

Алгоритмы роевого интеллекта и их применение для анализа образовательных данных

Целью работы является исследование современных подходов и перспектив применения алгоритмов роевого интеллекта для анализа образовательных данных, а также возможность применения модификаций муравьиного алгоритма для организации учебного контента в адаптивных системах проведения проектных семинаров.

Материалы и методы. Проведенное исследование включало обзор современных работ в области анализа образовательных данных на основе алгоритмов роевого интеллекта, рассмотрены подходы к решению задачи построения оптимального пути обучения (оптимальной организации учебного контента) на основе муравьиного алгоритма и его модификаций, позволяющих учитывать освоение учебного материала в процессе построения обучающего пути; применение алгоритма роя частиц и его модификации на основе алгоритма Роккио для снижения размерности данных в задаче автоматической классификации вопросов; применение муравьиного алгоритма, алгоритма колонии пчел и алгоритма летучих мышей для построения систем рекомендаций при выборе учебного контента; прогнозирования успеваемости обучающихся на основе алгоритма оптимизации роя частиц. Было предложено использование модификации муравьиного алгоритма для организации учебного контента на проектных семинарах.

Результаты. В ходе работы были исследованы современные подходы к решению задач в области анализа образовательных данных на основе алгоритмов роевого интеллекта. Изучены различные подходы к обновлению феромонов (их испарению) при построении оптимального обучающего пути с учетом освоения учебного материала и поиска групп «схожих» обучающихся; исследованы возможности применения гибридных алгоритмов роевого интеллекта для построения систем рекомендаций.

На основании модификации муравьиного алгоритма предложен подход к организации учебного контента на проектных семинарах в ходе реализации проектной деятельности с учетом индивидуальных предпочтений и уровня знаний обучающихся.

Были разработаны классы на языке python: класс для обработки статистических данных; класс для реализации модификации муравьиного алгоритма, учитывающего текущий уровень знаний и заинтересованность обучающихся в изучении определенной темы на проектных семинарах, класс для вывода оптимальной последовательности тем проектных семинаров для данного обучающегося. Разработанные классы позволяют создать адаптивную систему, помогающую обучающимся 1-го курса с выбором тем проектных семинаров.

Заключение. По результатам проведенного исследования можно сделать вывод об эффективности применения алгоритмов роевого интеллекта для решения широкого круга задач, возникающих при организации учебного контента и анализе данных об обучающихся в системах электронного обучения и о перспективах разработки гибридных подходов на основе алгоритмов роевого интеллекта для реализации адаптивных систем обучения на основе парадигмы «обучение по требованию».

Полученные результаты могут быть применены для автоматизации организации учебного контента при проведении проектных семинаров для обучающихся первых курсов, когда важно понимать базовый уровень знаний и заинтересованность обучающихся в освоении новых технологий.

Ключевые слова: анализ образовательных данных, оптимизация роя частиц, муравьиный алгоритм, пчелиный алгоритм, алгоритм летучих мышей, оптимальный обучающий путь

Yulia Y. Dyulichева

V.I. Vernadsky Crimean Federal University, Simferopol, Russia

The swarm intelligence algorithms and their application for the educational data analysis

The purpose of the paper is the investigation of the modern approaches and prospects for the application of swarm intelligence algorithms for educational data analysis, as well as the possibility of using of ant algorithm modifications for organizing educational content in adaptive systems for conducting project seminars.

Materials and methods. The review of the modern articles on the educational data analysis based on swarm intelligence algorithms is provided; the approaches to solving problem of the optimal learning path construction (optimal organization of the learning objects) based on the algorithm and its modifications taking into account the students' performance in the process of the optimal learning path construction are investigated; the application of particle swarm optimization and its modification based on Roccio algorithm for the reduction of curse dimension in the problem of the auto classifying questions; the application of ant algorithm, bee colony algorithm and bat algorithm for recommender system construction are studied; the prediction of students' performance based on particle swarm optimization is researched in the article. The modification of ant

algorithm for optimal organization of learning objects at projects seminars is proposed.

Results. The modern approaches based on swarm intelligence algorithms to problem solving in educational data analysis are investigated. The various approaches to pheromones updating (their evaporation) when building the optimal learning path based on students' performance data and search of group with "similar" students are studied; the abilities of the hybrid swarm intelligence algorithms for recommendation construction are investigated.

Based on the modification of ant algorithm, the approach to the learning content organization at project seminars with individual preferences and students' level of basic knowledge is proposed. The python classes are developed: the class for statistical data processing; the class for modification of ant algorithm, taking into account the current level of knowledge and interest of student in studying a specific topic at the project seminar; the class for optimal sequence of the project seminars' topics for students. The developed classes allow creating the adaptive system that helps first year students with a choice of topics of project seminars.

Conclusion. According to the results of the study, we can conclude about the effectiveness of swarm intelligence algorithms usage to solve a wide range of tasks connected with learning content and students' data analysis in the e-learning systems and perspectives to hybrid approaches development based on swarm intelligence algorithms for realizing the adaptive learning systems on the paradigm of "demand learning".

The results can be used to automate the organization of learning content during project seminars for the first-year students, when it is important to understand the basic level of knowledge and students' interest in learning new technologies.

Keywords: educational data analysis, particle swarm optimization, ant algorithm, bee colony algorithm, bat algorithm, optimal learning path

Введение

Популярность массовых открытых онлайн-курсов и развитие систем управления обучением привели к накоплению специфичных для образовательной среды данных. Анализ образовательных данных — новая область исследования, связанная с разработкой и применением компьютерных технологий для изучения и использования уникальных типов данных, поступающих из образовательных сред, с целью лучшего понимания процесса обучения и выявления закономерностей (шаблонов) в больших объемах образовательных данных, которые вручную было бы трудно или невозможно проанализировать [1]. Применение методов машинного обучения для анализа данных в системах электронного обучения позволяет по-новому взглянуть на возможности и перспективы развития этих систем по направлению к более глубокому пониманию сущности процесса обучения и самого обучающегося; факторов, влияющих на эффективность обучения и оценивания качества процесса обучения.

