

Моделирование контекста в задачах предиктивной аналитики

А.А. Романов¹, А.А. Филиппов¹

¹Ульяновский государственный технический университет, Северный Венец, 32, Ульяновск, Россия, 432027

Аннотация

Методы предиктивной аналитики приобретают особую значимость в современных системах управления. Актуальной задачей является повышение достоверности результатов моделирования в задачах прогностики. В совокупности с повышением качества прогнозов в работу методов необходимо привлекать дополнительную информацию – контекстные сведения из предметной области. Предлагаемые новые методы анализа, прогнозирования временных рядов с учетом контекста позволяют повысить эффективность решения управления за счет повышения качества принятия решений.

Ключевые слова

Предиктивная аналитика, онтология, временные ряды, контекст

1. Введение

В процессе деятельности любой современной организации возникает необходимость в своевременном принятии срочных управленческих решений, которое требует от специалиста глубокого знания особенностей проблемной области, умения использовать различные системы поддержки принятия решений и средства работы со знаниями. Желание максимально автоматизировать и ускорить процесс получения оценки состояния некоторой сложной системы вызывает потребность в методах анализа данных с учетом контекста некоторой предметной области, позволяющего учесть ее особенности и ограничения. Один и тот же набор данных в различных предметных областях будет иметь разные интерпретации и результаты анализа.

2. Применение моделей временных рядов с использованием контекста в предиктивной аналитике

Одним из распространенных типов данных в предиктивной аналитике являются временные ряды [1]. Методы моделирования и прогнозирования временных рядов нуждаются в компоненте, хранящим дополнительные знания о моделируемом процессе при решении задач аналитики и прогностики.

Для учета особенностей проблемной области при обнаружении и валидации информационных процессов планируется использовать методы инженерии знаний.

В работах [2,3] представлен подход к интеллектуальному анализу данных на основе анализа множества временных рядов, однако в данных работах не учитываются особенности проблемной области, что значительно снижает точность обнаружения процессов, так как особенности проблемной области не учитываются. Актуальной является задача исследования и разработки моделей и методов обнаружения и валидации информационных процессов на основе анализа временных рядов с использованием контекста предметной области. Развитие методов анализа данных должно вестись с опорой на 3 направления: машинное обучение (кластеризация, регрессия, анализ текстов, ...), анализ временных рядов (изменения показателей во времени), анализ контекста (работа моделей с привлечением баз знаний).

Анализ данных с учетом контекста предлагается строить по следующей схеме (рис. 1).

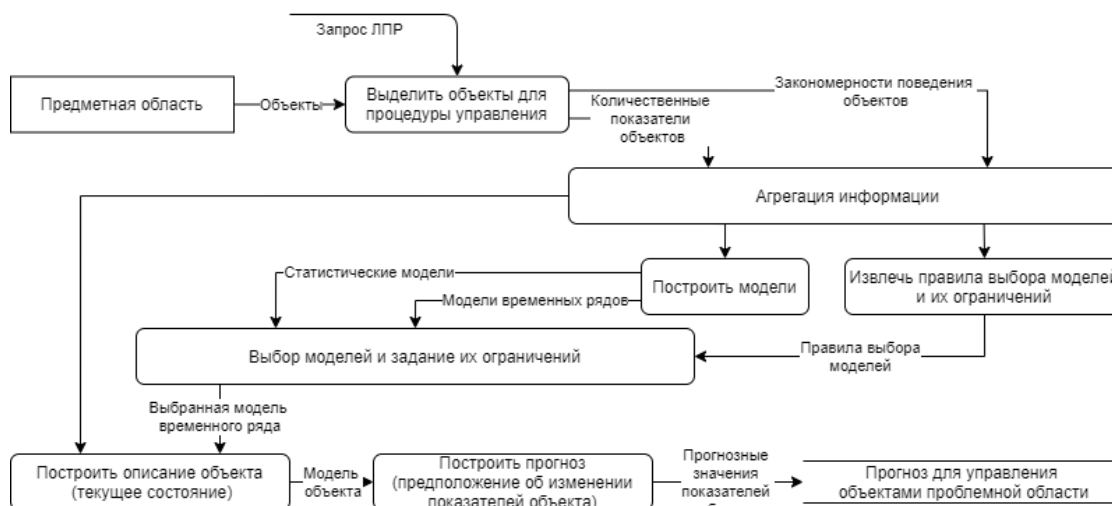


Рисунок 1: Анализ данных с использованием контекста и динамики

Контекст для анализа и прогнозирования временных рядов можно представить в виде онтологии, содержащей описание необходимых объектов рассматриваемой предметной области, их свойств и ограничений на допустимые значения свойств. Онтология может содержать логические правила и вспомогательные сущности для предписывающей аналитики. Также такая онтология может содержать описание доступных для использования методов анализа и прогнозирования временных рядов в разрезе свойств: тренд, сезонность, периодичность, длина временного ряда и т. д.

Формально онтологию для формирования контекста, используемого в процессе анализа и прогнозирования временных рядов, можно представить в виде следующего выражения:

$$Domain = \langle O, M, A, R \rangle,$$

в котором O – компонент для описания особенностей рассматриваемой предметной области; M – компонент для описания характеристик методов анализа и прогнозирования временных рядов; A – компонент для осуществления предписывающей аналитики; R – отношения между компонентами онтологии.

3. Заключение

Предлагаемый подход построения контекста и его использование в методах предиктивной аналитики позволяет учитывать ограничения предметной области и корректно интерпретировать исходные данные и результаты анализа.

4. Благодарности

Исследование выполнено при финансовой поддержке РФФИ и Правительства Ульяновской области в рамках научных проектов № 18-47-730022, 19-47-730005.

5. Литература

- [1] Ярушкина, Н.Г. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н.Г. Ярушкина, И.А. Андреев, Г.Ю. Гуськов. – Ульяновск: УлГТУ, 2020. – 220 с.
- [2] Dunkl, R. A method for analyzing time series data in process mining: application and extension of decision point analysis / R. Dunkl, S. Rinderle-Ma, W. Grossmann, K.A. Froschl // International Conference on Advanced Information Systems Engineering. – 2014. – P. 68-84.
- [3] Wang, D. A Two-Phase Development of Fuzzy Rule-Based Model and Their Analysis / D. Wang, W. Pedrycz, Z. Li // IEEE Access. – 2019. – Vol. 7. – P. 80328-80341.