

# Особенности использования методов интеллектуального анализа данных в обучающих информационных системах

С. В. Козлов

*Аннотация* – В статье обсуждается применение методов интеллектуального анализа данных в образовательном процессе. Анализируются возможности их внедрения в учебный процесс как функциональных инструментов обучающих информационных систем. Рассматриваются особенности их использования для построения оптимальных траекторий обучения учащихся. В качестве метода интеллектуального анализа данных в таких системах предлагается методология соответствия Галуа. Автором описаны варианты ее применения в случае представления предметной области в виде графовой модели  $G$  учебного материала. На примерах графовых моделей охарактеризованы различные возможные ситуации. Классификация случаев основана на выделении множеств неусвоенных элементов знаний, ассоциированных с вершинами графа  $G$ . В качестве критериев разбиения на подмножества выбраны параметры диаметра  $d$  и количества вершин. Для каждого из случаев  $d=0$ ,  $d=1$  и  $d=2$  дана содержательная интерпретация. Также разобран случай с  $n$  вершинами диаметра  $d$  более двух. На основании выделенных ситуаций сформулировано правило разбиения на подграфы  $G_1, G_2, \dots, G_n$  графа  $G$  по неусвоенным элементам знаний для каждого учащегося. Это позволило описать алгоритм применения соответствия Галуа как метод интеллектуального анализа данных в обучающих информационных системах. Автором последовательно дана характеристика каждого из этапов алгоритма. Особое значение при этом придается формированию оптимальных стратегий обучения в соответствии с выделенными случаями применения методологии Галуа для графовой модели  $G$  изучаемого материала. Актуальность статьи связана с необходимостью сопровождения

учебного процесса с помощью обучающих информационных систем.

*Ключевые слова* – информатика, информационные системы, базы данных, программирование, информационно-коммуникационные технологии, интеллектуальный анализ данных, имплекативные матрицы, соответствие Галуа, инвариантные методы, графовая модель.

## I. ВВЕДЕНИЕ

В настоящее время процессы распространения и внедрения различных информационных систем в разные области жизнедеятельности человека широко и многообразны [1, 2, 3]. Практически не осталось такого рода человеческой деятельности, в котором была бы возможна эффективная реализация системных решений без применения возможностей информационно-коммуникационных технологий. В тоже время степень использования средств ИТ-технологий напрямую зависит от отраслевых задач. При этом в ряде случаев необходимы относительно простые сбор и накопление информации, а в других детальный анализ полученных данных [4, 5]. Здесь следует отметить, что тенденции развития информационного пространства таковы, что для решения даже первой задачи на сегодняшний день уже требуется поиск и введение в практику использования методов работы с большими объемами данных. А размеры и уровни детализации данных постоянно возрастают, что в свою очередь обуславливает развитие тех инструментов анализа, которые позволяют выявлять структурные элементы, оказывающие ключевое влияние на поведение системы в целом и отдельных ее подсистем.

Если еще несколько десятилетий тому назад в некоторых областях еще и был возможен интуитивный подход в анализе данных, то реалии сегодняшнего дня таковы, что время принятия решений неумолимо уменьшается. В ряде случаев вообще необходимо оперативное реагирование на изменение текущей ситуации в режиме реального времени. Число таких

Статья получена 24 апреля 2020.

Козлов Сергей Валерьевич, Смоленский государственный университет, доцент кафедры информатики, кандидат педагогических наук, доцент (email: svkozlov1981@yandex.ru)

информационных систем с каждым годом все увеличивается. Это обуславливает поиск таких интеллектуальных методов изучения входящего потока данных, которые могли бы сопоставлять изменяющиеся данные в динамично нередко параллельно развивающихся подсистемах информационной среды [6].

Также необходимо отметить тот факт, что структура входных данных без относительно их объема также достаточно изменчива. Это следует учитывать при выборе инструментов для анализа данных. Иными словами инструменты интеллектуального изучения структурных элементов и их характеристик должны обеспечивать синтез информации различной природы. Вообще говоря, естественный образ информации и имеет многомерную структуру [7]. И если математические методы компьютерной обработки данных XX века позволяли, как правило, обрабатывать информацию в одном срезе, то сейчас такой анализ во многом не дает оценку ситуации в полном объеме. В контексте принятия управленческих решений это может повлечь необратимые последствия в функционировании структурных элементов информационной среды. Нередко именно латентные характеристики системных компонентов оказывают решающее влияние на их поведение. При этом зачастую это особенно критично при выборе оптимального пути развития информационной системы.

Одним из примеров такого рода информационных систем могут служить автоматизированные обучающие системы в сфере образования [8, 9]. Несмотря на специфику применения каждой из них общая особенность их использования предполагает непрерывный мониторинг изучаемых компонентов образовательной среды. Следовательно, методы исследования должны позволять отслеживать и оценивать изменения в информационной системе в режиме реального времени. За выбором того или иного решения стоит оптимальный вариант функционирования и развития ее компонентной базы. В частности, одним из таких элементов являются ученики как субъекты образовательного процесса. А одна из целей обучения состоит в максимальном развитии умений и навыков каждого учащегося. Таким образом, в среде информационной системы должны быть реализованы функциональные инструменты, которые позволяют реализовать стратегию общего и индивидуального обучения оптимальным образом. Для этого программная оболочка должна иметь встроенные функции на основе интеллектуального анализа данных.

