

УДК 004.9:[510.23+510.67]
**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ПРОБЛЕМНОЙ
ОБЛАСТИ ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ
ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ДАННЫХ**

П.И. Сагайда^{1,2}

¹Донецкий национальный технический университет, г. Покровск

²Донбасская государственная машиностроительная академия, г. Краматорск
e-mail: pavlo.sahaida@gmail.com

Общая постановка проблемы. В условиях быстрого развития таких направлений интеллектуальной обработки данных (ИОД), как On-Line Analysis Processing, или OLAP (на основе визуализации агрегированных данных), Data Mining и Data Science (на основе методов математической статистики, Machine Learning и искусственного интеллекта) [1], недостаточно развиты теоретические основы создания информационных систем для ИОД. Выбор и использование методов и алгоритмов обработки, исследуемых параметров все еще требуют высокой квалификации привлекаемых аналитиков и инженеров по знаниям, результаты обработки данных не приводят к извлечению зависимостей достаточной специфичности и полезности для агентов организационно-технических систем (ОТС). Накопленные в данной области знания, как теоретического, так и экспериментального характера, требуют соответствующей инженерии, организации в виде, доступном для автоматизации их обработки и модификации в ходе использования [2].

Целью данной работы является повышение качества и оперативности проектирования и реализации компьютеризированных информационных систем для интеллектуальной обработки данных, организованных с использованием методов инженерии знаний за счет разработки математической модели функционирования таких систем.

Для достижения поставленной цели было выполнено математическое моделирование проблемной области компьютеризированных информационных систем (КИС) для ИОД на основе холистического (т.е. обеспечивающего своеобразие и приоритет целого над его частями) представления образа гипотетического модельного пространства проблемной области и спецификации имеющихся знаний на основе наиболее универсального математического аппарата – теории категорий. Это позволяет использовать единый подход для математического моделирования объектов и процессов проблемной области, целей и задач обработки данных, работы измерительных каналов и процессов преобразования данных в КИС, процессов инженерии знаний. Акцент при моделировании с помощью категорий на преобразования

(морфизмы) между объектами различной физической природы и на преобразования между категориями (функторы) позволяет удобно и интуитивно понятно представить преобразования данных и моделей в рассматриваемых КИС. Универсальность разработанных в теории категорий математических объектов и отображений позволяет использовать единые топологические шаблоны при моделировании различных аспектов проблемной области, а также при построении и использовании онтологических моделей – основы инженерии знаний.

Так как КИС для ИОД и их пользователям приходится иметь дело с разнообразными ОТС и различными взаимодействующими объектами, то их деятельность рассматривается в различных аспектах, в условиях действия разнообразных возмущающих воздействий и возникающих при этом помех для анализа и прогноза процессов, протекающих в ОТС. Измерительный канал КИС для ИОД также испытывает влияние возмущающих воздействий, что приводит к искажению сигнала и, соответственно, к снижению информативности данных о работе ОТС, накапливаемых в соответствующих хранилищах данных. Основные компоненты и их соотношения для проблемной области КИС для ИОД приведены на рис. 1.

