



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.89

ОПИСАНИЕ ПОДХОДА К ИЗВЛЕЧЕНИЮ ЗНАНИЙ ОБ АНОМАЛИЯХ ЛОКАЛЬНЫХ ТЕНДЕНЦИЙ ВРЕМЕННОГО РЯДА

Афанасьева Т.В., Заварзин Д.В.

Ульяновский государственный технический университет,

г. Ульяновск, Россия

tv.afanasjeva@gmail.com

d.zavarzin91@gmail.com

В работе предлагается новый подход для извлечения знаний об аномалиях в данных которые представлены в виде локальных нечетких тенденций временных рядов (ВР). Работа основана на определениях гранулярного представления ВР, а также классификациях аномалий и свойств поведения нечетких тенденций ВР описанных в предыдущих работах.

Ключевые слова: извлечение знаний, поиск аномалий, временной ряд, нечеткая тенденция.

Введение

Целью процесса извлечения знаний из любого источника данных – является выявление в них нетривиальных зависимостей и закономерностей. Это необходимо для того, чтобы получить знания, пригодные для интерпретации человеком экспертом и для помощи в принятии решений человеку, не обладающими квалификацией эксперта в широком диапазоне сфер человеческой деятельности. Например, методы извлечения знаний типа могут применяться в следующих отраслях деятельности: торговле (анализ структуры клиентов и покупок, моделей поведения потребителей), в сферах защиты информации банков и страховых организаций (выявление случаев мошенничества, создание новых целевых программ и услуг), медицинских учреждений (выработка дорожной карты лечения, прогнозирование состояния здоровья).

Рассматривая примеры, приведенные выше, нетрудно заметить, что одной из наиболее востребованных задач в этих сферах является поиск отклонения от нормального поведения процесса, знания из данных которого мы намерены извлекать. Поведение процесса не свойственное картине нормального (типичного) протекания процесса характеризуется понятием – аномалия. Процедуру извлечения знаний о поведении анализируемого процесса необходимо автоматизировать и это связано в первую очередь, со сложностью систем мониторинга, сбора и накопления гетерогенных данных с применением традиционных средств анализа (визуальное наблюдение, экспертное

оценивание). Каждый экземпляр данных в этом случае может быть зависим от множества других, а также внешних воздействий.

Процесс извлечения знаний всегда сопровождается работой с данными внушительных объемов, распределенных по временным интервалам разной размерности – т.е. темпоральными данными. В современных аналитических системах данные представляются в виде временных рядов, как наиболее удобного средства представления темпоральных данных. Такая форма представления является удобной для решения задач экспертного и традиционного методов анализа процессов в настоящее время.

Анализом временных рядов, а также задачей извлечения знаний об аномалиях в наборах данных занимается специальное научное направление – теория временных рядов. В частности, извлечение знаний об аномалиях является важным для решения задач прогнозирования данных – идентификация и устранение аномалий, например, способствует повышению точности и глубины прогноза [ссылка]. В настоящей работе предлагается рассмотреть задачу извлечения знаний об аномалиях с точки зрения частного метода анализа временных рядов – метода нечетких тенденций.

В последнее время популярность получило относительно новое направление теории анализа временных рядов – теория нечетких временных рядов. Она рассматривает временные ряды не с точки зрения привычных числовых моделей, а опираясь на математический аппарат нечеткой логики [Афанасьева и др., 2009]. Наибольший

интерес в этой области представляет извлечение знаний о наличии редко встречающихся локальных тенденций ВР. Однако, задача представления знаний об аномалиях в локальных тенденциях не получила достаточного рассмотрения в исследованиях отечественных и зарубежных авторов.

В настоящей работе будет предложен новый подход к извлечению знаний об аномалиях в локальных тенденциях на основе их нечеткого представления и особенностей поведения ВР. В первой главе будет дано понятие временного ряда, нечеткой тенденции и аномалии, рассмотрены существующие методы извлечения знания об аномалиях в локальных тенденциях ВР, а также какие недостатки они содержат. Во второй главе будет описан новый подход к извлечению знаний об аномалиях локальных тенденций ВР на основе анализа поведения ВР.