## II. ПРИМЕНЕНИЕ СООТВЕТСТВИЯ ГАЛУА КАК МЕТОДА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ОБУЧАЮЩИХ ИНФОРМАЦИОННЫХ СИСТЕМАХ

Многообразии программных средств в области образовательной деятельности на сегодняшний день достаточно велико [10, 11]. Одни имеют вспомогательный характер, например, как системы документооборота. Вторые предназначены для решения экономических задач, например, программное обеспечение для проведения закупок. С помощью третьих систематизируют личные данные субъектов образовательного процесса – учащихся и преподавателей. Назначение четвертых представление информации об учреждении образования web-средствами. Пятую категорию составляют программные средства автоматизированной поддержки учебного процесса.

Как раз среди последней группы компьютерных средств в последнее время особое значение приобретают информационные системы интеллектуального анализа данных [12, 13]. С одной стороны востребованность программных инструментов данной категории объясняется ответом на запрос о профильной подготовке учащихся оптимальным образом. С другой стороны реалии современной образовательной системы как неотъемлемой части общего информационного пространства также диктуют внедрение передовых компьютерных технологий в контексте подготовки конкурентоспособных кадров для всех сфер жизнедеятельности человека.

До недавнего времени преподаватель исходя из своего жизненного опыта, профессиональной и специальной подготовки интуитивно определял траекторию обучения учащихся. При этом, хотя он и не учитывал всех возможных факторов в условиях относительно небольших информационных объемов, он оказывался нередко прав. В настоящее время потоки информации таковы, что интуитивный подход либо невозможен в принципе, либо требует подтверждения с помощью комплексного анализа полученных эмпирическим путем данных. Для этого требуются программные комплексы анализа статистических данных. При этом очередная выборка группы учащихся может оказаться нерепрезентативной относительно заложенных в компьютерную систему сценариев развития конкретных ситуаций.

Это возможно в силу целого ряда причин. Например, выборка учащихся может оказаться в целом сильнее и слабее по тем или иным показателям в обучении. А для достижения целей профилизации образовательного процесса

усредненный подход не приемлем. Также личностные запросы учащихся могут изменяться даже на относительно малом временном промежутке. Изменения же в групповом составе учащихся практически не предусмотрены. Ученик не может произвольно варьировать потребность в уровне обучения в сопоставлении с индивидуальными запросами. Он также не может изменять темп своего обучения и многие другие параметры обучения в рамках во многом остающейся классической системы образования. Ввиду этого все эти причины обуславливают внедрение образовательный процесс автоматизированных программных комплексов с функцией интеллектуального анализа данных.

Такие информационные системы в независимости от содержательной области представления учебного материала должны иметь инвариантный способ представления данных [14]. В большинстве случаев такое представление возможно с помощью инструментов и алгоритмов графового моделирования [15]. Так учебный материал можно систематизировать по элементам знаний и связанных с ними умений и навыков решения практических задач. Элементы знаний следует объединить в единую модель, в которой они соединены друг с другом прямой или косвенной связями. Прямая связь предполагает, что для понимания данного учебного материала необходимо владение непосредственно соединенным с ним элементами знаний. Косвенная связь может показывать опосредованное отношение элементов знаний с элементами другой темы. В этом случае можно либо расширить текущую рабочую модель отдельными компонентами либо объединить две модели с помощью таких косвенно связанных элементов.

Такую ситуацию легко проиллюстрировать моделью изучения отдельной учебной темы в математике. Например, пусть учащиеся изучают тему решения квадратного уравнения или вычисления значений производной при нахождении максимального или минимального значений функции на заданном отрезке. Однако проверка решения заданий показывает, что ошибки учащиеся совершают в вычислениях из-за неуверенного знания таблицы умножения и выполнения арифметических операций. При этом учащиеся могут отлично оперировать формулами нахождения заданных математических величин.

Интуитивно выявить причины неуспеваемости учебного материала учителю оказывается достаточно сложно. В тоже время при заявленных целях обучения это следует не только делать, но выполнять оперативно.

Следовательно, автоматизированная интеллектуальная информационная система должна позволять оперировать динамично изменяющимися моделями знаний учебного материала. Она должна предусматривать возможность построения индивидуальных моделей знаний, как отдельных учащихся, так и их групп. При этом данные этих моделей в программной среде информационной системы должны исследоваться методами интеллектуального анализа данных.

Рассмотрим особенности применения соответствия Галуа как математического метода интеллектуального анализа данных на графовой модели изучаемого учебного материала [16, 17]. Пусть модель изучаемой темы – графовая модель – представлена на рисунке 1. В ней элементы знаний обозначены в виде вершин графа. Прямые связи – в виде сплошных дуг. Косвенные связи в виде дуг, обозначенных пунктирными линиями.

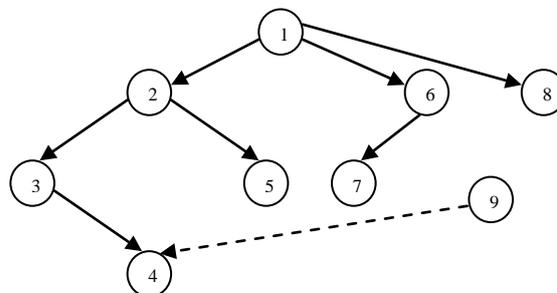


Рис. 1. Графовая модель  $G$  изучаемого материала

Обсудим возможные варианты применения соответствия Галуа на подграфе  $H$  графа  $G$ , порождённом подмножеством вершин, лежащих на расстоянии, не превышающем 1 от множества вершин  $M$ , и полустепенью захода не более 1 [18, 19].

1.  $H$  – подграф с  $n$  вершинами диаметра  $d$  не более двух.

Возможны следующие случаи – базовые. В них особое значение имеет содержательная интерпретация.

1.1.  $d = 0$ , подграф состоит из одной вершины  $X$ ,  $n = 1$ .

