моделей динамических процессов, порождающих временные ряды на соответствующем уровне абстракции.

3 ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА

Прикладная онтология представляет семантическое описание предметной области с использованием специализированного языка. В 1997 консорциум W3C определил спецификацию RDF (Resource Description Framework). RDF предоставляет простой, но мощный язык описания ресурсов, основанный на триплетях (triple-based) «Субъект-Предикат-Объект» и спецификации URI. Кроме RDF был разработан язык описания структурированных словарей для RDF – RDF Schema (RDFS). Он предоставляет минимальный набор средств для спецификации онтологий. Базовой

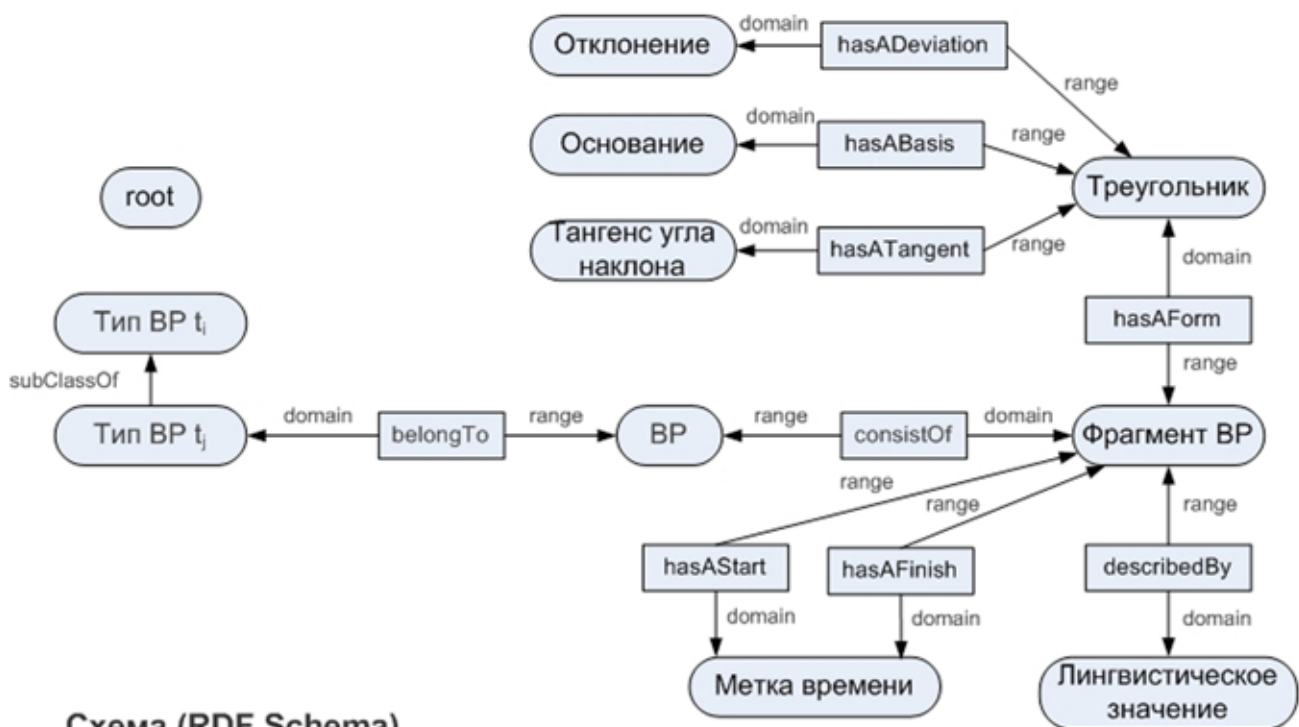


Схема (RDF Schema)

Онтология (RDF) (пример фрагмента)