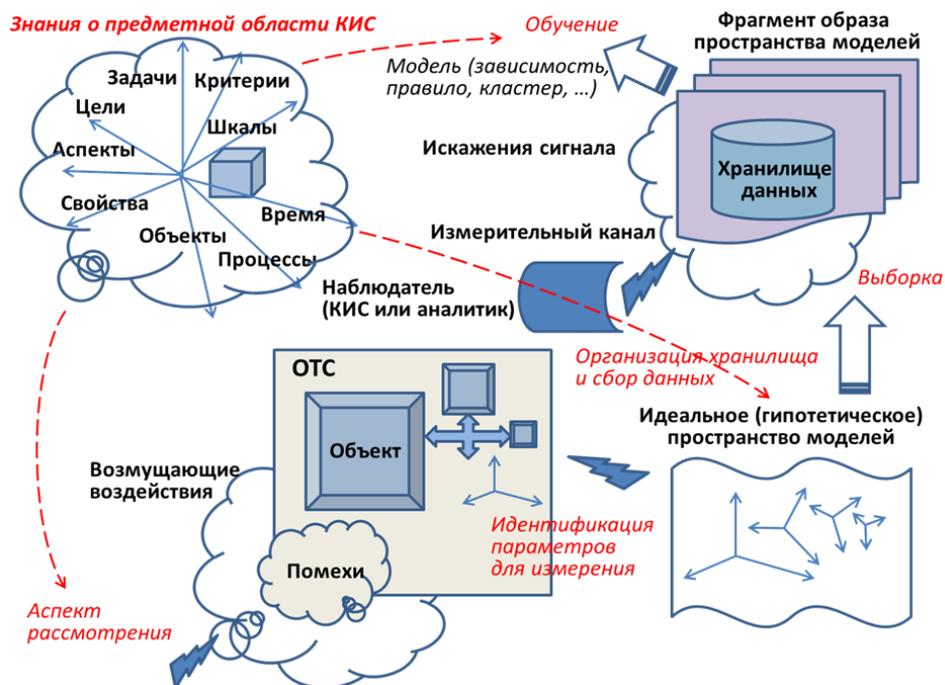


Рис. 1. Основные компоненты и их соотношения для проблемной области КИС для ИОД

Подготовленные к процессу анализа данные из этих хранилищ затем используются для обучения разнообразных моделей предметной области (Про). Прошедшие жизненный цикл в рамках процесса ИОД модели,

доказавшие свою относительную адекватность, затем используются для прогнозирования работы ПрО, поддержки принятия решений при выработке управляющих воздействий и, в конечном счете, оптимизации функционирования ПрО. Повысить качество и оперативность всего процесса извлечения моделей из данных можно, применив имеющиеся у разработчиков и аналитиков знания о процессах в ПрО, процессах, протекающих в ходе ИОД, возможных методах и алгоритмах обработки. При этом знания эти должны быть формализованы, организованы в виде хранилищ знаний и быть доступны с помощью языков запросов в реальном масштабе времени функционирования КИС для ИОД.

Все описанные выше подсистемы, процессы и преобразования в разработанной математической модели были представлены категориями и функторами, отображающими эти категории друг в друга, а также соответствующими объектами в рамках теории категорий: произведениями, копроизведениями, декартовыми и кодекартовыми квадратами над категориями и функторами. Коммутативность полученной диаграммы как результирующей модели обосновывает закономерность подхода к моделированию проблемной области и адекватность полученной модели [3].

Выводы. Разработанная модель проблемной области функционирования информационных систем для интеллектуальной обработки данных, ее представление в виде коммутативной диаграммы позволили выполнить моделирование обобщенных процессов получения и преобразования данных в компьютеризированных информационных системах на высоком уровне абстракции, в виде наиболее общего представления математической модели с использованием теории категорий. Такой подход дал возможность абстрагироваться от внутренней структуры отдельных объектов разработанных категорий проблемной области, рассмотреть функторы и их композиции, отображающие разработанные категории друг на друга, перечень и последовательность которых определяется топологическими свойствами теории категорий.

Література

1. Nettleton D. Commercial data mining: processing, analysis and modeling for predictive analytics projects / David Nettleton. – NY: Elsevier, 2014. – 288 p. - ISBN 978-0-12-416602-8
2. Feilmayr, C. An analysis of ontologies and their success factors for application to business / C. Feilmayr, W.Wöß // Data & Knowledge Engineering. – 2016. – v. 101 – pp. 1 – 23.
3. Сагайда П.И. Математическое моделирование компьютеризированных информационных систем для интеллектуальной обработки данных на основе теории категорий / П.И. Сагайда // Наукові праці ДонНТУ. Серія: «Обчислювальна техніка та автоматизація». – № 1(29). – 2016. – С. 147-157. ISSN 2075-4272