

НЕЧЕТКАЯ РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА С ХОЛОДНЫМ СТАРТОМ ДЛЯ ВЫБОРА ТРАЕКТОРИИ ОБУЧЕНИЯ

П.А. Головинский, А.О. Шаталова

Аннотация. Рассмотрены подходы к решению проблемы выбора образовательной траектории и определены преимущества использования рекомендательных систем для ее решения. Сформулирована задача холодного старта рекомендательных систем и выбран способ ее решения, основанный на создании гибридной рекомендательной системы, объединяющей нечеткую экспертную систему на основе правил и рекомендательную систему с использованием нечеткой совместной фильтрации. Общий подход реализуется применительно к проблеме выбора направления обучения на этапе поступления в высшее учебное заведение. В качестве исходных данных используется модифицированный тест Климова на выбор профессии. Приведены правила оценивания метрик и близости нечетких треугольных данных. На основе принятого нечеткого представления описываются алгоритмы нечеткой экспертной системы и нечеткой рекомендательной системы с совместной фильтрацией. Комбинация двух подходов достигается путем генерации псевдоданных с помощью экспертной системы. Это позволяет решить задачу холодного старта и получить рекомендательную систему, качество которой постепенно улучшается благодаря подстановке значений из реальных запросов пользователей в базу данных. Тестирование программ, реализующих предложенные алгоритмы, подтвердило работоспособность построенной системы нечетких рекомендаций.

Ключевые слова: экспертная система, рекомендательная система, нечеткое описание, нечеткая метрика, совместная фильтрация, холодный старт, траектория обучения.

ВВЕДЕНИЕ

Рекомендательные системы призваны быть советчиками в принятии решений, когда предстоит сделать выбор из многих альтернатив на основе анализа поведения других лиц, делавших выбор в подобных условиях [1]. Рекомендательная система подбирает и предлагает пользователю решения на основании имеющихся знаний о пользователе, пространстве решений и взаимодействии пользователя и пространства решений. Тем самым выбор решения определяется как свойствами пользователя, так и свойствами предлагаемых альтернатив. Алгоритмы таких систем отличаются по типу данных, на основе которых они формируют рекомендации. Методы, реализация которых связана с использованием данных о предыдущем применении системы, называются совместной фильтрацией. При совместной (коллоборативной) фильтрации определяется близкая группа пользователей, имеющая те же частные предпочтения, что и текущие предпочтения конкретного пользователя. Далее предпочтения пользователя расширяются до полного набора предпочтений всей близкой группы.

Контентная фильтрация опирается на метаданные, т. е. описания объектов из каталога пользователя. В зависимости от алгоритма система может формировать рекомендации путем подбора объектов, аналогичных тем, которые выбирались пользователем ранее, путем сопоставления объектов с данными из пользовательского профиля либо путем поиска схожих объектов из всего контента. Фактически цель разработки рекомендательных систем состоит в создании инструментов, позволяющих сделать выбор, максимально учитывающий предпочтения пользователя на основе данных [2]. Учитывая роль, которую в процессе выбора могут играть также психологические факторы, задача создания совершенных рекомендательных систем не только представляет несомненный интерес, но и сопряжена с преодолением существенных трудностей.

Формальная постановка задачи выглядит следующим образом. Пусть имеется множество пользователей и множество объектов. Для множества пользователей известна матрица предпочтений при выборе объектов. Задача рекомендательной системы – заполнить недостающие значения в векторе

предпочтений исследуемого пользователя, основываясь на анализе имеющихся данных матрицы рейтинга объектов для всего множества пользователей. Таким образом, в техническом отношении нужны инструменты, позволяющие заполнить пропуски в матрице. Для этого существуют различные алгоритмы [3–5], в том числе исследующие дополнительные данные об объектах и пользователях. Во многих практически важных случаях данные, на которых строится поиск рекомендации, носят неточный, размытый характер. Такие данные естественно описывать в рамках вероятностных или нечетких моделей. В то время как преимущества вероятностных моделей [6] проявляются при работе с большими данными, при малых начальных выборках для решения проблемы холодного старта приходится довольствоваться априорными экспертными оценками вероятностей, и в этом случае более релевантным становится подход, основанный на нечетких вычислениях [7].

Для нечетких рекомендательных систем развиты нечеткие варианты методов кластеризации и факторизации матрицы рейтингов [8–25]. Нечеткие экспертные рекомендательные системы достаточно удобны для разработки и применения [26–28], но в процессе их эксплуатации накопление данных не осуществляется. Обзор имеющихся решений показывает, что выбор алгоритма для рекомендательной системы существенным образом зависит от конкретной решаемой задачи.