а) если полустепень исхода вершины  $X$  равна 0, то в этом случае необходимо проверять знания, ассоциированные с самой вершиной  $X$  и непосредственно предшествующей ей вершиной  $Y$ .

Для графа  $G$  (рис. 1) такой случай возникает, например, если необходимо проверить усвоение элемента знаний ассоциированного с вершиной 7. Подграф  $S$  графа  $H$ , знания, ассоциированные с вершинами которого подлежат проверке, содержит множество вершин  $\{6, 7\}$  (рис. 2).

$$\Gamma(3) = \{6, 7\}, \Gamma'(6, 7) = \{6, 7\} \cap \{6, 7\} = \{6, 7\}.$$

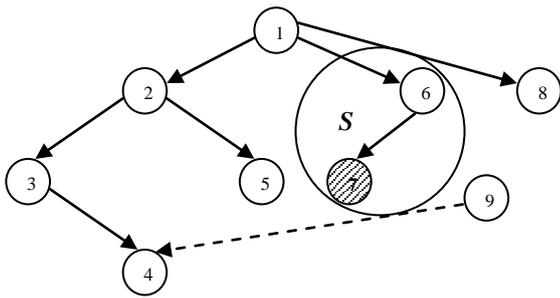


Рис. 2. Подграф  $S$  для элемента знания 7

Необходимость проверки знаний, ассоциированных не только с вершиной 7, но и с вершиной 6 обусловлена двумя причинами. Во-первых, учащийся мог действительно недостаточно отработать навыки выполнения заданий, связанные с вершиной 7. Во-вторых, он мог в недостаточной степени усвоить знания, ассоциированные с вершиной 6, и в силу данного фактора не справится с заданиями, ассоциированными с вершиной 7.

б) если полустепень исхода вершины  $X$  не равна 0, то в этом случае необходимо проверять знания, ассоциированные только с самой вершиной  $X$ .

Например, рассмотрим вершину 3 на графовой модели  $G$  изучаемого материала (рис. 3).

$$\Gamma(3) = \{2, 3, 4\}, \Gamma'(2, 3, 4) = \{2, 3\} \cap \{2, 3, 4\} \cap \{3, 4\} = \{3\}.$$

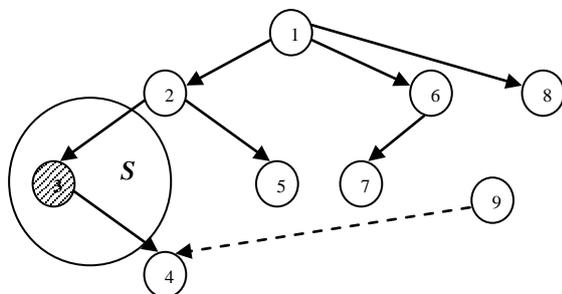


Рис. 3. Подграф  $S$  для элемента знания 3

В отличие от случая с конечной вершиной 7, в данной ситуации проверка только элемента знания 3 объясняется тем, что учащийся правильно выполнил задания, ассоциированные как с предыдущим, так и с последующим материалом. Следовательно, и его ошибка обусловлена не незнанием фактического материала, а в большей степени другими причинами, например, невнимательностью. Также некорректное выполнение заданий, связанных с усвоением знаний,

ассоциированных с вершиной 3, может быть вызвано пробелами в знаниях других тем. В таком случае необходимо проверять знания, определяемые на графе  $G$  с помощью косвенных связей.

1.2.  $d = 1$ , подграф состоит из двух вершин  $X$  и  $Y$ ,  $n = 2$ .

В этом случае необходимо проверять знания, ассоциированные с вершинами  $X$  и  $Y$ .

Так, например, выберем вершины 2 и 3 графовой модели  $G$  изучаемого материала (рис. 4).

$$\Gamma(2, 3) = \{1, 2, 3, 5\} \cap \{2, 3\} = \{2, 3\}, \Gamma'(2, 3) = \{1, 2, 3, 5\} \cap \{2, 3\} = \{2, 3\}.$$

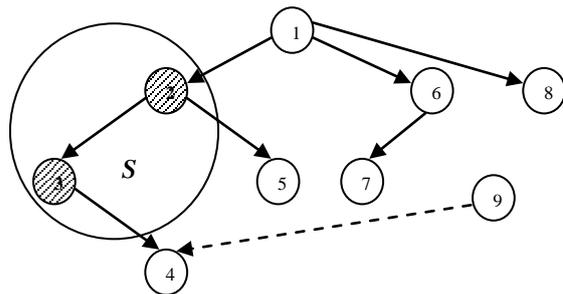


Рис. 4. Подграф  $S$  для элемента знания 2 и 3

В данном случае необходимо проверять не только знания, ассоциированные с каждой из вершин. Более важно обратить внимание на их взаимосвязь, то есть предложить учащемуся выполнить задания, которые затрагивают именно характер влияния одного элемента знаний на другой элемент. Иными словами необходимо проверка деятельности высшего порядка, затрагивающая синтезированные знания на основе двух элементов графовой модели информационной системы.

1.3.  $d = 2$ , подграф состоит из множества вершин,  $X_1, \dots, X_n, \dots$ ,  $n \geq 2$ , находящихся на расстоянии 2 друг от друга, и вершины  $A$ , находящейся на расстоянии 1 от вершин данного множества.