Важнейшую роль в электронном обучении играют системы управления обучением (LSM). В таких системах возникают задачи персонализации обучения, автоматизации создания оптимальной последовательности объектов обучения (лекций, заданий, модулей и т.п.) для достижения поставленных целей обучения, кластеризации онлайн-курсов, онлайн-лекций и прочего обучающего контента, автоматизации построения вопросов к учебному материалу, исследо-

вания эмоционального состояния обучающегося на основе отслеживания траектории движения глаз, времени просмотра контента; исследовании эмоций на лице обучающегося по видеоизображению и оценке эмоционального состояния обучающегося по голосу, прогнозирование успеваемости обучающихся и т.п. Задачи такого типа могут быть сформулированы как задачи классификации, кластеризации или регрессии. Методы машинного обучения такие как решающие деревья, нейронные сети, метод ближайших соседей, метод опорных векторов, байесовский метод, методы нечеткой логики, ассоциативные правила, генетические алгоритмы, нечеткие алгоритмы хорошо зарекомендовали себя в области анализа образовательных данных [2, 3]. Примеры применения среды Weka как инструмента для решения задач классификации и кластеризации образовательных данных рассматриваются в [4, 5]. В работе [6] приведен обзор исследований в период с 2000 по 2017 гг., посвященный применению методов машинного обучения в компьютерной аналитике образовательных данных в общем, и, в частности, в компьютерной прогностической аналитике образовательных данных, компьютерной поведенческой аналитики в системах электронного обучения и компьютерной визуализации образовательных данных.

Интеллектуальный анализ данных на основе методов машинного обучения позволяет решать сложные задачи, связанные с прогнозированием успеваемости обучающегося на основе данных анкетиро-

вания и данных, накопленных в информационной системе приемной комиссии, с возможностью подбора одного из алгоритмов классификации или кластеризации, например, решение задачи на основе метода опорных векторов, метода k ближайших соседей, наивного байесовского классификатора, дерева решений, решающего леса [7]. История появления анализа образовательных данных как отдельного направления исследования, основанного на применении методов машинного обучения, и обзор традиционных и специфичных методов анализа образовательных данных приведены в работе [8]. Применению метаэвристических алгоритмов для решения задачи анализа образовательных данных в русскоязычной литературе уделено недостаточно внимания.

В данной работе исследуются перспективы применения алгоритмов роевого интеллекта в области анализа образовательных данных и возможность применения муравьиного алгоритма для построения оптимального обучающего пути, определяющего темы для изучения проектной деятельности с учетом коэффициента, учитывающего заинтересованность обучающегося в изучении темы и влияющего на испарение феромонов в муравьином алгоритме. Таким образом, *актуальность темы данной статьи* определяется необходимостью исследования инновационных подходов для решения задач классификации, прогнозирования и построения оптимального пути, которые могут быть сформулированы как NP-трудные задачи, возникающие при про-

ектировании и реализации различных подсистем системы электронного обучения. Многообразие и большие объемы данных, накапливаемые в системах электронного обучения, требуют поиска новых алгоритмов предобработки данных с последующей автоматической классификацией и прогнозированием. В последние годы существенно возросло число научных работ, посвященных метаэвристическим алгоритмам оптимизации, принципы работы которых основаны на идеях, «подсмотренных» у природы [9, 10]. К таким «природным» алгоритмам относятся алгоритмы роевого интеллекта. Для настройки таких алгоритмов необходимо подобрать значения параметров таких как количество итераций, размер популяции, значения феромонов (для муравьиного алгоритма), значения коэффициентов понижения громкости и повышения частоты распространения импульсов (для алгоритма летучих мышей), и определить фитнес-функцию. Решение задачи анализа образовательных данных в каждом конкретном случае требует выбора методов предобработки данных и выбора способа калибровки параметров алгоритмов роевого интеллекта. Исследование эффективности применения алгоритмов роевого интеллекта позволит проектировать приложения, связанные с анализом образовательных данных больших объемов, на основе гибридных подходов, как правило, улучшающих качество решений.

Целью работы является исследование современных подходов и перспектив применения алгоритмов роевого интеллекта для анализа образовательных данных, а также возможность применения модификации муравьиного алгоритма с учетом предпочтений обучающихся для организации учебного контента в адаптивных системах при проведении проектных семинаров.

1. Образовательные данные и некоторые типы задач в системах электронного обучения

Развитие систем электронного обучения привело к накоплению больших массивов данных о поведении обучающихся: наличие электронных портфолио студентов в web-ориентированных системах обучения, запись информации о действиях пользователей электронной системы обучения в log-файлах, отзывы обучающихся о курсах и обучающем контенте системы электронного обучения и т.п. Сбор такой информации и ее анализ связан с желанием получить обратную связь от обучающихся в электронной системе обучения с последующим корректированием обучающего контента с целью увеличения эффективности взаимодействия между обучающимся и системой электронного обучения. Построение описания объекта обучения и обучающегося, во многом, определяется используемым в дальнейшем алгоритмом. Например, для описания объекта обучения используется уникальный идентификатор id, тип файла – аудио, видео-контент, презентация, текстовый документ, ожидаемый результат обучения, критерии оценивания. Для описания обучающегося используется профайл, в котором содержится уникальный идентификатор обучающегося, данные на основе анкетирования, индивидуальные достижения, приобретенные компетенции и т. п.

Анализ образовательных данных является отдельным направлением Data Mining, поскольку, несмотря на применимость методов машинного обучения для анализа образовательных данных, это направление требует разработки собственных методов предобработки данных на этапе математического описания индивидуальных качеств

обучающихся, их поведения, предпочтений стиля обучения, получения компетенций и выработки соответствующих им профессиональных навыков и их интерпретации с точки зрения достижения поставленных целей обучения и разработки критериев оценивания качества обучения, организации сбора информации в виде обратной связи от обучающегося с целью ее последующего анализа и динамической коррекции обучающего контента и способов взаимодействия обучающегося с системой электронного обучения. Задача анализа образовательных данных является подзадачей анализа больших данных, накапливаемых в результате функционирования онлайн-систем обучения, стремительно увеличивающихся в объеме в связи с ростом количества онлайн-курсов и онлайн-школ.

Рассмотрим задачи, возникающие при анализе образовательных данных, которые могут быть решены с помощью алгоритмов роевого интеллекта.