1. СОСТОЯНИЕ ПРОБЛЕМЫ И ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ

Разнообразие и сложность больших систем предопределяет потребность в разнообразных инструментах управления ими [29], включая методы машинного обучения. В последнее десятилетие возникла задача более гибкой реакции системы образования на динамические изменения в мире и сформировалась устойчивая тенденция к персонализации траектории обучения студентов высших учебных заведений. Сложность и масштаб такой задачи требуют создания интеллектуальных систем управления [30–33], включающих в себя рекомендательные системы, облегчающие рациональный выбор учащимися направлений подготовки и наполнения их курсами, которые позволяли бы увязать текущее обучение с будущей профессиональной карьерой. Реальная потребность, стоящая за данной задачей, состоит в наличии на рынке труда большого числа незаполненных вакансий из-за несоответствия подаваемых заявлений профессиональным требованиям работодате-

лей [34]. Поэтому предпринимаются целенаправленные усилия по созданию пользовательских систем, позволяющих определить лучшую траекторию обучения для последующей успешной карьеры [35, 36]. Существует множество факторов и атрибутов, которые могут помочь в определении карьерного пути. Сюда входят навыки, отношения, принятие решений, домашний адрес, образование, работа родителей и т. д.

Характерной особенностью рассматриваемой задачи создания рекомендательной системы для образования является наличие в числе исходных параметров нечетко определенных величин, отражающих такие факторы, как предпочтения, ожидания и другие качественные оценки [37]. Существуют два разных подхода к построению рекомендательных систем на основе нечеткой логики. Первый из них заключается в создании экспертных систем на основе правил [38, 39], в то время как второй подход опирается на большие данные [40–43]. Оба этих подхода имеют свои достоинства и недостатки. Метод, базирующийся на правилах, предполагает явную формулировку правил на основе опыта экспертов [44]. При этом процесс эксплуатации системы сам по себе никак не сказывается на ее качестве. Методы построения рекомендательных систем, основанные на машинном обучении, предполагают, что правила принятия решений заранее неизвестны, но требуют больших данных. Это означает, что если исходных данных в базе мало, то рекомендательная система, основанная на обучении, не способна сформулировать релевантные правила вывода. Это ограничение известно как проблема холодного старта [45]. В то же время, по мере увеличения количества данных качество такой рекомендательной системы будет расти, все в большей степени отражая реальные соотношения и связи. Цель настоящей работы состоит в преодолении ограничений, свойственных каждому из этих подходов, путем их объединения. Таким образом, в начальной стадии рекомендательная система будет работать на основе правил, генерируя соответствующую базу данных. По мере роста количества данных имитационные примеры в ней будут замещаться реальными данными и рекомендательная система перестроится на результаты машинного обучения, преодолевая трудности холодного старта.

2. МЕТОДЫ И ДАННЫЕ

Проблема выбора траектории обучения с учетом будущей карьеры в многоступенчатой системе образования возникает несколько раз: на этапе



определения направления подготовки и профиля образовательной программы при поступлении в высшее учебное заведение, при выборе магистратуры в случае продолжения высшего образования и, наконец, при поступлении в аспирантуру. Индивидуальная траектория обучения может также детализироваться с помощью набора элективных курсов. В настоящей работе предлагается рекомендательная система [46], которая может применяться на разных ступенях образования и способна адаптироваться к меняющейся структуре и приоритетам российского образования. Для обеспечения холодного старта системы используется гибридная нечеткая рекомендательная система, объединяющая экспертную систему, работающую на правилах, и рекомендательную систему, основанную на совместной фильтрации характеристик пользователя [9].

В качестве основы для апробации подхода выбрана методика Е.А. Климова «Определение типа будущей профессии» [44], в которой четкие оцен-

ки заменены нечеткими. Методика основана на классификации профессиональных интересов. Она позволяет установить, в какой области учащемуся лучше выбрать будущую специальность, по которой он будет проходить дальнейшее обучение. В соответствии с данной классификацией мир современных профессий можно разделить на пять основных типов: «человек – природа» (П – работа с природными объектами), «человек – техника» (Т – работа с техническими устройствами), «человек – знаковая система» (З – работа с абстрактными символическими системами, моделями, естественными и искусственными языками), «человек – искусство» (Х – творческие профессии), «человек – человек» (Ч – профессии, предполагающие непосредственное общение и взаимодействие с людьми). В таблице приведен тест [44], по результатам которого система принимает решение.

По каждому пункту таблицы проводится оценка по пятибалльной нечеткой шкале, результаты которой умножаются на вес, указанный в соответ-