В этом случае необходимо проверять знания, слабо ассоциированные с вершиной  $A$  и ассоциированные с множеством вершин  $A, X_1, \dots, X_n, \dots$ . Например, рассмотрим вершины 2 и 6 графовой модели  $G$  изучаемого материала (рис. 5).

$$\Gamma(2, 6) = \{1, 2, 3, 5\} \cap \{1, 6, 7\} = \{1\}, \Gamma'(1) = \{1, 2, 6\}.$$

$$\Gamma(2, 3, 4) = \{1, 2, 3, 5\} \cap \{2, 3, 4\} \cap \{3, 4\} = \{3\}, \Gamma'(3) = \{2, 3, 4\}.$$

$$\Gamma(2, 3, 5) = \{1, 2, 3, 5\} \cap \{2, 3, 4\} \cap \{2, 5\} = \{2\}, \Gamma'(2) = \{1, 2, 3, 5\}.$$

$$\Gamma(2, 6, 8) = \{1, 2, 3, 5\} \cap \{1, 6, 7\} \cap \{1, 8\} = \{1\}, \Gamma'(1) = \{1, 2, 6, 8\}.$$

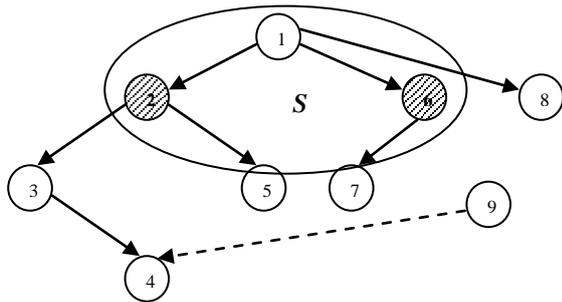


Рис. 5. Подграф  $S$  для элемента знания 2 и 6

В этом случае характер неправильного выполнения учащимся заданий, ассоциированных с вершинами 2 и 6, интуитивно определить уже достаточно сложно. Выбор вариантов предполагает анализ различных разбиений множеств элементов, ассоциированных или слабо ассоциированных с данными элементами. Учителю в совокупности для каждого отдельного ученика или группы учащихся выполнить данный анализ достаточно проблематично. При условии, что он не ошибется и не пропустит какого-либо, возможно наиболее значимого для учащегося, варианта временные рамки будут ограничивать его деятельность. Это в конечном итоге будет оказывать существенное влияние на принятие оптимальных решений.

2.  $H$  – подграф с  $n$  вершинами диаметра  $d$  более двух.

$d \geq 2$ , подграф состоит из множества вершин  $X, Y, Z$ , и т.д.,  $n \geq 2$

В этом случае отсутствие общей вершины  $A$ , находящейся на расстоянии 1 от вершин данного множества приводит к проверке знаний ассоциированных с каждой из вершин множества  $M$ .

Так, например, в рассматриваемой графовой модели  $G$  изучаемого материала можно указать множества из элементов 3 и 8 или 4 и 6.

$$\Gamma(3, 8) = \{2, 3\} \cap \{1, 8\} = \emptyset, \Gamma'(\emptyset) = M.$$

$$\Gamma(4, 6) = \{3, 4\} \cap \{1, 6, 7\} = \emptyset, \Gamma'(\emptyset) = M.$$

В подобных этим случаях анализ причин неправильного выполнения учащимися заданий, связанных с данным набором вершин, для учителя носит не только сложный неопределенный путь, но и имеет запутанный характер. Интуитивный выбор того или иного набора вершин, ассоциированных с данными, не может учесть всю вариативность разбиения множеств. При этом необходимо учитывать как прямые связи, так и косвенные связи, исследовать характер взаимосвязи между элементами, а также латентные параметры информационной системы в совокупности с иными характеристиками ее функционирования. В такой ситуации использование автоматизированных информационных систем, позволяющих применять методы

интеллектуального анализа данных, становится решающим фактором принятия оптимальных решений.

### III. АЛГОРИТМ ПРИМЕНЕНИЯ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В ПРОЦЕССЕ ФУНКЦИОНИРОВАНИЯ ОБУЧАЮЩИХ ИНФОРМАЦИОННЫХ СИСТЕМ

Процессы функционирования обучающих информационных систем достаточно многообразны. Они не сводятся только к исследованию ситуаций правильного выполнения заданий, ассоциированных с заданным набором вершин. В тоже время определение стратегии обучения учащихся является одной из ключевых функций в работе программной среды информационной системы. Без ее обоснованного выбора работа инструментов информационной системы не может быть выполнена оптимальным образом.

В связи с вышеизложенным, для конструирования индивидуального или групповых наборов заданий с помощью соответствия Галуа как метода интеллектуального анализа данных поступим следующим образом.

Разобьем всевозможными способами множество вершин  $M$  графа  $G$  на подграфы  $G_1, G_2, \dots, G_n$  так, чтобы выполнялись следующие условия:

- 1) объединение подграфов  $G_1, G_2, \dots, G_n$  есть граф  $G$ ;
- 2) попарное пересечение подграфов  $G_1, G_2, \dots, G_n$  есть пустое множество;
- 3) каждый из подграфов  $G_1, G_2, \dots, G_n$  удовлетворяет одному из случаев 1.1 – 1.3.

Такой подход оптимален для изучения всевозможных случаев выполнения заданий, ассоциированных с вершинами графовой модели  $G$  изучаемого материала. Во-первых, случаи 1.1 – 1.3 содержательно согласуются с профессиональным подходом обоснования выявленных причин неполного усвоения учебного материала. Во-вторых, они с высокой степенью корреляции соотносятся с интуитивным восприятием преподавателем образовательных процессов, протекающих в информационной обучающей системе. В-третьих, в ситуациях выделения подграфов с  $n$  вершинами диаметра  $d$  более двух, их можно путем всевозможных разбиений также свести к случаям 1.1 – 1.3. В-четвертых, исследование степени влияния косвенных связей на правильное выполнение заданий, ассоциированных с заданным набором вершин, также предполагает изучение факторов влияния

в совокупности с диаметром  $d$ , который не превышает двух.