1. Существующие методы и подходы к извлечению знаний об аномалиях локальных тенденций

Временной ряд – это последовательность дискретных упорядоченных в неслучайные равноотстоящие моменты времени измерений (показателей наблюдений) $X = \{x_t\}$, где $t = \overline{1, n}$, характеризующих уровни состояния изучаемого процесса. Практически любые наблюдения, полученные в результате мониторинга процессов в различных сферах человеческой деятельности (безопасность, медицина, экология и т.д.), могут быть представлены в виде некоторой последовательности, зависимой от времени – то есть в форме ВР. В теории нечетких систем временной ряд представляется в форме нечеткого временного ряда (НВР), где в равноотстоящие дискретные временные отсчеты $\tilde{X} = \{\tilde{x}_t\}$, где $t = \overline{1, n}$, фиксируются последовательности нечетких меток. Для НВР было формализовано понятие нечеткой тенденции, выражающей качественное изменение на временном интервале НВР.

Нечеткая тенденция (НТ) ВР $\tau \in \mathfrak{Z}$ НВР $\tilde{X} = \{\tilde{x}_t\}$, $t = \overline{1, n}$, может быть представлена структурно-лингвистической моделью в виде отношения, построенного на декартовом произведении свойств нечеткой тенденции $\tilde{V} \times \tilde{A} \times \Delta \tilde{T} \rightarrow \mathfrak{Z}$ [Афанасьева и др., 2011]:

$$\tau = \langle \tilde{v}, \tilde{a}, \Delta \tilde{t}, \mu \rangle \quad (1)$$

где \tilde{v} – функция принадлежности нечеткого множества типа НТ $\tilde{v} \in \tilde{V}$, выражающая направление изменения (например, $\tilde{v}_1 =$ «рост», $\tilde{v}_2 =$ «стабильность», $\tilde{v}_3 =$ «падение» и т.д.); \tilde{a} – функция принадлежности нечеткого множества интенсивности НТ, $\tilde{a} \in \tilde{A}$, выраженное лингвистическим термом (например, $\tilde{a}_1 =$

«большое», $\tilde{a}_2 =$ «среднее», $\tilde{a}_3 =$ «малое» и т.д.); $\Delta \tilde{t}$ – длительность НТ, где $\Delta \tilde{t} \in \Delta \tilde{T}$; μ – функция принадлежности участка НВР, ограниченного интервалом $\Delta \tilde{t}$ НТ.

Под аномалиями объектов ВР мы можем понимать следующие: четкие и нечеткие значения, локальные, элементарные и основные тенденции, а также те типы, которые могут быть найдены в зависимостях объектов ВР одного типа (правила следования) или разных типов (комбинированные) (рис. 1).

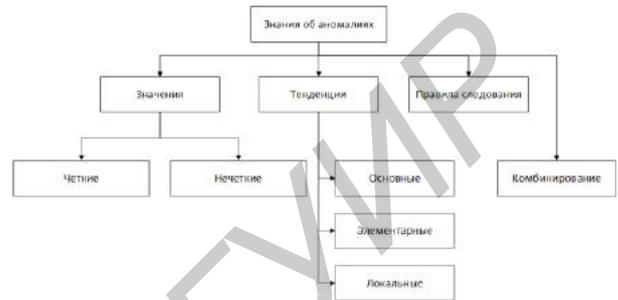


Рисунок 1 – Схема типологии знаний об аномалиях ВР

Опираясь на формулу (1) для представления знаний об аномалиях в локальных тенденциях в работе [Афанасьева и др., 2014] использовалась следующая классификация темпоральных процессов, которые могут быть описаны при помощи локальных нечетких тенденций ВР:

- процессы I рода – класс процессов, содержащих аномалии;
- процессы II рода – класс процессов, не содержащих аномалии.