Тест Е.А. Климова на выбор профессии

№ п/п	Характеристика пользователя	Тип профессий				
		П	Т	З	Х	Ч
1	Легко знакомлюсь с людьми	–	–	–	–	1
2	Охотно и подолгу могу что-нибудь мастерить	–	1	–	–	–
3	Люблю ходить в музеи, театры, на выставки	–	–	–	1	–
4	Охотно и постоянно ухаживаю за растениями, животными	1	–	–	–	–
5	Охотно и подолгу могу что-нибудь вычислять, чертить	–	–	1	–	–
6	С удовольствием общаюсь со сверстниками или малышами	–	–	–	–	1
7	С удовольствием ухаживаю за растениями и животными	1	–	–	–	–
8	Обычно делаю мало ошибок в письменных работах	–	–	1	–	–
9	Мои изделия обычно вызывают интерес у товарищей, старших	–	2	–	–	–
10	Люди считают, что у меня есть художественные способности	–	–	–	2	–
11	Охотно читаю о растениях, животных	1	–	–	–	–
12	Принимаю участие в спектаклях, концертах	–	–	–	1	–
13	Люблю читать об устройстве механизмов, приборов, машин	–	1	–	–	–
14	Подолгу могу разгадывать головоломки, задачи, ребусы	–	–	2	–	–
15	Легко улаживаю разногласия между людьми	–	–	–	–	2
16	Считают, что у меня есть способности к работе с техникой	–	2	–	–	–
17	Любят мое художественное творчество	–	–	–	2	–
18	У меня есть способности к работе с растениями и животными	2	–	–	–	–
19	Я могу ясно излагать свои мысли в письменной форме	–	–	2	–	–
20	Я почти никогда ни с кем не ссорюсь	–	–	–	–	1
21	Результаты моего технического творчества одобряют даже незнакомые люди	–	1	–	–	–
22	Без особого труда усваиваю иностранные языки	–	–	1	–	–
23	Мне часто случается помогать даже незнакомым людям	–	–	–	–	2
24	Подолгу могу заниматься музыкой, рисованием, читать книги и т. д.	–	–	–	1	–
25	Могу влиять на ход развития растений и животных	2	–	–	–	–
26	Люблю разбираться в устройстве механизмов, приборов	–	1	–	–	–
27	Мне обычно удается убедить людей в своей правоте	–	–	–	–	1
28	Охотно наблюдаю за растениями или животными	1	–	–	–	–
29	Охотно читаю научно-популярную, критическую литературу, публицистику	–	–	1	–	–
30	Стараюсь понять секреты мастерства и пробую свои силы в живописи, музыке и т. п.	–	–	–	1	–
Результаты		8	8	8	7	8

ствующей графе. Указывается действительное число в интервале от 0 до 5, адекватно отражающее собственную оценку ответа пользователя на каждый тестовый вопрос. Тест Климова предлагается дополнить средним баллом дисциплин аттестата по областям интересов, учитывая таким образом не только субъективные оценки пользователя, но и доступную точную и проверяемую внешнюю информацию с весом 2:

- тип П: биология, география, основы безопасности жизнедеятельности, информатика и ИКТ;
- тип Т: алгебра и начала анализа, геометрия, физика, информатика и ИКТ;
- тип З: алгебра и начала анализа, геометрия, русский язык, информатика и ИКТ;
- тип Х: русский язык, литература, всеобщая история, история России;
- тип Ч: обществознание, всеобщая история, история России, основы безопасности жизнедеятельности.

В результате формируется вектор нечетких данных, на основе выбора максимальной длины которых следует сделать вывод о предпочтении одного из пяти типов профессий. Выбор длины векторов нечетких оценок вместо простой суммы признаков позволяет увеличить вклад в оценку наиболее выраженных признаков. Как и в работе [9], используем треугольные нечеткие числа \tilde{a} с функцией принадлежности, определяемой с помощью триплета чисел (a^-, a, a^+) соотношением

$$\mu_{\tilde{a}}(x) = \begin{cases} 0, & x < a^-, \\ \frac{x - a^-}{a - a^-}, & a^- \leq x \leq a, \\ \frac{a^+ - x}{a^+ - a}, & a \leq x \leq a^+, \\ 0, & a^+ < x, \end{cases} \quad (1)$$

где μ – функция принадлежности, $x \in X$.

При выходе интервала задания нечеткости за границы отрезка $[c, d]$ определения нечетких параметров слева следует задать $a^- = c$, а справа задать $a^+ = d$.

Для треугольных нечетких чисел арифметические операции определяются простыми правилами преобразования триплетов [47, 48]

$$\tilde{a} + \tilde{b} = (a^- + b^-, a + b, a^+ + b^+),$$

$$\tilde{a} \cdot \tilde{b} = (a^- b^-, ab, a^+ b^+),$$

$$\tilde{a} - \tilde{b} = (a^- - b^+, a - b, a^+ - b^-),$$

$$\frac{\tilde{a}}{\tilde{b}} = \left(\frac{a^-}{b^+}, \frac{a}{b}, \frac{a^+}{b^-} \right).$$

Расстояние между нечеткими числами определяется соотношением [47]

$$d(\tilde{a}, \tilde{b}) = \sqrt{\frac{1}{3} \left[(a^- - b^-)^2 + (a - b)^2 + (a^+ - b^+)^2 \right]}. \quad (2)$$

Евклидово расстояние между нечеткими векторами $\tilde{\mathbf{a}}, \tilde{\mathbf{b}}$ [49] рассчитывается аналогичным образом с заменой квадратичной суммы для одной пары триплетов в уравнении (2) суммой квадратов разностей триплетов нечетких координат

$$d(\tilde{\mathbf{a}}, \tilde{\mathbf{b}}) = \sqrt{\frac{1}{3} \sum_{jk} (a_j^k - b_j^k)^2}. \quad (3)$$

Здесь с помощью индекса j нумеруются векторы, а с помощью индекса k – параметры триплета. Для точных чисел выражение (3) совпадает с формулой для обычного расстояния в евклидовом пространстве. Приведенных соотношений достаточно для обработки нечетких данных тестирования и выбора приоритетного решения.