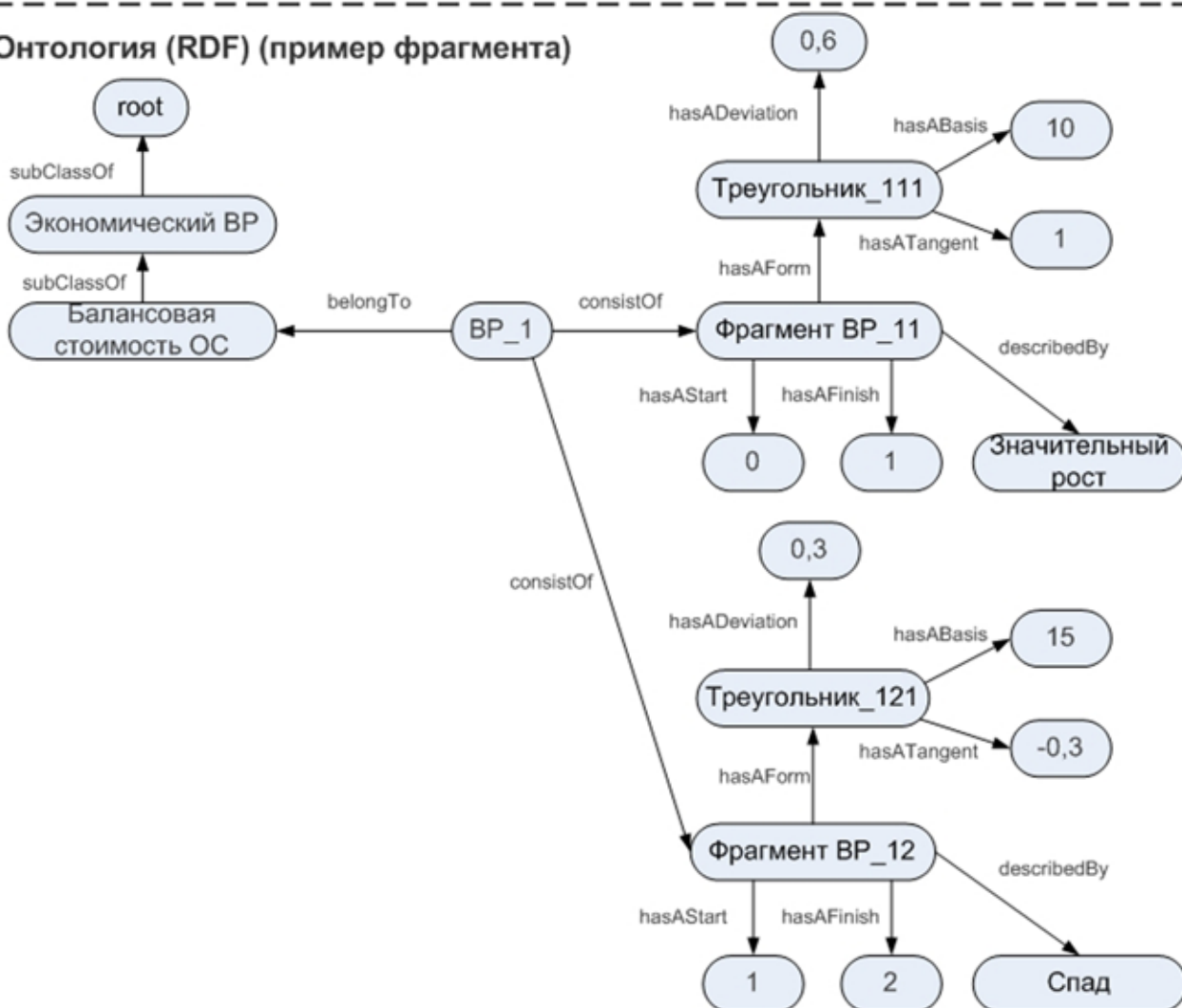


Рис. 1. RDF-схема и фрагмент онтологии

структурной единицей RDF является коллекция троек (или триплетов), каждая из которых состоит из субъекта, предиката и объекта (S, P, O). Набор триплетов называется RDF-графом. В качестве вершин графа выступают субъекты и объекты, в качестве дуг – предикаты (или свойства). Направление дуги, соответствующей предикату в данной тройке (S, P, O), всегда выбирается так, чтобы дуга вела от субъекта к объекту.