Пусть $\mathfrak{Z} = \{\tilde{\tau}_t\}$, где $t = \overline{1, m}$ – дискретный ВР, состоящий из m – локальных нечетких тенденций. Тогда, согласно формуле (1) условие обнаружения аномального поведения нечеткого ВР по компонентам нечетких локальных тенденций может быть выражены следующим образом:

$$\tilde{v}_t = \begin{cases} \tilde{v}_t^{valid}, \tilde{v}_t \notin \tilde{V}' \\ \tilde{v}_t^{invalid}, \tilde{v}_t \in \tilde{V}' \end{cases} \quad (2)$$

$$\tilde{a}_t = \begin{cases} \tilde{a}_t^{valid}, \tilde{a}_t \notin \tilde{A}' \\ \tilde{a}_t^{invalid}, \tilde{a}_t \in \tilde{A}' \end{cases} \quad (3)$$

$$\Delta \tilde{t}_t = \begin{cases} \Delta \tilde{t}_t^{valid}, \Delta \tilde{t}_t \notin \Delta \tilde{T}' \\ \Delta \tilde{t}_t^{invalid}, \Delta \tilde{t}_t \in \Delta \tilde{T}' \end{cases} \quad (4)$$

где \tilde{v}_t^{valid} – допустимое значение типа НТ \tilde{v}_t , $\tilde{v}_t^{invalid}$ – недопустимое значение типа НТ \tilde{v}_t , $\tilde{V}' = \{\tilde{v}'_1, \tilde{v}'_2, \dots, \tilde{v}'_n\}$ – множество аномальных типов НТ на ВР X;

\tilde{a}_t^{valid} – допустимое значение интенсивности НТ

$\tilde{a}_i, \tilde{a}_i^{invalid}$ – недопустимое значение интенсивности

НТ $\tilde{a}_i, \tilde{A}' = \{\tilde{a}'_1, \tilde{a}'_2, \dots, \tilde{a}'_n\}$ – множество аномальных интенсивностей НТ на ВР X;

$\Delta\tilde{t}_i^{valid}$ – допустимое значение продолжительности НТ $\Delta\tilde{t}_i, \Delta\tilde{t}_i^{invalid}$ – недопустимое значение продолжительности НТ $\Delta\tilde{t}_i, \Delta\tilde{T}' = \{\Delta\tilde{t}'_1, \Delta\tilde{t}'_2, \dots, \Delta\tilde{t}'_n\}$ – множество аномальных значений продолжительности НТ на ВР X.

В работе [Заварзин и др., 2015] был приведен пример классификации знаний об аномалиях с точки зрения анализа поведения ВР.

2. Описание алгоритма извлечения знаний об аномалии локальной тенденции ВР

В данной работе предлагается подход к извлечению знаний об аномалиях анализируя поведение локальных тенденции ВР. В отличие от других работ (например, [Заварзин и др., 2015]) в этой статье описан подход для поиска аномалий в условиях, когда множество возможных аномалий неизвестно. Идея нового подхода заключается в сравнении поведения анализируемой НЛТ относительно поведения основной НТ выраженной через свойство типа тенденции. Локальная тенденция тип которой противоречит поведению основной тенденции ВР – возможно является аномальной.

Для достижения своей цели подход использует ACL – шкалу для интерпретации локальной тенденции и её составляющих – типа, интенсивности и длительности в виде лингвистических термов. Более подробно метод разбиения при помощи ACL-шкалы описан в работе [Афанасьева, 2008].

Далее осуществляется анализ НЛТ относительно основной тенденции ВР (табл. 1).

Таблица 1 – Пример извлечения знаний об аномалиях из локальных тенденций ВР

Поведение ВР, представленное в виде типа основной тенденции		Пример	Возможный тип аномальной локальной тенденции
Тип = «стабильность»		Рис. 2	рост, падение
Тип = «рост»	ступенчатый	Рис. 3	падение
	регулярный	Рис. 4	падение
Тип «падение»	ступенчатое	Рис. 6	рост
	регулярное	Рис. 7	рост