Экспертная система, построенная на нечетких параметрах, позволяет делать заключения на основе данных, полученных от пользователя, но не учитывает историю запросов, которая может содержать информацию, отличающуюся от экспертных оценок. Одной из причин такого отличия может стать наличие существенного скрытого фактора, который отсутствовал или не был учтен во время создания экспертной системы, а затем приводит к систематическому искажению результата.

Для учета информации, которая содержится в истории запросов с учетом принятых решений, выбрана нечеткая система совместной фильтрации данных, основанная на косинусной мере сходства нечетких данных

$$\text{sim}(\mathbf{a}, \mathbf{b}) = \frac{\sum_{jk} a_j^k b_j^k}{\sqrt{\sum_{jk} (a_j^k)^2 \sum_{jk} (b_j^k)^2}}.$$

Этот критерий аналогичен классическому коэффициенту корреляции Пирсона между векторами и совпадает с ним при переходе к четким данным. Максимальное значение, равное единице, он принимает при совпадении направлений сравниваемых векторов. Это свойство обеспечивает поиск в базе данных прецедента, наиболее близкого к рассматриваемому случаю. Для небольших баз дан-



ных такое решение является наилучшим. В случае больших данных целесообразно предварительно провести кластеризацию данных и принимать решение путем сравнения с центрами кластеров.

3. ПРЕДЛАГАЕМЫЙ АЛГОРИТМ И РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Для программной реализации нечеткой рекомендательной системы с холодным запуском требуются алгоритмы работы экспертной системы, основанной на правилах, рекомендательной системы, построенной на совместной фильтрации, и их сопряжение, генерирующее псевдоданные на основе экспертной системы и заданной статистики принимаемых решений.

Алгоритм нечеткой экспертной системы, основанной на свойствах пользователя и правилах, имеет следующий вид.

Шаг 1. Ввод данных, характеризующих пользователя.

Шаг 2. Преобразование четких данных в нечеткие.

Шаг 3. Определение на основании нечеткого представления и правил весов решений как модулей нечетких векторов характеристических показателей, относящихся к разным решениям.

Шаг 4. Нормировка весов решений, их сравнение и вывод результата в виде ранжированных рейтингов рекомендаций.

Для генерации базы псевдоданных воспользуемся случайной генерацией пяти блоков таких данных, в каждом из которых преобладает тот или иной тип профессиональных предпочтений, в то время как остальные решения присутствуют в сниженной, но близкой по величине пропорции. Вариации оценок отдельных свойств внутри данного типа меньше, чем отличие между главным типом и второстепенными типами для блока.

Алгоритм нечеткой рекомендательной системы, построенной на совместной фильтрации запроса пользователя и исторических данных о запросах в базе данных, задает следующую последовательность действий.

Шаг 1. Формирование базы данных и ее загрузка.

Шаг 2. Преобразование показателей в нечеткие характеристики.

Шаг 3. Введение данных текущего пользователя.

Шаг 4. Преобразование четких данных текущего пользователя в нечеткие.

Шаг 5. Поиск в базе нечетких данных наиболее близкого экземпляра.

Шаг 6. Вывод решения для найденного экземпляра в качестве рекомендуемого.

Для запуска алгоритма совместной фильтрации необходимо иметь релевантную базу данных. Она с большой долей вероятности может отсутствовать в начальный период эксплуатации системы, создавая известную проблему холодного старта. Для решения этой задачи с помощью нечеткой экспертной системы, основанной на правилах, генерируем исходную базу псевдоданных, которая отражает наблюдаемую статистику принимаемых пользователями решений. Далее в процессе эксплуатации второй рекомендательной системы, построенной на совместной фильтрации, псевдоданные постепенно заменяются данными, взятыми из реальных примеров. Таким образом, релевантность всей системы в процессе эксплуатации повышается.

Перед началом эксплуатации системы необходимо проверить качество базы псевдоданных путем сравнения ответов, которые выдают на одинаковых тестовых примерах обе рекомендательные системы. Интерфейс программ, реализующих в пакете MATLAB нечеткую экспертную систему FuzzyExpert, основанную на правилах, и нечеткую рекомендательную систему FuzzyRecommend, в которой используется совместная фильтрация, одинаковый и представляет собой столбец меню.

Приведем пример матрицы A входных данных, распределенных по строкам в соответствии с типом профессии:

$$A = \begin{pmatrix} 2,00 & 2,00 & 4,00 & 2,00 & 2,00 & 3,00 & 5,00 \\ 4,00 & 4,00 & 4,00 & 4,00 & 5,00 & 4,00 & 5,00 \\ 5,00 & 4,00 & 4,00 & 4,00 & 0,50 & 4,00 & 4,75 \\ 3,00 & 1,00 & 2,00 & 0 & 4,00 & 1,00 & 4,75 \\ 3,00 & 2,00 & 3,00 & 3,00 & 2,00 & 3,00 & 5,00 \end{pmatrix}.$$