Таким образом, алгоритм формирования индивидуальных или групповых траекторий с помощью соответствия Галуа в обучающих информационных системах может быть сведен к следующему. На первом этапе алгоритма необходимо провести диагностику входных знаний учащихся. Это необходимо для определения исходных компетенций обучаемых и позволит в дальнейшем индивидуально задать косвенные связи в графовой модели  $G$  изучаемого материала. Также на данном этапе следует выявить с помощью различных средств мониторинга мотивацию учащихся к изучаемой теме. Это обусловлено необходимостью определения целей и задач обучения и конечных результатов усвоения учебного материала.

На втором этапе алгоритма следует обеспечить возможность изучения учащимися данной темы различными методами и способами. Это можно осуществить как с помощью классических методологий организации учебного процесса, так и средствами информационно-коммуникационных технологий. Заметим, что второй подход, ввиду непрерывно возрастающих потоков информации, показывает не только свою необходимость, но и эффективность. Представление и систематизация данных в компьютерном виде позволяет дифференцировать учебный процесс в даже рамках одной группы обучения. В зависимости от параметров изучения учебного материала, определенных на первом этапе алгоритма, можно системно варьировать уровень знаний и овладения навыками в рамках данной темы. При этом выбранная траектория обучения должна изначально соответствовать логике построения базового графа  $G$  изучаемого материала. Также заметим, что в случае необходимости структуру графа  $G$  можно дополнять новыми вершинами либо изменять для укрупнения или детализации единиц учебного материала.

На третьем этапе алгоритма после изучения всей темы или ее завершенной части необходимо провести контроль знаний и умений учащихся. При этом объем знаний должен соответствовать ассоциированному с вершинами графа  $G$  либо с вершинами одного из графов  $G_1, G_2, \dots, G_n$  учебному материалу. Проверку знаний также целесообразно осуществлять в среде обучающей информационной системы. Это позволит системно накапливать данные о каждом учащемся и своевременно анализировать их в полном объеме. Только такой автоматизированный анализ с помощью компьютерных средств раскрывает дальнейшие

перспективы построения индивидуальных и групповых траекторий обучения оптимальным образом.

Четвертый этап алгоритма собственно состоит в применении соответствия Галуа как метода интеллектуального анализа данных в обучающей информационной системе. На данном этапе в соответствии с проверяемыми знаниями, ассоциированными с набором вершин графа  $G$ , выделяются неувоенные элементы изучаемого материала. Они формируют множество  $M$  вершин графа. Полученное таким образом множество  $M$  вершин графа  $G$  анализируется с точки зрения выполнения описанных выше условий. Рассматриваются всевозможные разбиения на подграфы  $G_1, G_2, \dots, G_n$  множества неувоенных элементов знаний  $M$ . При этом особое значение играет выполнение критериев из случаев 1.1 – 1.3. В соответствии с каждым из критериев полученные разбиения на подграфы  $G_1, G_2, \dots, G_n$  определяют варианты дальнейшей стратегии изучения учебного материала для каждого учащегося. Также отметим, что генерирование разбиений множества  $M$  на подграфы  $G_1, G_2, \dots, G_n$  и проверка критериев 1.1 – 1.3 выполняются в компьютерной среде обучающей информационной системы автоматизировано.

На пятом этапе соответствие Галуа применяется для анализа полученных разбиений для всех учащихся в совокупности. Это необходимо для выявления различных категорий учащихся, объединенных по тем или иным признакам неувоенных элементов знаний учебного материала. По результатам полученных данных можно выявить общие закономерности в формировании дальнейших индивидуальных и групповых стратегий обучения.

Элементы множества учеников также могут быть представлены с помощью графового моделирования. Это в свою очередь позволяет, применяя соответствие Галуа с разбиениями на множества, удовлетворяющими условиям 1.1 – 1.3, выделять учащихся, характеризующих обучение некоторой группы. По результатам обучения этих учащихся можно прогнозировать дальнейшее изучение учебного материала, ассоциированной с ними совокупности учеников. Таким образом, становится возможным генерирование различных траекторий обучения учащихся оптимальным образом. Учитываются как индивидуальные запросы, так и текущие результаты усвоения учебного материала для дальнейшего обучения. А также следующие этапы обучения могут быть уже реализованы в индивидуальном временном графике, наиболее точно отвечающем

возможностям и потребностям отдельно взятого учащегося.

Следующий, шестой этап алгоритма, представляет собой дальнейшее дифференцированное изучение учебного материала в малых группах в соответствии со сформированными по заданным параметрам траекториями обучения [2]. Далее этапы алгоритма с третьего по шестой повторяются для каждого учащегося до тех пор, пока не будут максимально достигнуты цели и задачи изучения данной темы. При этом может оказаться, что результаты обучения на очередном витке циклических этапов не показывают положительной динамики обучения. В таком случае следует обратить внимание на косвенные связи графовой модели учебного материала и выявить степень их влияния на изучение данной темы.

Причина в ряде случаев успешного усвоения учебного блока может напрямую зависеть от исходной базы знаний учащегося. В связи с этим компьютерная среда обучающей информационной системы должна предусматривать возможность изменения набора вершин графа и связей между ними. С помощью добавления новых вершин, ассоциированных с базовым учебным материалом, и косвенных связей, можно точно скорректировать индивидуальные траектории обучения учащихся. Следует также заметить, что в отдельных случаях может также понадобиться корректировка поставленных целей и задач обучения в совокупности с личностными запросами учащегося. Таким образом, алгоритм применения соответствия Галуа как метода интеллектуального анализа данных в процессе функционирования обучающих информационных систем целостно охватывает все этапы изучения выбранной для изучения темы учебного материала.