Задача определения оптимальной последовательности объектов обучения (оптимального обучающего пути). Для анализа поведения обучающихся в электронной системе обучения необходимо организовать сбор данных об обучающемся, его предпочтениях и ожиданиях относительно результатов обучения. Сведения об обучающемся формируется либо на основе заполнения электронного портфолио либо на основе анкетирования. В системах обучения часто используется анкетирование Фелдера и Соломона, позволяющее определить стилевые параметры обучения на основе 44 вопросов [11]. Например, пользователю системы электронного обучения предлагаются следующие вопросы с вариантами ответов: «Я понимаю что-то лучше, если...» (1) попробую это; 2) продумаю до конца); «Я предпочитаю получать новую ин-

формацию на основе...» (1) рисунков, диаграмм, графиков, карт; 2) на основе письменных указаний или устной информации) и т.п. Таким образом, такое анкетирование позволяет выяснить предпочтения относительно стиля обучения на основе пяти «размерностей»: активное/рефлексивное, чувственное/интуитивное, визуальное/вербальное, последовательное/глобальное, индуктивное/дедуктивное. Использование различных видов анкетирования позволяет собрать комплексную информацию для эффективного определения индивидуального обучающего пути и построить признаковое пространство для описания предпочтений и индивидуальных характеристик каждого обучающегося в виде числового вектора. Для обеспечения качества обучения строится кривая обучения, показывающая изменение количества полученных навыков обучающегося с течением времени.

Задача построения оптимального обучающего пути связана с определением уровня знаний изученного модуля или курса и подбора дальнейших модулей или курсов для обучения в зависимости от достигнутого уровня с последующим отображением актуальной информации о приобретенных навыках в профиле обучающегося. Для построения числового вектора, описывающего навыки обучающегося, сначала используются следующие характеристики: число очков после прохождения глобального теста, время, затраченное на прохождение теста и число попыток, которые затем заменяются на характеристики по итогам изученного модуля: оценка за тест, соответствующий модулю, время, затраченное на изучение контента модуля, количество попыток для прохождения теста и время, затраченное на заполнение правильных ответов теста [12]. Разработка адаптивных

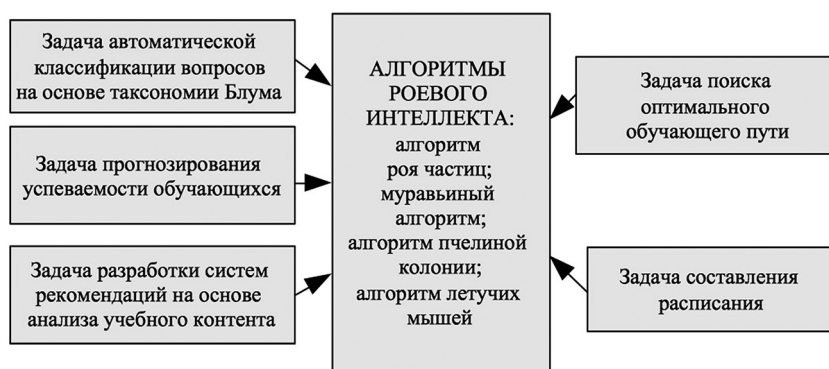


Рис. 1. Типы задач анализа образовательных данных на основе алгоритмов роевого интеллекта

веб-платформ электронного обучения связана с решением задачи подбора учебного материала в зависимости от поставленной цели. В таких системах реализуется принцип «обучения по требованию». Например, студенту необходимо в кратчайшие сроки решить задачу классификации изображений на основе имеющейся обучающей выборки. В адаптивных системах подбора учебного контента в зависимости от поставленной задачи предлагается не полное изучение курса по машинному обучению или анализу данных, а, например, изучение готового решения на основе библиотеки Scikit-learn. Таким образом, обучающийся может оперативно получить необходимый ему навык для решения поставленной задачи на основе индивидуального обучающего пути.

Задача классификации вопросов учителя согласно таксономии Блума, модели Варка или индикатора типа Майерса-Бриггса (Myers-Briggs Type Indicator). Задача автогенерации и классификации вопросов является одной из важнейших подзадач электронного обучения. Качество задаваемых вопросов, в значительной мере, определяет достижение целей обучения, дает возможность задуматься о квинтэссенции изученного учебного материала, приобрести необходимые практические навыки. Блум построил иерархию мыслительных процессов: запомина-

ние, понимание, применение, анализ, синтез и оценка. Каждому уровню мыслительного процесса можно поставить в соответствие глаголы (например, для уровня «применение» типичными являются глаголы: решите, используйте, постройте, классифицируйте и т.п.) и на их основе сформулировать вопросы [13]. Кроме таксономии Блума, популярными для выявления предпочитаемого стиля обучения являются модель Варка, которая рассматривает четыре стиля освоения информации: визуальный, вербальный, чтение/написание и кинестический (тактильный) и индикатор типа Майерса-Бриггса, позволяющий на основе анкетирования выявить личностные характеристики: экстравертный (E) / интровертный (I), сенсорный (S) / интуитивный (N), мышление (T) / чувство (F), суждение (J) / восприятие (P) [14].

Задача прогнозирования успеваемости студентов. Задача прогнозирования успеваемости студентов, понимание причин отсутствия мотивации, задача выявления групп студентов с различным уровнем начальной подготовки, выявление групп одаренных студентов и групп отстающих студентов становятся первоочередными задачами, от эффективного решения которых зависит качество обучения, улучшение которого тесно связано с персонификацией предлагаемых методик обучения.

Задача разработки системы рекомендаций учебного контента. Решение задачи разработки рекомендаций позволяет в системах обучения предлагать видеолекции, задания и другой учебный контент в зависимости от предпочтений обучающегося на основе анализа сделанного выбора другими студентами, имеющими схожие с ним предпочтения. Выделение групп схожих обучающихся осуществляется на основе матрицы оценок, которые можно получить как от пользователей системы электронного обучения непосредственно путем голосования за учебный контент, так и в результате оценивания времени, затрачиваемого обучающимся для просмотра фрагментов учебного контента. Дополнительно, анализируя время просмотра видеолекции, можно получить информацию о наиболее сложных частях лекции, которую пользователь просматривал неоднократно и получить информацию об уровне подготовки или неинтересной (или уже изученной) части видео-лекции, основываясь на тех фрагментах, которые пользователь пропускал при просмотре обучающего контента.