В первых шести столбцах содержатся 30 ответов пользователя на вопросы теста Климова. В последнем столбце находятся средние баллы аттестата по пяти выделенным профессиональным группам из четырех дисциплин. Величина нечеткости, определяющая разброс оценки параметров тестирования в каждую сторону, взята равной 1. Данные по средним баллам дисциплин рассматриваются в процессе обработки как четкие величины. Матрица весов оценок имеет вид

$$r = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 2 & 1 & 1 \end{pmatrix}.$$

На основании матрицы A и матрицы r вычисляются векторы предпочтения i -й профессии $B_{ij} = \{A_{ij} \cdot r_{ij}\}$ с проекциями $j = 1, 2, \dots, 7$ по числу показателей сравнения. Далее образуются нечеткие треугольные векторы предпочтения $\tilde{R}_{ij} = \{B_{ij}^-, B_{ij}, B_{ij}^+\}$, где $B_{ij}^\pm = B_{ij} \pm \delta$, $\delta = 1$. Если значение границы треугольного числа выходит за диапазон $[0, 5]$, то она заменяется соответствующим граничным значением интервала 0 или 5. После этого вычисляются длины векторов $R_i = d(\tilde{B}_i, 0)$ по формуле (3) и проводится нормировка полученных длин на их сумму, что дает выходной вектор $R_i / \sum_k R_k$ рекомендаций в процентах: 15,4695; 26,4630; 24,3394; 14,2747; 19,4534. Итоги обработки результатов тестирования показывают, что предпочтительными для дальнейшего обучения рассматриваемого пользователя являются профессии, связанные техническими устройствами, или, с несколько меньшим рейтингом, работа с абстрактными символьными системами, моделями, естественными и искусственными языками.

Полученные рейтинговые оценки согласуются с результатами индивидуальных собеседований, проведенных экспертами с разными пользователями. Использование нечеткого представления вместо исходных оценок в баллах повышает согласованность конечных результатов с оценками экспертов, раздвигая средние по величине оценки и смягчая предельные оценки на концах интервала. Повторное тестирование показывает робастность приоритетного выбора по ранжированию результатов.

Для работы с нечеткой рекомендательной системой, основанной на совместной фильтрации, была сгенерирована база псевдоданных, имеющая сто записей. Численное моделирование показало, что такое количество случайных записей обеспечивает робастное распознавание приоритетного профессионального направления. Тестирование работы системы показало, что в случае ограничения числа записей несколькими десятками сгенерированные данные не позволяют воспроизвести правильную расстановку приоритетов.

ЗАКЛЮЧЕНИЕ

С учетом большого разнообразия направлений подготовки и читаемых в университете курсов вы-

бор студентом направления и собственной траектории обучения представляется серьезной задачей. Для эффективной помощи в этом вопросе нужно либо создавать штат дорогостоящих квалифицированных консультантов, либо искать решение в реализации специализированной компьютерной рекомендательной системы.

Авторами был разработан подход к выработке рекомендаций по выбору общего направления профессионального обучения учащегося на основе использования данных теста Климова. Исходные данные рассматриваются как нечеткие величины, которые обрабатываются на основе нечеткой метрики и нечетких сравнений. Для решения проблемы холодного старта применяется алгоритм экспертной системы оценки с помощью интегрального критерия длины нечеткого вектора оценок. Программа нечеткой экспертной оценки может использоваться самостоятельно и независимо от баз данных. Проблема холодного старта решена путем генерации случайной базы псевдоданных с использованием распределения ответов по приоритетам.

Тестирование релевантности базы псевдоданных проводилось путем сопоставления результатов нечеткой совместной фильтрации с ответами, которые дает нечеткая экспертная система. Результаты численного моделирования показали, что оптимальные размеры базы псевдоданных, обеспечивающей быструю работу одновременно с надежным получением ответа, составляют около ста записей.

При эксплуатации системы с участием реальных пользователей псевдоданные обновляются путем замены записей новыми данными на основе проведенных тестов. Актуализация базы данных позволяет поддерживать ее правильное функционирование и осуществлять ее адаптацию к изменениям, происходящим в контингенте пользователей.

Отметим возможность альтернативного подхода для определения образовательной траектории с использованием нечетких искусственных нейронных сетей путем решения задачи кластеризации данных об абитуриентах, а также обучения с учителем на этапе накопления больших данных. Преимущество выбранного подхода по отношению к методике, связанной с применением искусственных нейронных сетей, состоит в возможности его функционирования при объемах данных, недостаточных для машинного обучения в период холодного старта.

Планируется применение разработанной системы в Воронежском государственном техническом



университете (ВГТУ). Соответствующие рекомендуемые направления и профили подготовки для поступающих в ВГТУ разделены по группам:

– тип П: геодезия и дистанционное зондирование, землеустройство и кадастры, природообустройство и водопользование;

– тип Т: строительство, радиотехника, приборостроение;

– тип З: экономика, прикладная информатика, информатика и вычислительная техника;

– тип Х: архитектура, дизайн архитектурной среды, реконструкция и реставрация архитектурного наследия;

– тип Ч: журналистика, реклама и связи с общественностью.