#### IV. ЗАКЛЮЧЕНИЕ

Внедрение в образовательный процесс подобных новейших средств интеллектуальных информационных технологий должно предусматривать целостный характер обучения. Использование возможностей методов интеллектуального анализа данных необходимо комплексно осуществлять на протяжении всего цикла обучения. Только в таком случае будет накапливаться кумулятивный эффект от применения такого рода интеллектуальных инструментов информационно-коммуникационных технологий. При этом подобный подход в обучении можно применять как индивидуально, например, при организации дистанционного обучения, так и в целях

дифференциации классического учебного процесса.

Особенности же применения математических методов интеллектуального анализа данных в обучающих информационных системах должны раскрывать перспективы гибкого моделирования образовательного процесса при любой форме его реализации. Особенно методы, подобные математическому аппарату соответствия Галуа, следует активно использовать в учебном процессе в целях профилизации обучения. Они позволяют не только мотивировать учащихся, но точно выявлять и более полно раскрывать их потенциальные возможности в той или иной сфере обучения.

Соответствие Галуа, как метод интеллектуального анализа данных можно использовать не только на этапе формирования индивидуальных наборов заданий при построении оптимальной траектории обучения для отдельного учащегося. Во-первых, соответствие Галуа в обучающих информационных системах, например, такой как «Advanced Tester», применяется для анализа характеристик группового обучения [20]. Компьютерная среда информационной системы хранит исходную графовую модель  $G$  изучаемого материала и индивидуализированные модели учащихся освоения ее элементов. Совокупность всех индивидуализированных моделей в разбиении, удовлетворяющих случаям 1.1 – 1.3 формирует групповые модели изучения компонентов учебного материала, ассоциированных с вершинами графа  $G$ . Это позволяет варьировать групповое обучение в рамках отдельно взятого блока данной темы.

Во-вторых, в аналогичных программных комплексах, как «Advanced Tester», интеллектуальные методы анализа данных можно использовать на этапах проектирования системных модулей [21]. Так методология соответствия Галуа совместно с инвариантами теории графов на модели учебного материала позволяет выявить характер взаимосвязей между ее элементами. При этом определенная методологически степень влияния одних элементов графовой модели на другие дает возможность скорректировать связи между группами ее компонентов, проанализировать латентные связи и характер их влияния на систему.

В-третьих, соответствие Галуа и подобные ему интеллектуальные методы анализа данных раскрывают возможности управления поведением информационной системы. Они позволяют на основе системного анализа информационных потоков выработать различные сценарии поведения [22, 23]. При

этом подобно наборам возможных заданий в обучающей подсистеме компьютерной программной среды система автоматизировано будет генерировать возможные варианты развития ситуаций в соответствии с заданными критериями. В контексте необходимости оперативного принятия эффективных решений такого рода инструменты становятся незаменимым средством моделирования системных ситуаций.

В-четвертых, представление данных в любой информационной системе многомерно. Характеристики любого объекта природы или процесса, протекающего в ней, определяют их неделимо, целостно. В связи с этим изучение только одного набора характеристик может оказаться недостаточным для их комплексного понимания. Соответствие Галуа, как методология интеллектуального анализа данных обладает такими возможностями всестороннего исследования системных объектов и процессов. Она позволяет совместно с математическим аппаратом импликативных матриц анализировать поведение объектов в информационной системе и протекающих в ней процессов по совокупностям различных элементов и их свойств [24]. Это оказывается в целом ряде случаев необходимым условием нормального функционирования системных компонентов информационной среды.

Таким образом, методы интеллектуального анализа данных, таких как соответствие Галуа, выступают инструментом не только оптимальной работы информационной системы, но и функциональным средством ее непрерывного развития. Подобные методы системного анализа являются незаменимыми во многих информационных средах. В обучающих информационных системах они должны быть реализованы как необходимые средства реализации целей и задач, стоящих перед современной образовательной средой. Только использование таких интеллектуальных методологий на основе информационно-коммуникационных технологий позволит эффективно развивать обучающие информационные системы.

#### БИБЛИОГРАФИЯ

- [1] Размахнина А. Н., Баженов Р. И. О применении экспертных систем в различных областях // Постулат. – 2017. – № 1 (15). – С. 38.
- [2] Козлов С. В. Использование функциональных возможностей информационных систем в производственной сфере // Энергетика, информатика, инновации – 2017 (электроэнергетика, электротехника и теплоэнергетика, математическое моделирование и информационные технологии в производстве): сборник трудов VII-ой Международной научно-технической конференции. В 3 томах. – Смоленск. – 2017. – Т. 1. – С. 298-301.
- [3] Кошевенко С. В., Сильченкова С. В. Структура информационной системы образовательного менеджера // Инновации на основе информационных и коммуникационных технологий. – 2015. – Т. 1. – С. 69-71.
- [4] Максимова Н. А. Моделирование информационно-образовательной среды учебного заведения // Концепт. – 2016. – № 5. – С. 195-200.
- [5] Козлов С. В. Перспективы внедрения интеллектуальных цифровых технологий в процессы управления // Цифровой регион: опыт, компетенции, проекты: сборник статей Международной научно-практической конференции (г. Брянск, 30 ноября 2018 г.) [Электронный ресурс]. – Брянск: Брян. гос. инженерно-технол. ун-т, 2018. – С. 236-240.
- [6] Мунерман В. И. Реализация параллельной обработки данных в облачных системах // Современные информационные технологии и ИТ-образование. – 2017. Т. 13. № 2. – С. 57-63.
- [7] Соколов Н. П. Введение в теорию многомерных матриц. – Киев: Наукова думка, 1972. – 176 с.
- [8] Тихонова Л. П. Актуальные вопросы разработки современных средств мониторинга и контроля качества обучения // Инновации в профессиональном и профессионально-педагогическом образовании: материалы 23-й Международной научно-практической конференции. – Екатеринбург, 2018. – С. 602-605.
- [9] Киселева О. М. Реализация принципа индивидуализации образовательного процесса с использованием программы «Траектория обучения» // Современные научные исследования и инновации. – 2014. № 5-2 (37). – С. 41.
- [10] Максимова Н. А., Гаврилова Т. И. Методические особенности применения развивающих компьютерных игр в учебном процессе // Научно-методический электронный журнал Концепт. – 2015. № 8. – С. 61-65.
- [11] Баженов Р. И., Штепа Ю. П., Шевченко Н. В. Организация проектно-исследовательской деятельности школьников средствами образовательной робототехники // Информатика в школе. – 2017. – № 10 (133). – С. 25-27.