2. Краткое описание алгоритмов роевого интеллекта

Алгоритмы роевого интеллекта такие как муравьиные алгоритмы, алгоритмы пчелиной колонии, алгоритмы стаи рыб и т.п. основаны на изучении поведения колоний живых организмов, реализующих «оптимальное» для всей колонии поведение и эффективно зарекомендовавшие себя для решения сложных комбинаторных оптимизационных задач и, в частности, для поиска оптимальных путей в графах и сетях. Общая схема алгоритмов роевого интеллекта включает следующие основные этапы: 1) Инициализация популя-

ции (на этом этапе инициализируется (часто случайно) популяция агентов (муравьев, пчел, рыб, летучих мышей и т.п.), которые представляют собой начальные приближения к искомому решению задачи оптимизации; 2) Миграция агентов (перемещение агентов в пространстве поиска таким образом, чтобы приблизиться к искомому экстремуму оптимизируемой функции, при этом важно соблюдать компромисс между интенсификацией (скоростью сходимости алгоритма) и диверсификацией («широтой» поиска в пространстве возможных решений); 3) Завершение поиска (на каждой итерации проверяется условие останова (это может быть ограничение по времени работы алгоритма, ограничение по количеству популяций), после чего в качестве «наилучшего» предъявляется решение, описывающее наилучшее положение среди агентов популяции, при этом важную роль играет фитнес-функция, оценивающая качество агентов популяции) [17]

Метод оптимизации роя частиц основан на следующей идее. Пусть i -ая частица представляет собой числовой вектор в k -мерном пространстве $X_i = \{x_{i1}, \dots, x_{ik}\}$, определяющий положение частицы в пространстве решений — потенциальное решение. Рой представляет собой совокупность из n частиц $S = \{X_1, \dots, X_n\}$. Для каждой i -ой частицы определяется вектор скорости $V_i = \{v_{i1}, \dots, v_{ik}\}$, определяющий направление движения частицы. Метод роя частиц основан на вычислении на каждой итерации координат центра и скоростей. По сути решается задача минимизации фитнес-функции: чем меньше значение фитнес-функции на каждой итерации, тем ближе результат к оптимальному центру роя частиц. Фитнес-функция определяется как среднее значение ев-

клидовых расстояний между вектором частицы и вектором, определяющим координаты центра [15].

Идея муравьиного алгоритма заключается в вычислении на каждой итерации вероятности выбора ребра в графе, которая определяется расстоянием между вершинами и количеством феромонов на ребрах, с последующим пересчетом феромонов — значений, показывающих значимость каждого ребра в построении оптимального маршрута. Свойство испарения с течением времени феромонов является важнейшим свойством генерации различных путей для достижения цели и выбора кратчайшего из них в системе муравьиной колонии.

При инициализации пчелиного алгоритма исследуется K точек в пространстве поиска с наилучшим значением целевой функции с последующим запоминанием в улье, затем исследуется окрестность этих точек определенного радиуса и при нахождении точек с «лучшим» значением целевой функции, происходит обновление информации в улье о наилучших возможных решениях в пространстве поиска [16].

Алгоритм летучих мышей основан на моделировании природной эхолокации, позволяющей анализировать расстояние и различать добычу и препятствия. На первом этапе инициализируется популяция летучих мышей, каждая из которых характеризуется позицией, скоростью, частотой, длиной волны испускаемого импульса и громкостью для поиска добычи. Поиск локального решения осуществляется с учетом среднего значения громкости всех летучих мышей на некотором временном шаге. Если локальное решение меньше, чем лучшее решение на текущем шаге, то выполняется понижение громкости и увеличение скорости распространения импульса [17].

3. Решение задач в системах электронного обучения на основе алгоритмов роевого интеллекта

Эффективность использования муравьиных алгоритмов для реализации адаптивных систем электронного обучения, направленных на динамическое обновление контента с учетом текущих достижений обучающегося, объясняется одним из главных свойств, проявляемых в муравьиной колонии — свойство к самоорганизации. В работе [13] описывается адаптивная и персонифицированная система электронного обучения (APeLS), направленная на построение оптимального индивидуального обучающего пути. На основании электронных профайлов в этой системе генерируются курсы, которые рассматриваются как вершины графа, взаимосвязи между которыми устанавливаются на основе выбора курсов предыдущими пользователями системы электронного обучения. Для нового пользователя электронной системы обучения на основании применения муравьиного алгоритма строится оптимальная последовательность курсов (оптимальный путь в графе). Для реализации муравьиного алгоритма в системе адаптивного обучения APeLS вводится коэффициент посещаемости вершины (заинтересованности обучающимися в курсе) и интенсивность значения феромона на каждом ребре графа обновляется каждым пользователем системы электронного обучения на основе максимума интенсивности посещения вершин, образующих ребро. Сходство пути может быть определено с помощью следующего выражения $\text{learning_path}(n1, n2) = 2 * \text{dep_max}(\text{len}(n1, n2))$.

Подход к построению оптимального обучающего пути, определяемого как совокупность курсов для изучения,

может быть построен на основе алгоритма DYLPА (Dynamic Path Advisor) [18]. Алгоритм DYLPА осуществляет поиск похожих на нового пользователя обучающихся, уже воспользовавшихся электронной системой обучения, и на основе муравьиного алгоритма осуществляет поиск оптимального пути, представляющего собой последовательность курсов, среди путей, выбранных похожими на данного пользователя обучающимися. Алгоритм DYLPА позволяет рассмотреть альтернативные обучающие пути в зависимости от выбранного стиля обучения. Следующий объект обучения (вершина графа) предлагается на основе весовых коэффициентов, «пройденных» вершин графа (изученных объектов обучения). Каждый из ранее обучавшихся в системе электронного обучения рассматривается в качестве агента-муравья, «проходящего» по определенному пути из последовательности курсов. Обновление феромонов происходит на основе подбора для данного пользователя наиболее близких, по приобретенным навыкам, пользователей системы обучения, при этом вводятся весовые коэффициенты для различного учета влияния приобретенных навыков и затраченного времени на освоение объекта обучения.