Дальнейшее уточнение траектории обучения учащегося после поступления в университет происходит путем выбора элективных дисциплин, а также после окончания бакалавриата или специалитета при поступлении в магистратуру, а далее в аспирантуру. Эта задача является актуальной в связи с изменениями в организации системы высшего образования в России. Она предполагает проведение дальнейших исследований с целью разработки инструментов интеллектуальной поддержки формирования соответствующей индивидуальной дорожной карты студента.

ЛИТЕРАТУРА

1. *Фальк К.* Рекомендательные системы на практике. – М.: ДМК Пресс, 2020. – 448 с. [Falk, K. Practical Recommender Systems. – New York: Manning Publications, 2019. – 432 s.]
2. *Кутянин А.Р.* Рекомендательные системы: обзор основных постановок и результатов // Интеллектуальные системы. Теория и приложения. – 2017. – Т. 21, № 4. – С.18–30. [Kutyainin, A.R. Recommender Systems: Overview of Main Statements and Results // Intelligent Systems. Theory and Applications. – 2017. – Vol. 21, no. 4. – P. 18–30. (In Russian)]
3. *Koren, Y., Bell, R., Volinsky, C.* Matrix Factorization Techniques for Recommender Systems // Computer. – 2009. – Vol. 42, iss. 8. – P. 30–37.
4. *Noratiqah, M.A., Mohd, A.A., Nurul, F.R.* Comparison between Content-Based and Collaborative Filtering Recommendation System for Movie Suggestions // AIP Conference Proceedings. – 2013. – DOI: 10.1063/1.5054256.
5. *Ramlatchan, A., Yang, M., Liu, Q., et al.* A Survey of Matrix Completion Methods for Recommendation Systems // Big Data Mining and Analytics. – 2018. – Vol. 1, no. 4. – P. 308–323.
6. *Sheng, B., Gengxin, S.* Matrix Factorization Recommendation Algorithm Based on Multiple Social Relationships // Mathematical Problems in Engineering. – Vol. 2021. – Art. ID 6610645.
7. *Thakera, S., Nagori, V.* Analysis of Fuzzification Process in Fuzzy Expert System // Procedia Computer Science. – 2018. – Vol. 132. – P. 1308–1316.
8. *Lucas, J.P., Laurent, A., Moreno, M.N., Teisseire, M.* A Fuzzy Associative Classification Approach for Recommender Systems // International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. – 2012. – Vol. 20, no. 4. – P. 579–617.
9. *Zhang, Z., Lin, H., Liu, K., et al.* A Hybrid Fuzzy-Based Personalized Recommender System for Telecom Products/Services // Information Sciences. – 2013. – Vol. 235. – P. 117–129.
10. *Rahman Siddiquee, M.M., Haider, N., Rahman, R.M.* A Fuzzy Based Recommendation System with Collaborative Filtering // Proceedings of the 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA 2014). – Dhaka, 2014. – P. 1–8. – DOI:10.1109/SKIMA.2014.7083524
11. *Guangquan, L., Jiali, X., and Siyu, M.* Learning Resource Recommendation Method Based on Fuzzy Logic // Journal of Engineering Science and Technology Review. – 2018. – Vol. 11, no. 4. – P. 146–153.
12. *Sundus, A., Usman, Q., Raheel, N.* HCF-CRS: A Hybrid Content Based Fuzzy Conformal Recommender System for Providing Recommendations with Confidence // PLoS ONE. – 2018. – Vol. 13, no. 10. – Art. no. e0204849.
13. *Hamada, M., Ometere, A.L., Bridget, O.N., et al.* A Fuzzy-Based Approach and Adaptive Genetic Algorithm in Multi-Criteria Recommender Systems // Advances in Science, Technology and Engineering Systems Journal. – 2019. – Vol. 4, no. 4. – P. 449–457.
14. *Calderon-Vilca, H., Chavez, N.M., Guimarey, J.M.R.* Recommendation of Videogames with Fuzzy Logic // Proceedings of 27th Conference of Open Innovations Association (FRUCT). – Trento, 2020. – P. 27–37.
15. *Yang, Y., Zhang, Y.* Collaborative Filtering Recommendation Model Based on Fuzzy Clustering Algorithm // AIP Conference Proceedings. – 2018. – Vol. 1967, iss. 1. – Art. no. 040050. – DOI: <https://doi.org/10.1063/1.5039124>.
16. *Lee, S.* Fuzzy Clustering with Optimization for Collaborative Filtering-Based Recommender Systems // Journal of Ambient Intelligence and Humanized Computing. – 2022. – Vol. 13, no. 2. – P. 4189–4206.
17. *Comendador, B.E.V., Becbec, W.F.C., Guzman, J.R.P.* Implementation of Fuzzy Logic Technique in a Decision Support Tool: Basis for Choosing Appropriate Career Path // International Journal of Machine Learning and Computing. – 2020. – Vol. 10, no. 2. – P. 339–345.
18. *Duan, L., Wang, W., Han, B.* A Hybrid Recommendation System based on Fuzzy C-Means Clustering and Supervised Learning // KSII Transactions on Internet and Information Systems. – 2021. – Vol. 15, no. 7. – P. 2399–2413.
19. *Hasanzadeh, S., Fakhrahmad, S., Taheri, M.* A Fuzzy Approach to Review-Based Recommendation: Design and Optimization of a Fuzzy Classification Scheme Based on Implicit Features of Textual Reviews // Iranian Journal of Fuzzy Systems. – 2021. – Vol. 18, iss. 6. – P. 83–99.
20. *Cui, C., Li, J., Zang, Z.* Measuring Product Similarity with Hesitant Fuzzy Set for Recommendation // Mathematics. – 2021. – Vol. 9, iss. 21. – DOI 10.3390/math9212657.
21. *Chen, J., Lu, Y., Shang, F., Wang, Y.* A Fuzzy Matrix Factor Recommendation Method with Forgetting Function and User Features // Applied Soft Computing. – 2021. – Vol. 100. – Art. no. 106910.
22. *Yin, M., Liu, Y., Zhou, X., Sun, G.* A Fuzzy Clustering Based Collaborative Filtering Algorithm for Time-aware POI Recommendation // Journal of Physics: Conference Series. – 2021. – Vol. 1746. – Art. no. 012037. – DOI 10.1088/1742-6596/1746/1/012037.
23. *Jin, B., Liu, D., Li, L.* Research on Social Recommendation Algorithm Based on Fuzzy Subjective Trust // Connection Science. – 2022. – Vol. 34, no.1. – P. 1540–1555.

