

Закключение. Из анализа исследований зарубежных и отечественных учёных инновационных методов обучения, в частности, общетехнических дисциплин, следует, что, несмотря на их многочисленность, достаточный срок разработок (начиная с 80-х гг. XX в.), глубину проработки, высокую научность и другие положительные факторы, на сегодня отсутствует единый методический инновационный подход к обучению общетехническим дисциплинам в учреждениях высшего образования технического профиля.

Рассмотренная технология инновационного обучения, реализуемая в рамках проблемно-ориентированного и контекстного обучения, междисциплинарного подхода и обучения на основе дидактического принципа единства фундаментальности и профессиональной направленности, обеспечит формирование у студентов способности к инновационной деятельности, а после окончания всего цикла обучения — подготовку высококвалифицированных специалистов, готовых к инновационной инженерной деятельности.

Эти задачи решаются путём разработки методической системы обучения техническому творчеству, которая совместно с системой обучения в олимпиадной среде, обучения во время внеаудиторной работы в научных кружках и других молодёжных студенческих творческих коллективах, а также совместно с учебным процессом в учреждении высшего образования представляет целостную систему интегрированной технологии обучения общетехническим дисциплинам.

Список цитируемых источников

1. Наумкин Н. И. Инновационные методы обучения в техническом вузе. Саранск : Изд-во Мордов. ун-та, 2007. 122 с.
2. Вербицкий А. А. Активное обучение в высшей школе: контекстный подход : метод. пособие. М. : Высш. шк., 1991 ; Лаврентьева Н. Б. Контекстное обучение как инновационная технология. Барнаул : АГУ, 1998. 156 с.
3. Лаврентьева Н. Б. Контекстное обучение как инновационная технология. Барнаул : АГУ, 1998. 156 с.
4. Вербицкий А. А. Активное обучение в высшей школе: контекстный подход.
5. Там же.
6. Там же.

УДК 004.422.632.001.2

Г. А. Исакова

*Республиканское государственное предприятие на праве хозяйственного ведения
«Евразийский национальный университет имени Л. Н. Гумилёва»,
Астана, Республика Казахстан*

МЕТОД КЛАСТЕРИЗАЦИИ АНАЛИЗА И ОБРАБОТКИ БОЛЬШИХ ДАННЫХ, ПРЕДСТАВЛЕННЫХ В МАТРИЧНОЙ ФОРМЕ

Одним из подходов анализа и обработки больших данных (BigData) является использование матричного представления данных, матричного исчисления и методов оптимизации и на этой основе разработка соответствующих инструментальных средств. Рассмотрен подход, модели и методы кластеризации больших данных, представленных в матричной форме.

One approach to the analysis and processing of large data (BigData) is to use a matrix of data, matrix calculus and methods of optimization and on this basis the development of appropriate tools. Examine approaches, models and methods for clustering large data represented in matrix form.

Введение. Парадигма BigData является мировым трендом в области информационных технологий, но к настоящему времени мало изучена. Результаты исследований в области технологии больших данных весьма востребованы как для отраслей мировой экономики, так и для реальных секторов экономики Казахстана. Сегодня ни одна отрасль и ни одна организация не могут эффективно функционировать без достижений в области информационных технологий. Курс на экономическую индустриализацию страны и максимальное удовлетворение спроса социальной сферы определяют новые вызовы и требования к информационным технологиям и обуславливают необходимость повсеместного внедрения самых современных технологий в социально-экономические системы, к которым, безусловно, относятся технологии анализа и обработки больших данных.

В существующих технологиях и продуктах, разработанных мировыми вендорами, используемые методы являются, как правило, конфиденциальными, они вложены в соответствующие коммерческие инструменты. Одной из основных проблем в использовании BigData является анализ огромных массивов данных для извлечения полезной информации для принятия решения.

В настоящее время для анализа и обработки больших данных необходима разработка теоретических основ, методов и алгоритмов в целях создания эффективных инструментов для их применения. В связи с этим указанные проблемы весьма актуальны.