При применении муравьиного алгоритма в работе [19] для учета феромонов ребрам приписывался вес с меткой S — объект изучен, F — объект не изучен, которая определялась после прохождения обучающимся вершины — объекта обучения на основании тестирования знаний, полученных при изучении объекта обучения. Для реализации уменьшения феромонов вдоль ребер используется механизм обратного распространения, при котором по мере удаленности ребра от рассматриваемой вершины вдоль пути, определяющего последовательность курсов,

знаменатель коэффициента всякий раз увеличивается на единицу, т.е. если это непосредственно предшествующее ребро для данной вершины, то коэффициент равен α , если это второе предшествующее ребро, то учитывается коэффициент $\alpha/2$ и т.п.

Усовершенствование настройки параметров и принципа обновления феромонов муравьиного алгоритма рассматривается в работе [20], в которой предлагается алгоритм ISACS (Improved the Style base Ant Colony System). В качестве обучающего объекта рассматривается пара обучающий контент c_i — задание e_j .

В таблице представлены различные подходы к вычислению (обновлению) феромонов и определению меры близости между новым обучающимся и предыдущими обучающимися системы электронного обучения.

Задача классификации вопросов на основании таксономии Блума, определяющей когнитивный уровень знаний, возникает в различных областях анализа образовательных данных. На основании таксономии Блума решается задача автоматической классификации учебного контента от контента электронных пособий до описаний обучающих мобильных приложений. Для увеличения эффективности применения метода роя частиц в задаче автоматической классификации вопросов на основе таксономии Блума применяется алгоритм Роккио, используемый для исключения влияния «проклятья» размерности данных на производительность метода роя частиц. Перед применением метода роя частиц выполняется этап предобработки текста, включающий токенизацию, стемминг, выбор термов, приписывание весовых коэффициентов термам на основе значимости для описания класса вопросов с помощью метода tfidf [21].

Таблица 1

Подходы к вычислению феромонов в муравьином алгоритме для решения задачи поиска оптимальной последовательности обучающих объектов

Авторы	Подходы к обновлению феромонов
L.-H. Wong, C.-K. Looi, 2009 г.	Вычисление меры схожести между текущим обучающимся c и предыдущими обучающимися a $S(c, a) = \left[\sum w_x (x_c - x_a)^2 + w_t (t_c - t_a)^2 \right]^{1/2} \quad (1.1),$ где w_x — весовые коэффициенты, отражающие значимость приобретенных навыков, w_t — весовые коэффициенты для учета времени обучения, позволяющие учесть «испарение» феромонов на ребре графа
S. Allach, M. Ben Ahmed, A. Ghadi, M. Essaïdi, 2012 г.	Испарение феромонов учитывается как для успешного изучения S_i объекта обучения, так и для неуспешного F_i изучения объекта обучения по формулам $S_i = T \times S_{i-1}$, $F_i = T \times F_{i-1}$, где T определяет коэффициент испарения феромонов, а x — период, с которым происходит испарение.
Dr. N. Sivakumar, R. Praveena, 2015 г.	Максимальная частота посещений вершины графа рассматривается как интенсивность феромона, а частота посещения вершины V_r вычисляется как отношение числа индивидуальных посещений вершины Nvt к общему числу вершин графа Nt $V_r = \frac{Nvt}{Nt}$
M. Rastegarmoghadam, K. Ziarati, 2016 г.	Обновление феромонов для ребра (i, j) k -м «лучшим» муравьем: $\Delta T_{ij}^k = Q \left(R_{ij} + \left(\frac{S_{ij}^k}{S_{\max}} \times \frac{1}{T_{ij}^k} \right) \right),$ где S_{ij}^k — оценка, которую получил k -й обучающийся, T_{ij}^k — время, затраченное на изучение c_i и выполнение задания e_j , $S_{\max}$ — наибольшая оценка, полученная обучающимися, «прошедшими» по ребру (i, j) , R_{ij} — коэффициент памяти, отражающий забывание учебного материала с течением времени

Рассмотрим решение задачи разработки рекомендательной системы на основе муравьиного алгоритма, пчелиного алгоритма и алгоритма летучих мышей. В работе [22] реализовано несколько алгоритмов роевого интеллекта для решения различных задач: муравьиный алгоритм — для решения задачи кластеризации обучающихся со схожими интересами относительно стиля обучения и предпочитаемых для изучения курсов, пчелиный алгоритм — для определения наилучшего обучающегося в каждом кластере студентов на основе вычисления рейтинга в системе рекомендаций, алгоритм летучих мышей — для определения полезных и неинтересных объектов обучения, основываясь на предпочтениях наилучшего студента в кластере предоставляются «наилучшие» рекомендации для обучающихся это-

го кластера. Таким образом, алгоритм коллаборативной фильтрации, основанный на алгоритмах роевого интеллекта, позволяет решать задачу персонализации предложений в системах электронного обучения.

Для извлечения скрытых закономерностей, определяющих успеваемость обучающихся, используется алгоритм на основе метода S3PSO оптимизации роя частиц в дискретном пространстве, предложенный в [23] и направленный на построение легко интерпретируемых правил классификации по типу решающих деревьев. Каждая частица описывается в виде вектора, состоящего из двух частей: часть для описания параметров и часть для контроля наличия или отсутствия признака в определенных правилах классификации и вводится фитнес-функция,

позволяющая учитывать понятность каждого правила на основе критериев поддержки, доверия и понятности.

Одним из наиболее изученных алгоритмов роевого интеллекта для решения NP-трудной задачи составления расписания является муравьиный алгоритм и его модификации. В работе [24] исследуется эффективность применения системы муравьиной колонии и максиминной муравьиной системы (MAX-MIN Ant System) к решению задачи составления расписания университетских курсов. Для реализации модификаций муравьиного алгоритма, отличающихся способом обновления феромонов, каждой упорядоченной паре событие-временной интервал приписываются феромоны. В максиминной муравьиной системе баланс между интенсификацией поиска и диверсификацией достигается за счет управления максимальной разницей между максимальным и минимальным уровнями феромона. В системе муравьиной колонии для достижения баланса между интенсификацией и диверсификацией поиска реализуются глобальное и локальное правила обновления феромонов, при этом применение локального правила обновления феромонов направлено на выбор последующими муравьями разных временных интервалов для одного и того же события из расписания.