24. *Malandri, L., Porcel, C., Xing, F., et al.* Soft Computing for Recommender Systems and Sentiment Analysis // *Applied Soft Computing*. – 2022. – Vol. 118, no. 3. – Art. no. 108246. – DOI: <https://doi.org/10.1016/j.asoc.2021.108246>.
25. *Xin, Y., Henan, B., Jianmin, N. et al.* Coating Matching Recommendation Based on Improved Fuzzy Comprehensive Evaluation and Collaborative Filtering Algorithm // *Scientific Reports*. – 2021. – Vol. 11. – Art. no. 14035.
26. *Dai, B., Chen, R.-C., Zhu, S.-Z., Huang, C.-Y.* A Fuzzy Recommendation System for Daily Water Intake // *Advances in Mechanical Engineering*. – 2016. – Vol. 8, no. 5. – DOI: <https://doi.org/10.1016/j.eswa.2020.113738>.
27. *Liu, Y., Eckert, C.M., Earl, C.* A Review of Fuzzy AHP Methods for Decision-Making with Subjective Judgements // *Expert Systems with Applications*. – 2020. – Vol. 161. – Art. no. 113738.
28. *Akbar, M.A., Khan, A.A., Huang, Z.* Multicriteria Decision Making Taxonomy of Code Recommendation System Challenges: A Fuzzy-AHP Analysis // *Information Technology and Management*. – 2023. – Vol. 24. – P. 115–131.
29. *Теория управления (дополнительные главы): Учебное пособие / Под ред. Д.А. Новикова. – М.: ЛЕНАНД, 2019. – 552 с. [Teoriya upravleniya (dopolnitel'nye glavy). Uchebnoye posobiye. Ed. by D.A. Novikov. – M.: LENAND, 2019. – 552 s. (In Russian)]*
30. *Беляева К.А.* Сервис «Индивидуальная траектория студента» в личном кабинете студента iStudent.urfu.ru. Руководство пользователя. – Екатеринбург: УРФУ, 2016. [*Belyaeva, K.A.* Service «Individual Trajectory of the Student» in the Personal Account of the Student iStudent.urfu.ru. User Guide. – Yekaterinburg: URFU, 2016. (In Russian)]
31. *Арапова Е.А., Крамаров С.О., Сахарова Л.В.* Разработка концепции интеллектуальной платформы для реализации индивидуальной траектории обучения с учетом базового уровня знаний и психотипа обучающегося // *Вестник кибернетики*. – 2022. – № 1 (45). – С. 6–15. [*Arapova, E.A., Kramarov, S.O., Sakharova, L.V.* Concept Development of an Intelligent Platform Aimed at Implementing an Individual Learning Path According to the Student's Basic Level of Knowledge and Psychological Type // *Proceedings in Cybernetics*. – 2022. – No. 1 (45). – P. 6–15. (In Russian)]
32. *Куприянов Р.Б., Агранат Д.Л., Сулейманов Р.С.* Использование технологий искусственного интеллекта для выстраивания индивидуальных образовательных траекторий обучающихся // *Вестник Российского университета дружбы народов. Серия: Информатизация образования*. – 2021. – Т. 18, № 1. – С. 27–35. [*Kupriyanov, R.B., Agranat, D.L., Suleimanov, R.S.* The Use of Artificial Intelligence Technologies for Building Individual Educational Trajectories of Students // *Bulletin of the Peoples' Friendship University of Russia. Series: Informatization of education*. – 2021. – Vol. 18, no. 1. – P. 27–35. (In Russian)]
33. *Guseva, A.I., Kireev, V.S., Bochkarev, P.V., et al.* Scientific and Educational Recommender Systems. Information Technologies in Education of the XXI Century (ITE-XXI) // *AIP Conf. Proc.* – 2017. – Vol. 1797, iss. 1. – P. 2–11.
34. *Estrada, J.E., Bernabe, G.G., Lopez, J.S., Potestades, J.A.* Model Development in Assessing the Career Path of Senior High School Students in Philippine Setting // *International Journal of Information and Education Technology*. – 2018. – Vol. 8. – P. 459–461.
35. *Khorasani, E.S., Zhenge, Z., Champaign, J.* A Markov Chain Collaborative Filtering Model for Course Enrollment Recommendations // *Proceedings of 2016 IEEE International Conference on Big Data*. – Washington, DC, 2016. – P. 3484–3490.