Для описания знаний экспертов используется формат RDF [14]. На рисунке 1 представлены схема онтологии содержательной интерпретации временных рядов и один из фрагментов онтологии. Согласно схеме корневым элементом является ресурс, указывающий на предметную область. Предметная область, в свою очередь, включает некоторые показатели. Например, предметная область «Экономический временной ряд» включает показатель «Балансовая стоимость основных средств». Каждый показатель предметной области базируется на одном или нескольких типовых временных рядах. Иными словами, на основании этих временных рядов формируются новые знания о данном показателе.

Каждый временной ряд может быть разбит на фрагменты (интервалы разбиения). Каждый фрагмент имеет несовпадающие начальную и конечную точки. Стоит отметить, что номер конечной точки фрагмента всегда больше номера начальной точки. Каждому фрагменту дается лингвистическое описание, характеризующее динамику рассматриваемого показателя за промежуток времени, соответствующий этому фрагменту.

Отдельный фрагмент временного ряда может быть рассмотрен как прямоугольный треугольник, начальная и конечная точки отрезка гипотенузы которого соответствуют уровням временного ряда в начальной и конечной точке фрагмента. Каждый ресурс-треугольник модели имеет такие свойства, как длина основания, тангенс угла наклона и абсолютное значение ошибки.

Длина основания может быть вычислена как произведение количества промежутков между измерениями во фрагменте (число измерений, уменьшенное на единицу) и масштаба временного ряда (величина промежутка времени между двумя измерениями).

Тангенс угла наклона находится для угла между гипотенузой и отрезком временной оси, находящимся правее точки пересечения с гипотенузой треугольника. Таким образом, если на рассматриваемом фрагменте наблюдается рост показателя, то тангенс угла наклона будет положительным, а если падение – отрицательным.

В рассматриваемых треугольниках начальная и конечная точки отрезка гипотенузы совпадают с соответствующими уровнями временного ряда. Промежуточные же точки гипотенузы, как правило, не соответствуют уровням временного ряда. Совокупность абсолютных значений разности между реальными и полученными в ходе аппроксимации уровнями временного ряда для данного треугольника представляет собой абсолютное значение ошибки аппроксимации.

4 МЕТОДИКА ПОСТРОЕНИЯ НЕЧЕТКИХ ОНТОЛОГИЙ FuzzyOWL

Формальная семантика OWL описывает, как получить логические следствия, имея такую онтологию, т.е. получить факты, которые не представлены в онтологии буквально, но следуют из ее семантики.

OWL (англ. Web Ontology Language) – язык описания онтологий, позволяющий представлять классы и отношения между ними. В основе языка – представление действительности в модели данных «объект – свойство». OWL является переформулировкой дискрипционной логики (DL) с использованием синтаксиса XML. В настоящее время существует наиболее универсальный и имеющий достаточную поддержку со стороны разработчиков интеллектуальных экспертных систем подход к построению нечетких онтологий – методология FuzzyOWL.

Синтаксис FuzzyOWL предполагает три формата используемого алфавита: для представления нечетких классов (fuzzy concepts), нечетких отношений (fuzzy roles) и нечетких объектов классов онтологии (fuzzy individuals).

В отличие от OWL DL, реализующей разновидность дискрипционной логики $SHOIN(D)$, концепция FuzzyOWL охватывает все положения логики $SROIQ(D)$:

- S – это набор базовых классов, имен свойств и имен объектов, является подмножеством DL в связке с транзитивными свойствами;

- R – возможность определения составных аксиом вложенности ролей в наборе аксиом перечисленных видов: $R \circ S \subseteq R, R \circ S \subseteq S$, где $R \circ S$ – композиция ролей;

- O – наличие номиналов (так называемых классов единичных объектов);

- I – наличие обратных свойств;

- Q – качественные ограничения кардинальности ролей;

(D) – типы данных, свойства, которые связывают объекты с типами данных.