4. Применение модификации муравьиного алгоритма для организации проектных семинаров

Начиная с 2018 года, в ФГАОУ ВО «Крымский федеральный университет имени В.И. Вернадского» была введена проектная деятельность для обучающихся первых курсов. На факультете математики и информатики Таврической академии (структурное подразделение) ФГАОУ ВО «Крым-

ский федеральный университет имени В.И. Вернадского» основным направлением проектной деятельности стало введение обучающихся в изучение современных информационных технологий на уровне рассмотрения простейших задач, не требующих глубоких знаний высшей математики. Изучение вводных лекций о современных технологиях и новых востребованных профессиях, появляющихся в связи с развитием инновационных технологий и требованиях к будущим специалистам, для которых глубокое знание математики и программирования становится определяющим фактором, дающим преимущество в конкурентной среде, направлено на повышение мотивации изучения дисциплин образовательной программы бакалавриата на факультете математики и информатики Таврической академии (структурное подразделение) ФГАОУ ВО «Крымский федеральный университет имени В.И. Вернадского».

На основе анализа работ [18,19,20], посвященных разработке адаптивных систем на основе муравьиных алгоритмов, для программной реализации определения оптимальной последовательности тем на проектных семинарах был выбран алгоритм, предложенный в [18] со следующими модификациями:

1. Построение корпуса данных на основании анкетирования обучающихся второго курса, ранее участвовавших в проектной деятельности (выявление базового уровня знаний математики и программирования, выявление тем проектных семинаров, изученных в ходе проектной деятельности, вызвавших наибольший интерес) и анкетирования обучающихся первого курса (выявление базового уровня знаний математики и программирования). Оценивание базового уровня знаний математики и програм-

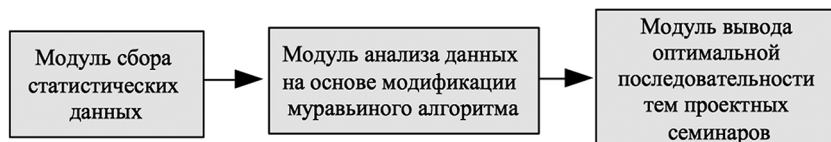


Рис. 2. Основные модули приложения для адаптивного выбора тем проектных семинаров

мирования проводится в игровой форме на основании очков, набранных при решении задач игры «Математическое домино». Эта игра позволяет не только выявить уровень знаний и оценить индивидуальный вклад обучающегося, но и получить навык командной работы.

2. Для вычисления схожести обучающегося 1-го курса и обучающихся 2-го курса, проявивших интерес при изучении тем проектных семинаров, сопоставленных вершине графа, может быть применена формула (1.1) со следующей модификацией: весовой коэффициент, отражающий субъективную заинтересованность обучающихся второго курса в изучении темы проектного семинара, «приписанной» данной вершине графа.

3. Следующая вершина рекомендуется на основании феромонов, ранее вычисленных для обучающихся 2-го курса, «схожих» по уровню знаний с обучающимся, для которого

рекомендуется следующая для изучения тема проектного семинара, при этом, чем больше значение феромонов, связанных со следующей вершиной, тем выше вероятность того, что эта вершина будет выбрана.

Архитектура приложения для адаптивного выбора тем проектных семинаров представлена на рисунке 2.

Для построения и визуализации графа использовались python-библиотека NetworkX, предназначенная для анализа сложных сетевых структур, и библиотека matplotlib. Визуализация графа и фрагмент оптимального пути для обучающегося с id = 47 представлены на рис. 3.

Предложенный подход позволяет выявить наиболее интересные темы проектных семинаров для изучения обучающимися первого курса и адаптировать проектную деятельность с учетом индивидуальных предпочтений и уровня знаний обучающихся.

student_id=47: t12->t3->t5->t7->t1->t4->t11->t2->t14->t8

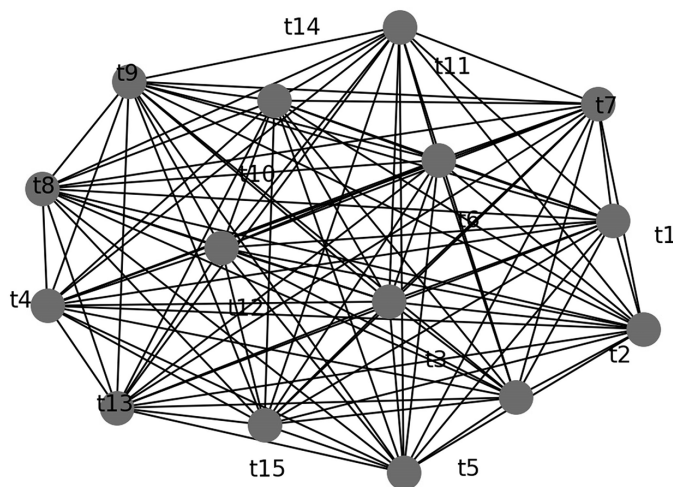


Рис. 3. Граф с вершинами, соответствующими темам проектных семинаров, и фрагмент оптимального пути для student_id=47

Заключение

Эффективный анализ образовательных данных направлен на создание систем электронного обучения нового поколения, в основу которых поставлены интересы и мотивация обучающегося, гибкие механизмы взаимодействия с обучающимся и управление учебным контентом для достижения поставленной цели обучения. Парадигма «обучения по требованию» для выживания в конкурентной среде требует от специалистов быстрого освоения новых технологий, поэтому важнейшей задачей в системах электронного обучения является задача адаптивности учебного контента для приобретения, в кратчайшие сроки, необходимых обучающемуся (специалисту) знаний

и навыков для решения поставленной задачи.

В процессе исследования в работе:

1) изучены типы задач в области анализа образовательных данных, для решения которых могут быть применены роевые алгоритмы;

2) изучены модификации муравьиного алгоритма для определения оптимальной последовательности объектов обучения с учетом текущих достижений обучающегося и информации об обучающихся и, полученных ими навыках, ранее проходивших обучение в электронной системе обучения. Рассмотрены подходы к определению испарения феромонов в муравьином алгоритме, как важного фактора для реагирования муравьиной колонии на постоянно изме-

няющиеся данные в системе электронного обучения;

3) предложена модификация муравьиного алгоритма на основании работы [9], учитывающая субъективный фактор заинтересованности обучающихся в изучении тем проектных семинаров и реализована архитектура адаптивной системы для построения оптимальной последовательности тем проектных семинаров для обучающихся 1-го курса.