- [12] Киселева О. М., Тимофеева Н. М., Быков А. А. Формализация элементов образовательного процесса на основе математических методов // *Современные проблемы науки и образования*. – 2013. – № 1. – С. 224.
- [13] Баженов Р. И., Лопатин Д. К. О применении современных технологий в разработке интеллектуальных систем // *Журнал научных публикаций аспирантов и докторантов*. – 2014. – № 3 (93). – С. 263-264.
- [14] Емельченков Е. П., Киселева О. М. О представлении предметных областей с помощью семантических сетей // *NovaInfo.Ru*. – 2016. – Т. 2. № 42. – С. 17-23.
- [15] Кристофидес Н. Теория графов. Алгоритмический подход. – М.: Мир, 1978. – 432 с.
- [16] Кон П. М. Универсальная алгебра / П. М. Кон; пер. с англ. Т. М. Баранович; под ред. А. Г. Куроша. – М.: Мир, 1968. – 351 с.
- [17] Парватов Н. Г. Соответствие Галуа для замкнутых классов дискретных функций // *Прикладная дискретная математика*. – 2010. – №2(8). – С. 10-15.
- [18] Козлов С. В. Интерпретация инвариантов теории графов в контексте применения соответствия Галуа при создании и сопровождении информационных систем // *International Journal of Open Information Technologies*. – 2016. – Т. 4. № 7. – С. 38-44.
- [19] Зыков А. А. Основы теории графов. – М.: Вузовская книга, 2004. – 664 с.
- [20] Козлов С. В. Математические особенности использования возможностей программного комплекса «Advanced Tester» как инструмента функционального анализа системных данных // *International Journal of Open Information Technologies*. – 2019. – Т. 7. № 2. – С. 21-30.
- [21] Ересь А. В., Баженов Р. И. Разработка web-ориентированной экспертной системы профориентационного тестирования // *Постулат*. – 2018. – № 6 (32). – С. 80.
- [22] Козлов С. В., Суин И. А. О некоторых аспектах применения инвариантных методов функционального анализа данных в различных предметных областях // *Системы компьютерной математики и их приложения*. – 2019. – № 20-1. – С. 199-205.
- [23] Муха В. С. Математические модели многомерных данных // *Доклады Белорусского государственного университета информатики и радиоэлектроники*. – 2014. – № 2 (80). – С. 143-158.
- [24] Козлов С. В. Использование математического аппарата импликативных матриц при создании и сопровождении информационных систем // *International Journal of Open Information Technologies*. – 2017. – Т. 5. № 12. – С. 16-23.

# Features of using data mining methods in training information systems

Kozlov S.V.

**Abstract** – The article discusses the application of data mining methods in the educational process. The possibilities of their implementation in the training process as functional tools of training information systems are analyzed. Features of their use to build optimal learning paths of students are considered. Galois's compliance methodology is proposed as a method of data mining in such systems. The author describes versions of its application in case of representation of the subject area in the form of a graph model  $G$  of educational material. On the examples of graph models, various possible situations are characterized. Case classification is based on the allocation of sets of non-learned knowledge elements associated with the vertices of graph  $G$ . Parameters of diameter  $d$  and number of vertices are selected as criteria of division into subsets. For each of the cases  $d = 0$ ,  $d = 1$  and  $d = 2$  a meaningful interpretation is given. The case with  $n$  vertices of diameter  $d$  of more than two is also disassembled. Based on the identified situations, a rule is formulated for the subgraphs  $G_1, G_2, \dots, G_n$  of the graph  $G$  according to unlearned knowledge elements for each student. This allowed to describe the algorithm of application of Galois correspondence as a method of data mining in training information systems. The author consistently gives a characteristic of each of the stages of the algorithm. Particular importance is attached to the formation of optimal training strategies in accordance with the identified cases of application of the Galois methodology for the graph model  $G$  of the material under study. The relevance of the article is related to the need to support the educational process with the help of training information systems.

**Keywords** – informatics, information systems, databases, programming, information and communication technologies, data mining, implicative matrixes, compliance of Galois, invariant methods, graph model.