Основополагающим элементом методологии построения нечетких онтологий с точки зрения структуры OWL является наличие свойства-аннотации `fuzzyLabel`, которое хранит параметры и значения функции принадлежности, соотносимые с конкретным классом, объектом, отношением внутри заданного нечеткого множества данных.

Методология FuzzyOWL с помощью необходимых параметров нечеткости в аннотациях к элементам онтологии предполагает возможность делать нечеткими:

- классы;
- объекты классов онтологий;
- свойства (свойства типов данных и свойства объектов);
- аксиомы онтологий [11].

Алгоритм задания нечеткого отношения, согласно методике FuzzyOWL, схематично представлен на рисунке 2.

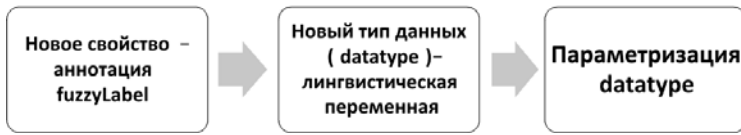


Рис. 2. Определение нечеткого отношения согласно методике FuzzyOWL

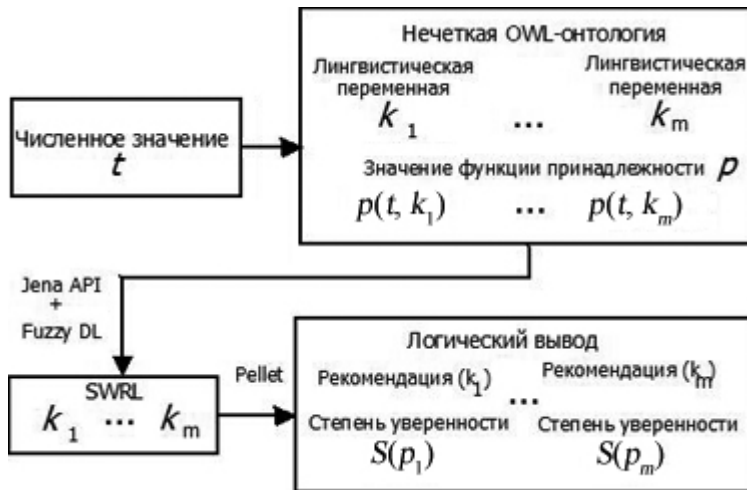


Рис. 3. Логический вывод на основе нечеткой онтологии

Алгоритм выглядит следующим образом:

1. Создается новое свойство-аннотация fuzzyLabel, в котором будут задаваться параметры нечеткости каждого элемента онтологии.
2. Лингвистические переменные определяются в форме новых созданных типов данных (datatype).
3. Для каждого созданного типа данных задаются верхняя и нижняя границы принимаемых значений, выбирается тип функции принадлежности, и для нее задаются соответствующие параметры [11, 12]. Логический вывод рекомендаций с использованием онтологии, содержащей нечеткие отношения, классы, а также объекты классов, имеет существенные отличия по сравнению с извлечением знаний из четкой онтологии (рис. 3).

В качестве формы знаний, позволяющей делать непосредственный вывод из содержащихся в онтологии данных и дополнять ее новыми, используется набор SWRL-правил. Значениям лингвистических переменных, заданным в теле SWRL-правила, соответствуют некоторые степени принадлежности, определенные в OWL-онтологии и в дальнейшем формирующие некоторую величину степени уверенности в процессе вывода сработавшего правила.