Применение алгоритмов роевого интеллекта позволяет решать важные задачи персонификации предложений в системах электронного обучения на основе анализа обучающихся со схожими предпочтениями и поведением, что повышает адаптируемость и динамичность систем электронного обучения.

Литература

1. Romero C., Romero J. R. & Ventura S. A survey on pre-processing educational data. *Educational Data Mining*. Cham: Springer, 2014. P. 29–64.
2. Castro F., Vellido A., Nebot À., Mugica F. Applying Data Mining Techniques to e-Learning Problems. In: Jain L.C., Tedman R.A., Tedman D.K. (eds) *Evolution of Teaching and Learning Paradigms in Intelligent Environment*. Studies in Computational Intelligence. Heidelberg: Springer, 2007. Vol. 62. P. 183–221. DOI: https://doi.org/10.1007/978-3-540-71974-8_8
3. A.V. Manjarres, L.G.M. Sandoval, M.J.S. Suarez Data Mining Techniques Applied in Educational Environments: Literature Review // *Digital Education Review*. 2018. No.33. P. 235–266.
4. S. Lakshmi Prabha, Dr.A.R. Mohamed Shanavas Application of Educational Data mining techniques in e-Learning A Case Study // *International Journal of Computer Science and Information Technologies*. 2015. Vol.6 (5). P. 4440–4443.
5. Atta-Ur-Rahman, K. Sultan, N. Aldhafferi A. Alqahtani Educational Data Mining for Enhanced teaching and learning // *Journal of Theoretical and Applied Information Technology*. 2018. Vol. 96. No. 14. P. 4417–4427.
6. H. Aldowah, H. Al-Samarraie, W.M. Fauzy Educational data mining and learning analytics for 21st century higher education: A review and synthesis // *Telematics and Informatics*. 2019. 37. P.13–49. DOI: 10.1016/j.tele.2019.01.007.
7. Горлушкина Н.Н., Коцюба И.Ю., Хлопотов М.В. Задачи и методы интеллектуального анализа образовательных данных для поддержки принятия решений // *Образовательные технологии и общество*. 2015. Т. 18. № 1. С. 472–482.
8. Белоножко П.П., Карпенко А.П., Храмов Д.А. Анализ образовательных данных: направления и перспективы применения [Электрон. ресурс] // Интернет-журнал «НАУКОВЕДЕНИЕ». 2017. Том 9. № 4. Режим доступа: <http://naukovedenie.ru/PDF/15TVN417.pdf>
9. H.Radwan Ahmed, J.I. Glasgow Swarm Intelligence: Concepts, Models and Applications // *Technical Report 2012*. 585. 52 p. DOI: 10.13140/2.1.1320.2568.
10. A. Chakraborty, A. Kumar Kar Swarm Intelligence: A Review of Algorithms // Chapter in book *Adaptive virtual topology control based on attractor selection*. 2017. P. 475–494. DOI: 10.1007/978-3-319-50920-4_19.
11. R.M. Felder, B.A. Soloman Index of Learning Styles Questionnaire [Электрон. ресурс] // North Carolina State University. Режим доступа: <https://www.webtools.ncsu.edu/learningstyles/> (Дата обращения: 21.08.2019).
12. N.C. Benabdellah, M. Gharbi Ant colony algorithm and new pheromone to adapt units sequence to learners' profiles // *International Journal of Computer Science and Applications*. 2015. Vol. 12. No. 1. P. 26–39.
13. Dr.K. Yugandhar Interlacing Bloom's Taxonomy and E-learning to Improve the Quality of English Class // *International Journal on Studies*

in English Language and Literature. 2016. Vol. 4. Issue 1. P. 12–16.

14. Dr. N. Sivakumar, R. Praveena Determining Optimized Learning Path For an E-learning System using Ant Colony Optimization // International Journal of Computer Science & Engineering Technology. 2015. Vol. 6. No. 2. P. 61–66.

15. А.П. Карпенко Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой. Учебное пособие. 2-е изд. Москва.: МГТУ им. Н.Э. Баумана, 2017. 446 с.

16. В.М. Курейчик, А.А. Кажаров Использование пчелиных алгоритмов для решения комбинаторных задач // Искусственный интеллект. 2010. № 3. С. 583–589.

17. Selim Yilmaz Ecir U. Kucuksille A New Modification Approach on Bat Algorithm for Solving Optimization Problems // Applied Soft Computing. 2015. Vol. 28. P. 259–275. DOI: 10.1016/j.asoc.2014.11.029.

18. Lung-Hsiang Wong, Chee-Kit Looi Adaptable Learning Pathway Generation with Ant Colony Optimization // Educational Technology & Society. 2009. 12(3). P. 309–326.

19. S. Allach, M. Ben Ahmed, A. Ghadi, M. Essaïdi Modeling e-learning based on Ant Colony

algorithm // International Journal of Networks Systems. 2012. Vol. 1. No. 1. P. 37–42.

20. M. Rastegarmoghadam, K. Ziarati Improved modeling of intelligent tutoring systems using ant colony optimization // Educational Information Technology. 2016. DOI: 10.1007/s10639-016-9472-2.

21. A.A. Yahya Swarm Intelligence-based approach for educational data classification // Journal of King University. Computer and Information Sciences. 2019. 31. P. 35–51.

22. D.K. Saini, L.S. Prakash Nature Inspired Recommender Algorithms for Collaborative Web based Learning Environments // International Journal of Computer Applications. 2015. Vol. 11. No. 14. P. 16–22. DOI:10.5120/20046-2048.

23. S.M.-H. Hasheminejad, M. Sarvmili S3PSO: Students' Performance Prediction Based on Particle Swarm Optimization // Journal of AI and Data Mining. 2019. Vol. 7. No. 1. P. 77–96. DOI: 10.22044/JADM.2018.5506.1662.

24. K. Socha M. Sampels M. Manfrin Ant Algorithms for the University Course Timetabling Problem with Regard to the State-of-the-Art // Workshops on Applications of Evolutionary Computation. EvoWorkshop 2003: Applications of Evolutionary Computing. 2003. P. 334–345. DOI: 10.1007/3-540-36605-9_31.

References

1. Romero C., Romero J. R. & Ventura S. A survey on pre-processing educational data. Educational Data Mining. Cham: Springer; 2014. 29-64.