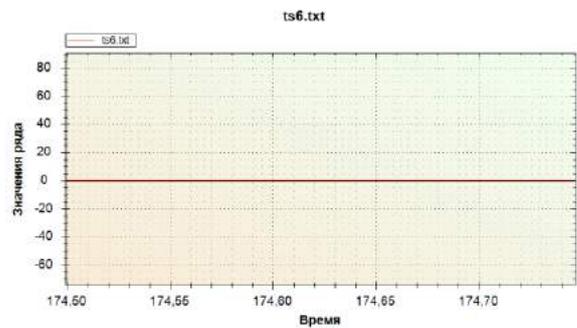


Рисунок 2 – График временного ряда, с типом основной тенденции – стабильность

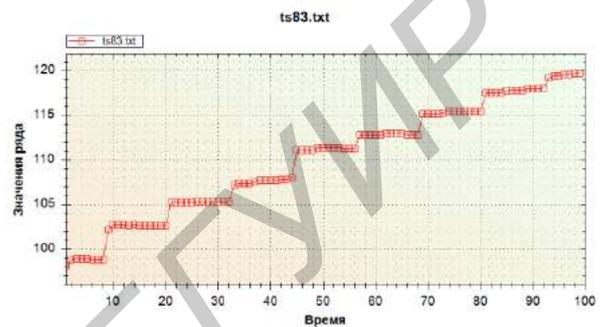


Рисунок 3 – График временного ряда, с типом основной тенденции – ступенчатый рост

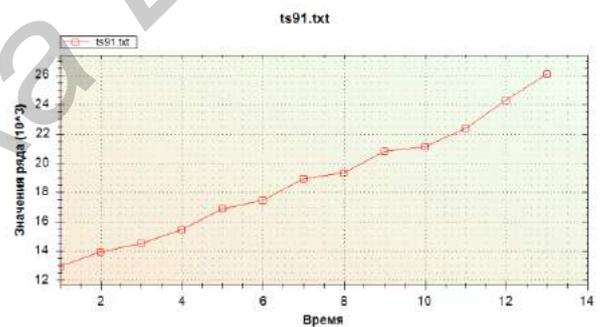


Рисунок 4 – График временного ряда, с типом основной тенденции – регулярный рост

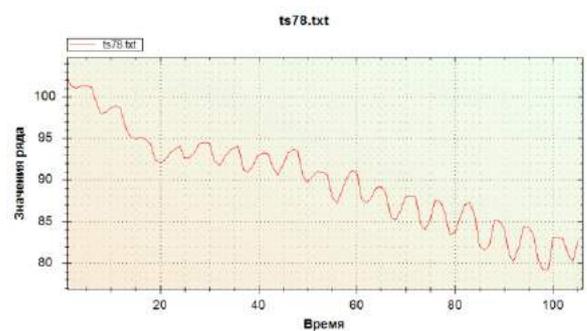


Рисунок 5 – График временного ряда, с типом основной тенденции – ступенчатое падение

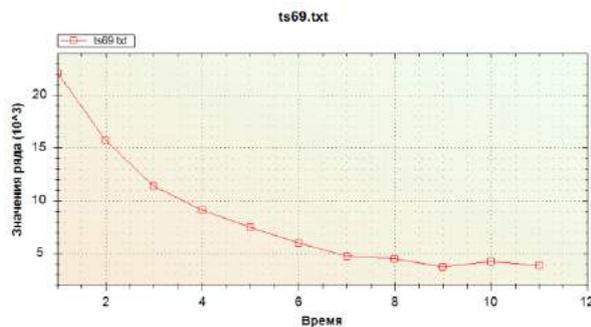


Рисунок 6 – График временного ряда, с типом основной тенденции – регулярное падение

Как видно в таблице 1, мы вполне однозначно можем утверждать об аномальности типа тенденции ВР, если она противоречит целевому типу основной тенденции. Однако, в случае с преобладающими типами «роста» и «падения» мы не можем дать утвердительной ответ на аномальность анализируемой локальной тенденции ВР. Одним из вариантов решения этой проблемы может быть более глубокая детализация целевой ЛТ, на более узкие интервалы ЛТ или элементарные тенденции.