Основная последовательность действий в процессе логического вывода рекомендации на основе взаимодействия нечеткой OWL-онтологии и системы продукционных SWRL-правил включает в себя следующие этапы:

1. Проверка входных данных на наличие в онтологии соответствующих классов, отношений, объектов.
2. Определение лингвистических переменных и соответствующих значений функции принадлежности по чис-

ленному значению входного параметра, определенного в онтологии.

3. Запрос на выполнение набора правил анализа ситуации и вывода рекомендации.
4. Запрос на получение связанных данных (обращение к «родителям» объектов).
5. Запрос на выполнение правил с учетом полученных связанных данных.
6. Вывод результата анализа (одна или несколько рекомендаций) в виде объекта разработанной OWL-онтологии с учетом показателя степени уверенности S .

При этом показатель степени уверенности является функцией от поступающих значений функций принадлежности, относящихся к заложенным в соответствующем правиле значениям предметной области:

$$S_{Rule} = \min \mu(t, k_j), \quad (1)$$

где t – термин проблемной области, присутствующий в i -м сработавшем SWRL-правиле;

k – лингвистическая переменная, $j = [1..m]$;

$\mu(t, k)$ – значение соответствующей функции принадлежности.

В целях повышения наглядности полученного результата необходимо провести нормализацию показателей степеней уверенности всех сработавших правил:

$$S_{n\text{Norm}} = \frac{S_n}{\max(S_n)}, \quad (2)$$

где $S_{n\text{Norm}}$ – нормализованное значение степени уверенности рекомендации каждого n -го правила.

Таким образом, наиболее вероятное решение по управлению той или иной ситуации, моделируемое значениями входных параметров, будет иметь значение степени уверенности, равное 1.

В конечном итоге пользователь экспертной системы получит на выходе набор рекомендаций на естественном языке, каждой из которых будет поставлено в соответствие некоторое значение $S = [0..1]$.

5 КОМБИНАЦИИ И КОЛЛЕКТИВЫ МОДЕЛЕЙ

Одним из наиболее успешных направлений в прогнозировании временных рядов является объединение их в комбинации или коллективы. Как показывает практика, объединение моделей, будь то комбинация или коллектив, позволяет получить результат лучше, чем результат каждой из моделей в отдельности. Для осуществления автоматизированного принятия решения о лучшем методе анализа ВР целесообразно использовать нечеткую онтологию анализа временных рядов. В данном проекте были реализованы комбинация моделей с использованием весов, полученных в результате информационных критериев (ИК) [7], комбинация с использованием нечетких весов [8] и коллектив, полученный агрегирующим алгоритмом В. Вовка [2]. Основной идеей данного подхода к комбина-

Результаты прогнозирования

Лучший результат	Веса ИК	Веса по ИК (Байес)	Веса по Акаике	Нечеткие веса	Алгоритм В. Вовка
27.3847	27.2887	27.2887	27.2887	28.3996	29.8256
23.3531	27.3576	27.3576	27.3576	41.801	24.0437
6.48463	6.49801	6.49801	6.49801	8.66947	9.99404
8.39089	7.64793	7.64793	7.64793	8.52974	10.6689
5.494	5.3245	5.3245	5.3245	11.9307	6.54996

ции является вычисление весов для моделей на базе ИК. В проекте реализован способ расчета весов, рассмотренный в [8]. Для расчета весов определяется разница между наименьшим значением критерия и текущим значением, например, критерия Акаике:

$$\Delta_{AIC}(M) = AIC(M) - \min_{N \in \mathcal{M}} (AIC(N)).$$

Далее рассчитывается функция правдоподобия:

$$Likelihood = \exp\left(-\frac{1}{2} \Delta_{AIC}(M)\right).$$

Агрегирующий алгоритм В. Вовка [2] рассматривает процесс прогнозирования в виде тройки $\langle \Omega, \Gamma, \lambda \rangle$, где Ω – множество исходов, Γ – множество допустимых предсказаний, $\lambda: \Omega \times \Gamma \rightarrow \mathbb{R}^+$ – функция потерь. Алгоритм прогнозирования представляется в виде функции $A: \Omega^* \rightarrow \Gamma$, которая выдает прогноз $\gamma(T+1)$ для следующего элемента последовательности. При этом потери при прогнозировании определяются как $LossA(\omega_1, \dots, \omega_T) = \sum_{t=1}^T \lambda(\omega_t, \gamma_t)$. Агрегирующий алгоритм оперирует с несколькими моделями, взвешивая в экспоненциальном пространстве ошибки прогнозов экспертов, а затем порождает прогноз с помощью решающего правила:

$$\gamma_{T+1} = \sum \log \left[\sum_{j=1}^N \exp(-\mu \cdot Loss_{A_j}(\omega)) p_0^{(j)} \right],$$

где $p_0^{(j)}$ – вес эксперта j в начальный момент времени;

$\mu > 0$ – параметр алгоритма.

В ходе реализации данного алгоритма было получено несколько модификаций. Для корректной работы алгоритма необходимо нормализовать данные временного ряда. Экспериментально был подобран отрезок нормализации от 0 до 1, в случае большой разницы между точками (порядка нескольких тысяч) происходило вырождение пересчёта весов. С учетом представленных моделей для агрегации также производится сдвиг временного ряда из области отрицательных значений. Кроме того, были реализованы методы для подготовки данных перед прогнозированием, а именно логарифмирование ряда и метод Box-Cox [14]. В таблице 1 приведены результаты тестирования реализованных агрегационных алгоритмов с указанием лучшего его результата по отдельной модели. Для тестирования использовались ряды из источника [6].

ЗАКЛЮЧЕНИЕ

Приведенная в данной статье интеграция методов анализа временных рядов и онтологического анализа показывает конкурентоспособность моделей нечетких тенденций, основанных на взаимном усилении онтологических и нечетко-гранулярных методов. Применение онтологии для решения задач интеллектуального анализа временных рядов позволяет выполнять содержательную интерпретацию фрагментов временного ряда без привлечения эксперта.

СПИСОК ЛИТЕРАТУРЫ

1. Ярушкина Н.Г. Основы теории нечетких и гибридных систем. – М.: Финансы и статистика, 2004. – 320 с.
2. V.G. Vovk Universal Forecasting Algorithms // Journal of Information and Computation, Volume 96, 1992, pp. 245–277.
3. Афанасьева Т. В., Ярушкина Н.Г. Нечеткий динамический процесс с нечеткими тенденциями в анализе временных рядов // Вестник РГУПС. – 2011. – № 3. – С. 6–15.
4. Song, Q. Fuzzy time series and its models // Fuzzy Sets and Systems. – 1993. – № 54. – pp. 269–277.
5. Şah, M. Forecasting Enrollment Model Based on First-Order Fuzzy Time Series // Proc. Int. Conf. Computational Intelligence (ICCI) (2004). – pp. 375–378.
6. Тестовая выборка временных рядов [Электронный ресурс]. – Режим доступа: <http://irafm.osu.cz/cif/cif-dataset.txt>.
7. Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, Journal of the Royal Statistical Society, Series B, 26, pp. 211–252.
8. Kolassa S. Combining exponential smoothing forecasts using Akaike weights. International Journal of Forecasting Volume 27, Issue 2, April–June 2011, pp. 238–251.
9. Viertl R. Statistical Methods for Fuzzy Data. John Wiley & Sons, Ltd, 2011, 270 p.
10. Huarng, K. Effective lengths of intervals to improve forecasting in fuzzy time series // Fuzzy Sets and Systems. – 2001. – № 123. – pp. 387–394.
11. Ярушкина Н.Г., Мошкин В.С. Применение онтологического подхода к анализу состояния локальной вычислительной сети // Радиотехника. – 2014. – № 7. – С. 120–124.
12. Yarushkina N.G. Soft Computing and complex system analysis // International Journal of General Systems. – 2001. – Т. 30. – № 1. – С. 71–88.