## REFERENCES

- [1] Razmahina A. N., Bazhenov R. I. O primeneni jekspertnyh sistem v razlichnyh oblastjah // Postulat. – 2017. – # 1 (15). – S. 38.
- [2] Kozlov S. V. Ispol'zovanie funkcional'nyh vozmozhnostej informacionnyh sistem v proizvodstvennoj sfere // Jenergetika, informatika, innovacii – 2017 (jelektrojenergetika, jelektrotehnika i teplojenergetika, matematicheskoe modelirovanie i informacionnye tehnologii v proizvodstve): sbornik trudov VII-oj Mezhdunarodnoj nauchno-tehnicheskoy konferencii. V 3 tomah. – Smolensk. – 2017. – T. 1. – S. 298-301.
- [3] Koshevenko S. V., Sil'chenkova S. V. Struktura informacionnoj sistemy obrazovatel'nogo menedzhera // Innovacii na osnove informacionnyh i kommunikacionnyh tehnologij. – 2015. – T. 1. – S. 69-71.
- [4] Maksimova N. A. Modelirovanie informacionno-obrazovatel'noj sredy uchebnogo zavedenija // Koncept. – 2016. – # 5. – S. 195-200.
- [5] Kozlov S. V. Perspektivy vnedrenija intellektual'nyh cifrovyyh tehnologij v processy upravlenija // Cifrovoy region: opyt, kompetencii, proekty: sbornik statej Mezhdunarodnoj nauchno-prakticheskoy konferencii (g. Brjansk, 30 nojabrja 2018 g.) [Jelektronnyj resurs]. – Brjansk: Brjan. gos. inzhenerno-tehnol. un-t, 2018. – S. 236-240.
- [6] Munerman V. I. Realizacija parallel'noj obrabotki dannyh v oblachnyh sistemah // Sovremennye informacionnye tehnologii i IT-obrazovanie. – 2017. T. 13. # 2. – S. 57-63.
- [7] Sokolov N. P. Vvedenie v teoriju mnogomernyyh matric. – Kiev: Naukova dumka, 1972. – 176 s.
- [8] Tihonova L. P. Aktual'nye voprosy razrabotki sovremennyh sredstv monitoringa i kontrolja kachestva obuchenija // Innovacii v professional'nom i professional'no-pedagogicheskom obrazovanii: materialy 23-j Mezhdunarodnoj nauchno-prakticheskoy konferencii. – Ekaterinburg, 2018. – S. 602-605.
- [9] Kiseleva O. M. Realizacija principa individualizacii obrazovatel'nogo processa s ispol'zovaniem programmy «Traektorija obuchenija» // Sovremennye nauchnye issledovanija i innovacii. – 2014. # 5-2 (37). – S. 41.
- [10] Maksimova N. A., Gavrilova T. I. Metodicheskie osobennosti primeneniya razvivajushhih komp'yuternyyh igr v uchebnom processe // Nauchno-metodicheskij jelektronnyj zhurnal Koncept. – 2015. # 8. – S. 61-65.

- [11] Bazhenov R. I., Shtepa Ju. P., Shevchenko N. V. Organizacija proektno-issledovatel'skoj dejatel'nosti shkol'nikov sredstvami obrazovatel'noj robototehniki // Informatika v shkole. – 2017. – # 10 (133). – S. 25-27.
- [12] Kiseleva O. M., Timofeeva N. M., Bykov A. A. Formalizacija jelementov obrazovatel'nogo processa na osnove matematicheskikh metodov // Sovremennye problemy nauki i obrazovanija. – 2013. – # 1. – S. 224.
- [13] Bazhenov R. I., Lopatin D. K. O primenenii sovremennyh tehnologij v razrabotke intellektual'nyh sistem // Zhurnal nauchnyh publikacij aspirantov i doktorantov. – 2014. – # 3 (93). – S. 263-264.
- [14] Emel'chenkov E. P., Kiseleva O. M. O predstavlenii predmetnyh oblastej s pomoshh'ju semanticheskikh setej // NovaInfo.Ru. – 2016. – T. 2. # 42. – S. 17-23.
- [15] Kristofides N. Teorija grafov. Algoritmicheskij podhod. – M.: Mir, 1978. – 432 s.
- [16] Kon P. M. Universal'naja algebra / P. M. Kon; per. s angl. T. M. Baranovich; pod red. A. G. Kurosha. – M.: Mir, 1968. – 351 s.
- [17] Parvatov N. G. Sootvetstvie Galua dlja zamknutyh klassov diskretnyh funkcij // Prikladnaja diskretnaja matematika. – 2010. – #2(8). – S. 10-15.
- [18] Kozlov S. V. Interpretacija invariantov teorii grafov v kontekste primenenija sootvetstvija Galua pri sozdanii i soprovozhdenii informacionnyh sistem // International Journal of Open Information Technologies. – 2016. – T. 4. # 7. – S. 38-44.
- [19] Zykov A. A. Osnovy teorii grafov. – M: Vuzovskaja kniga, 2004. – 664 s.
- [20] Kozlov S. V. Matematicheskie osobennosti ispol'zovanija vozmozhnostej programmno kompleksa «Advanced Tester» kak instrumenta funkcional'nogo analiza sistemnyh dannyh // International Journal of Open Information Technologies. – 2019. – T. 7. # 2. – S. 21-30.
- [21] Eres' A. V., Bazhenov R. I. Razrabotka web-orientirovannoj jekspertnoj sistemy proforientacionnogo testirovanija // Postulat. – 2018. – # 6 (32). – S. 80.
- [22] Kozlov S. V., Suin I. A. O nekotoryh aspektah primenenija invariantnyh metodov funkcional'nogo analiza dannyh v razlichnyh predmetnyh oblastjah // Sistemy komp'juternoj matematiki i ih prilozhenija. – 2019. – # 20-1. – S. 199-205.
- [23] Muha V. S. Matematicheskie modeli mnogomernyh dannyh // Doklady Belorusskogo gosudarstvennogo universiteta informatiki i radioelektroniki. – 2014. – # 2 (80). – S. 143-158.
- [24] Kozlov S. V. Ispol'zovanie matematicheskogo apparata implikativnyh matric pri sozdanii i soprovozhdenii informacionnyh sistem // International Journal of Open Information Technologies. – 2017. – T. 5. # 12. – S. 16-23.