2. Castro F., Vellido A., Nebot À., Mugica F. Applying Data Mining Techniques to e-Learning Problems. In: Jain L.C., Tedman R.A., Tedman D.K. (eds) Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence. Heidelberg: Springer, 2007; 62: 183-221. DOI: https://doi.org/10.1007/978-3-540-71974-8_8

3. A.V. Manjarres, L.G.M. Sandoval, M.J.S. Suarez Data Mining Techniques Applied in Educational Environments: Literature Review. Digital Education Review. 2018; 33: 235-266.

4. S. Lakshmi Prabha, Dr.A.R. Mohamed Shanavas Application of Educational Data mining techniques in e-Learning A Case Study. International Journal of Computer Science and Information Technologies. 2015; 6 (5): 4440-4443.

5. Atta-Ur-Rahman, K. Sultan, N. Aldhafferi A. Alqahtani Educational Data Mining for Enhanced teaching and learning . Journal of Theoretical and Applied Information Technology. 2018; 96; 14: 4417-4427.

6. H. Aldowah, H. Al-Samarraie, W.M. Fauzy Educational data mining and learning analytics for 21st century higher education: A review and synthesis. Telematics and Informatics. 2019; 37: 13-49. DOI: 10.1016/j.tele.2019.01.007.

7. Gorlushkina N.N., Kotsyuba I.YU., Khlopotov M.V. Tasks and methods of intellectual analysis of educational data to support decision-making. Obrazovatel'nyye tekhnologii i obshchestvo = Educational technologies and society. 2015; 18; 1: 472-482. (In Russ.)

8. Belonozhko P.P., Karpenko A.P., Khramov D.A. Analysis of educational data: directions and prospects of application [Electron. resource] Internet-zhurnal «NAUKOVEDENIYE» = Internet journal «SCIENCE». 2017; 9; 4. Available from: <http://naukovedenie.ru/PDF/15TVN417.pdf>(In Russ.)

9. H.Radwan Ahmed, J.I. Glasgow Swarm Intelligence: Concepts, Models and Applications. Technical Report 2012; 585; 52. DOI: 10.13140/2.1.1320.2568.

10. A. Chakraborty, A. Kumar Kar Swarm Intelligence: A Review of Algorithms. Chapter in book Adaptive virtual topology control based on attractor selection. 2017: 475-494. DOI: 10.1007/978-3-319-50920-4_19.

11. R.M. Felder, B.A. Soloman Index of Learning Styles Questionnaire [Internet]. North Carolina State University. Available from: <https://www.webtools.ncsu.edu/learningstyles/> (Дата обращения: 21.08.2019).

12. N.C. Benabdellah, M. Gharbi Ant colont algorithm and new pheromone to adapt units sequence to learners' profiles. International Journal of Computer Science and Applications. 2015; 12; 1:26-39.

13. Dr.K. Yugandhar Interlacing Bloom's Taxonomy and E-learning to Improve the Quality of English Class. International Journal on Studies in English Language and Literature. 2016; 4. 1:12-16.
14. Dr. N. Sivakumar, R. Praveena Determining Optimized Learning Path For an E-learning System using Ant Colony Optimization International Journal of Computer Science & Engineering Technology. 2015; 6; 2: 61-66.
15. A.P. Karpenko Sovremennyye algoritmy poiskovoy optimizatsii. Algoritmy, vdokhnovlennyye prirodoy. Uchebnoye posobiye. 2-ye izd = Modern search engine optimization algorithms. Algorithms inspired by nature. Tutorial. 2nd ed. Moscow: MSTU im. N.E. Bauman; 2017. 446 p. (In Russ.)
16. V.M. Kureychik, A.A. Kazharov Using bee algorithms to solve combinatorial problems. Iskustvennyy intellekt = Artificial Intelligence. 2010; 3: 583-589. (In Russ.)
17. Selim Yilmaz Ecir U. Kucuksille A New Modification Approach on Bat Algorithm for Solving Optimization Problems. Applied Soft Computing. 2015; 28: 259-275. DOI: 10.1016/j.asoc.2014.11.029.
18. Lung-Hsiang Wong, Chee-Kit Looi Adaptable Learning Pathway Generation with Ant Colony Optimization. Educational Technology & Society. 2009; 12(3): 309-326.
19. S. Allach, M.Ben Ahmed, A. Ghadi, M. Essaaidi Modeling e-learning based on Ant Colony algorithm. International Journal of Networks Systems. 2012; 1; 1: 37-42.
20. M. Rastegarmoghadam, K. Ziarati Improved modeling of intelligent tutoring systems using ant colony optimization. Educational Information Technology. 2016. DOI: 10.1007/s10639-016-9472-2.
21. A.A. Yahya Swarm Intelligence-based approach for educational data classification. Journal of King University. Computer and Information Sciences. 2019; 31:35-51.
22. D.K. Saini, L.S. Prakash Nature Inspired Recommender Algorithms for Collaborative Web based Learning Environments. International Journal of Computer Applications. 2015; 11; 14: 16-22. DOI:10.5120/20046-2048.
23. S.M.-H. Hasheminejad, M. Sarvmili S3PSO: Students' Performance Prediction Based on Particle Swarm Optimization. Journal of AI and Data Mining. 2019; 7; 1: 77-96. DOI: 10.22044/JADM.2018.5506.1662.
24. K. Socha M. Sampels M. Manfrin Ant Algorithms for the University Course Timetabling Problem with Regard to the State-of-the-Art. Workshops on Applications of Evolutionary Computation. EvoWorkshop 2003: Applications of Evolutionary Computing. 2003: 334-345. DOI: 10.1007/3-540-36605-9_31.

Сведения об авторе

Юлия Юрьевна Дюlicheva

к.ф.-м.н., доцент, доцент кафедры прикладной математики

Крымский федеральный университет

имени В.И. Вернадского,

Симферополь, Россия

Эл. почта: dyulicheva_yu@mail.ru

Information about the author

Yulia Y. Dyulicheva

Cand. Sci. (Phys.-Math.), Associate Professor,
Associate Professor of the Department of Applied
Mathematics

V.I. Vernadsky Crimean Federal University,

Simferopol, Russia

E-mail: dyulicheva_yu@mail.ru