13. Ярушкіна Н.Г., Перфільева И.Г., Афанасьева Т.В. Интеграция нечетких моделей для анализа временных рядов // Известия Самарского научного центра Российской академии наук. – 2010. – Т. 12. – № 4–2. – С. 506–509.

14. Прогнозирование временных рядов: нечеткие модели / Т.В. Афанасьева, А.М. Наместников, И.Г. Перфильева, А.А. Романов, Н.Г. Ярушкіна; под науч. ред. Н.Г. Ярушкіной. – Ульяновск: УлГТУ, 2014. – 145 с.

REFERENCES

1. Yarushkina N.G. *Osnovy teorii nechetkikh i gibridnykh sistem* [Fundamentals of the Fuzzy and Hybrid Systems Theory]. Moscow, Finansy i statistika Publ., 2004. 320 p.

2. Vovk V.G. Universal Forecasting Algorithms. *Journal of Information and Computation*, vol. 96, 1992, pp. 245–277.

3. Afanaseva T. V. Yarushkina N. G. Nechetkii dinamicheskii protsess s nechetkimi tendentsiyami v analize vremennykh riadov [Fuzzy Time Series with Fuzzy Tendency]. *Vestnik RGUPS* [Scientific-technical Journal 'Vestnik Gosudarstvennogo Universiteta Putey Soobshcheniya], 2011, no. 3, pp. 6–15.

4. Song Q., Chissom B. Fuzzy Time Series and Its Models. *Fuzzy Sets and Systems*, no. 54, 1993, pp. 269–277.

5. Şah, M., Degtiarev K. Y. Forecasting Enrollment Model Based on First-Order Fuzzy Time Series. *Proc. Int. Conf. Computational Intelligence. ICCI*, 2004, pp. 375–378.

6. *Testovaiia vyborka vremennykh riadov* [Test Set of Time Series]. Available at: <http://irafm.osu.cz/cif/cif-dataset.txt>.

7. Box G. E. P. and Cox D. R. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 1964, pp. 211–252.

8. Kolassa S. Combining Exponential Smoothing Forecasts Using Akaike Weights. *International Journal of Forecasting*, vol. 27, Iss. 2, 2011, pp. 238–251.

9. Viertl R. *Statistical Methods for Fuzzy Data*. John Wiley & Sons, Ltd. Publ., 2011. 270 p.

10. Huarng, K. Effective Lengths of Intervals to Improve Forecasting in Fuzzy Time Series. *Fuzzy Sets and Systems*, 2001, no. 123, pp. 387–394.

11. Yarushkina N.G., Moshkin V.S. Primenenie ontologicheskogo podkhoda k analizu sostoiianiia lokalnoi vychislitelnoi seti [Applying Ontological Approach to the Analysis of the State of Local Area Network]. *Radiotekhnika* [Radioengineering], 2014, no. 7, pp. 120–124.

12. Yarushkina N. G. Soft Computing and Complex System Analysis. *International Journal of General Systems*, 2001, vol. 30, no. 1, pp. 71–88.

13. Yarushkina N.G., Perfileva I.G., Afanaseva T.V. Integratsiia nechetkikh modelei dlia analiza vremennykh riadov [Fuzzy Models Integration in Time Series Analysis]. *Izvestiia Samarskogo nauchnogo tsentra Rossiiskoi akademii nauk* [Bulletin of Samara Scientific Centre of the Russian Academy of Sciences], 2010, vol. 12, no. 4–2, pp. 506–509.

14. Afanaseva T.V., Namestnikov A.M., Perfileva I.G., Romanov A.A., Yarushkina N.G. *Prognozirovaniie vremennykh riadov: nechetkie modeli*; pod nach. red. N.G. Yarushkinoi [Time Series Forecasting: Fuzzy Models; under the sci. editorship of N. Yarushkina]. Ulyanovsk, UlSTU Publ., 2014. 145 p.