Основная часть. Сегодня у ряда вендоров имеются разработки в области анализа и обработки BigData, но они не получили широкого распространения для эффективного их применения в процессе принятия решений и управления. В связи с этим возникает необходимость в новых подходах и методах, одним из которых является использование матричного представления данных, матричного исчисления и методов оптимизации, а на этой основе — разработка соответствующих инструментальных средств. Приведём основные задачи анализа и обработки BigData: анализ методов и технологий обработки больших данных; анализ особенностей и свойств структурированных и неструктурированных данных большого объёма и многообразия; разработка и исследование методов анализа больших данных, размещённых в хранилищах и представленных в матричной форме; разработка методов кластеризации неструктурированных, структурированных и гибридных данных для их эффективного представления; разработка новых подходов, методов и средств обработки больших данных; разработка инструментальных средств анализа и обработки больших данных; разработка базовых методов и моделей анализа и кластеризации на основе матричного исчисления; разработка методологических основ и рекомендаций для применения разработанных подходов, методов и инструментальных средств анализа и обработки больших данных в секторах национальных экономики.

Рассмотрим подход, модели и методы кластеризации больших данных, представленных в матричной форме для их анализа и обработки. При этом под данными можно понимать отдельные атрибуты, массивы данных, базы данных и данные в структурированных и слабоструктурированных форматах.

Предположим, что данные большого объёма представлены в матричной форме в виде одной матрицы или совокупности матриц. Сформулируем следующую задачу. Пусть $Z = \{z_i; i = \overline{1, I}\}$ — множество запросов поиска и извлечения данных $D = \{d_j; j = \overline{1, J}\}$ — множество атрибутов различных

форматов для удовлетворения запросов. Матрица $A = \|a_{ij}\|, i = \overline{1, I}; j = \overline{1, J}$ отражает необходимость поиска и извлечения j -го атрибута и матрицы больших данных для удовлетворения i -го запроса. Необходимо осуществить декомпозицию множества запросов и используемых ими данных таким образом, чтобы минимизировать взаимосвязи по данным между финальными кластерами запросов [1].

Для математической постановки задачи введём следующие переменные и обозначения.

$$\text{Пусть } x_{mi} = \begin{cases} 1, & \text{если } i\text{-й запрос распределяется в } m\text{-й кластер} \\ 0, & \text{в противном случае} \end{cases},$$

$$y_{jn} = \begin{cases} 1, & \text{если } j\text{-й запрос распределяется в } n\text{-й кластер} \\ 0, & \text{в противном случае} \end{cases}.$$

Тогда целевую функцию кластеризации запросов и данных можно сформулировать следующим образом:

$$\sum_{m=1}^M \sum_{i=1}^I \sum_{j=1}^J \sum_{n=1}^N x_{mi} a_{ij} y_{jn} \rightarrow \min$$

при ограничениях:

- на дублирование запросов в кластерах: $\sum_{m=1}^M x_{mi} = 1, i = \overline{1, I};$
- число запросов в каждом кластере запросов: $\sum_{i=1}^I x_{mi} \leq p_m; m = \overline{1, M};$
- дублирование данных для удовлетворения запросов: $\sum_{n=1}^N y_{jn} = 1, j = \overline{1, J};$
- число данных в каждом кластере данных: $\sum_{j=1}^J y_{jn} \leq S_n, n = \overline{1, N}.$

Сформулированная задача относится к блочно-симметричным задачам дискретного программирования и может быть решена с использованием алгоритма итеративных отображений [2].

Заключение. Как показал анализ технологий BigData, данное направление в самом начале развития. Результаты заключаются в разработке теоретической базы на основе матричного исчисления и множества эффективных инструментов анализа и обработки больших данных. Конечным результатом будут модели и методы, основанные на матричном исчислении, и программные инструменты, которые могут быть реализованы на современной технике и достижениях производителей вычислительных систем.

Список цитируемых источников

1. Казиев Г. З., Набиева Г. С., Даурова Б. М. Декомпозиция прикладных задач и исходных документов систем обработки данных на этапе технического проектирования // Издeніе-Поиск. Сер. естеств.-техн. наук. 2012. № 1 (1). С. 285—290.
2. Модели и методы дискретного программирования. Блочнo-симметричные модели — эффективный класс задач дискретного программирования / Г. З. Казиев [и др.] // Вестн. КБТУ. 2010. № 3. С. 60—66.