Представление знания о локальной нечеткой тенденции может быть выражено следующим образом – отношения, задаваемые между характеристикой поведения локальными тенденциями и основной тенденцией ВР можно формализовать при помощи модели нечеткого логического вывода в виде множества нечетких правил «Если-То»:

Правило 1. Если ОТ = «стабильность» И НЛТ = «падение» ИЛИ НЛТ = «рост», ТО АЛТ = «истина»

Правило 2. Если ОТ = «ступенчатый рост» И НЛТ = «падение», ТО АЛТ = «истина».

Правило 3. Если ОТ = «регулярный рост» И НЛТ = «падение», ТО АЛТ = «истина».

Правило 4. Если ОТ = «ступенчатое падение» И НЛТ = «рост», ТО АЛТ = «истина».

Правило 5. Если ОТ = «регулярное падение» И НЛТ = «рост», ТО АЛТ = «истина».

где АЛТ – аномалия локальной тенденции ВР, ОТ – основная тенденция ВР, НЛТ – локальная нечеткая тенденция.

Используя базу нечетких правил, можно извлечь знание об аномальности анализируемого ВР.

Заключение

В настоящей работе предложен новый подход для извлечения знаний об аномалиях нечетких локальных тенденций ВР. Данный подход предполагает акцентирование внимания на анализе текущего поведения ВР, а не на сравнении с существующей моделью нормального поведения.

Данный подход, однако, в силу отсутствия экспериментального подтверждения и работы с объектами – локальными тенденциями, которые не раскрывают закономерностей поведения ВР в

полной мере может быть расширен и дополнен с учетом полной классификации, представленной в работе [Заварзин и др., 2015]. Кроме того, данный подход может быть применен при решении задачи определения адекватности прогноза исходного ВР, по анализу его локальной тенденции.

Таким образом следующий этап развития этого подхода заключается в его алгоритмизации, реализации и проведении эксперимента, идентификация трудоемкости его исполнения и быстрей действия.

Библиографический список

[Афанасьева и др., 2009] Афанасьева Т.В., Ярушкина Н.Г. Нечеткое моделирование временных рядов и анализ нечетких тенденций. – Ульяновск: УлГТУ, 2009. – 299 с.

[Афанасьева и др., 2011] Афанасьева Т.В., Ярушкина Н. Г. Нечеткий динамический процесс с нечеткими тенденциями в анализе временных рядов // Вестник Ростовского государственного университета путей сообщения, -2011, -№3. -С. 7-16.

[Афанасьева и др., 2014] Афанасьева Т.В., Заварзин Д.В. Модель локальных тенденций в задачах анализа аномалий для BIG DATA // Информатика, моделирование, автоматизация проектирования: сборник научных трудов / под ред. А. Н. Афанасьева. - Ульяновск: УлГТУ, 2014. - 228 с.

[Заварзин и др., 2015] Заварзин Д.В., Курлов А.П. Классификация аномалий временного ряда на основе анализа его тенденций // Интеллектуальные системы и технологии: современное состояние и перспективы. Сборник научных трудов III-ей Международной летней школы-семинара по искусственному интеллекту для студентов, аспирантов и молодых ученых (Тверь – Протасово, 1-5 июля 2015 г.) – Тверь: Изд-во Тверского государственного технического университета, 2015. – 234 с.

[Афанасьева, 2008] Афанасьева Т.В. Модель ACL-шкалы для генерации лингвистических оценок в принятии решений // Вопросы современной науки и практики. – Изд. Университет им. В.И. Вернадского. – №4(14). – 2008. – т. 2

DESCRIPTION APPROACH TO EXTRACTING KNOWLEDGE ABOUT GLOBAL TRENDS ANOMALIES TIME SERIES

Afansieva T.V., Zavarzin D.V.

*Ulyanovsk State Technical University,
Ulyanovsk, Russia Federation*

tv.afanasjeva@gmail.com

d.zavarzin91@gmail.com

The paper proposes a new approach for extracting knowledge of the anomalies in the data presented in the form of local fuzzy time-series trends. The work is based on the definitions of granular representation of time series, as well as classifications of anomalies and behavior properties of fuzzy BP trends described in previous papers.