36. *Kamal, A., Naushad, B., Rafiq, H., Tahzeeb, S.* Smart Career Guidance System // *Proceedings of 4th International Conference on Computing & Information Sciences (ICIS)*. – Karachi, 2021. – P. 58–61.
37. *Ilahi, R., Widiaty, I., Abdullah, A.G.* Fuzzy System Application in Education // *IOP Conf. Series: Materials Science and Engineering*. – 2018. – Vol. 434. – Art. no. 012308. – DOI: [10.1088/1757-899X/434/1/012308](https://doi.org/10.1088/1757-899X/434/1/012308).
38. *Selva Rani, B. and Ananda Kumar, S.* Recommendation System for under Graduate Students Using FSES-TOPSIS // *The International Journal of Electrical Engineering & Education*. – 2019. – DOI: <https://doi.org/10.1177/0020720919879385>.
39. *Benilda, E.V., Comendador, W.F., Becbec, C., John, R.P. de Guzman.* Implementation of Fuzzy Logic Technique in a Decision Support Tool: Basis for Choosing Appropriate Career Path // *International Journal of Machine Learning and Computing*. – 2020. – Vol. 10, no. 2. – P. 339–345. – DOI: [10.18178/ijmlc.2020.10.2.940](https://doi.org/10.18178/ijmlc.2020.10.2.940).
40. *Natividad, M.C.B., Gerardo, B.D., Medina, R.P.* A Fuzzy-Based Career Recommender System for Senior High School Students in K to 12 Education // *IOP Conference Series: Materials Science and Engineering*. – 2019. – Vol. 482. – Art. no. 012025. – DOI: [10.1088/1757-899x/482/1/012025](https://doi.org/10.1088/1757-899x/482/1/012025).
41. *Qamhie, M., Sammaneh, H., Demaidi, M.N.* PCRS: Personalized Career-Path Recommender System for Engineering Students // *IEEE Access*. – 2020. – Vol. 8. – P. 214039–214049. – DOI: [10.1109/ACCESS.2020.3040338](https://doi.org/10.1109/ACCESS.2020.3040338).
42. *Hernandez, R., Atienza, R.* Career Track Prediction Using Deep Learning Model Based on Discrete Series of Quantitative Classification // *Applied Computer Science*. – 2021. – Vol. 17, no. 4. – P. 55–74. – DOI: [10.23743/acs-2021-29](https://doi.org/10.23743/acs-2021-29).
43. *Nghiem, T.L., Dinh, T.H., Nguyen, T.L.* A Fuzzy Logic Approach to Career Orientation for Students: A Case Study in Human Resource Management // In: *Global Changes and Sustainable Development in Asian Emerging Market Economies*. Ed. By Nguyen, A.T., Hens, L. – Cham: Springer, 2022. – Vol. 1. – DOI: [10.1007/978-3-030-81435-9_17](https://doi.org/10.1007/978-3-030-81435-9_17).
44. *Климов Е.А.* Психология профессионального самоопределения. – М.: «Академия», 2004. – 304 с. [*Klimov, E.A.* Psychology of Professional Self-determination. – M.: «Academy», 2004. – 304 p. (In Russian)]
45. *Pereira, A.L.V., Hruschka, E.R.* Simultaneous Co-clustering and Learning to Address the Cold Start Problem in Recommender Systems // *Knowledge-Based Systems*. – 2015. – Vol. 82. – P. 11–19. – DOI: [10.1016/j.knosys.2015.02.016](https://doi.org/10.1016/j.knosys.2015.02.016)
46. *Шаталова А.О., Головинский П.А.* Нечеткая экспертная система оценки выбора профессии. Свидетельство о регистрации программы для ЭВМ № 2023662881 : заявл. 27.03.2023; зарег. 15.06.2023. [*Shatalova, A.O., Golovinsky, P.A.* Fuzzy Expert System for Assessing Career Choice. Certificate of registration of a computer program. № 2023662881: appl. 27.03.2023; reg. 15.06.2023. (In Russian)]
47. *Коротеев М.В., Терелянский П.В., Иванюк В.А.* Арифметика нечетких чисел в обобщенно трапециевидной форме // *Совр. матем. и ее приложения*. – 2015. – P. 94–99. [*Koroteev, M.V., Terelyansky, P.V., Ivanyuk, V.A.* Arithmetic of



- Fuzzy Numbers in Generalized Trapezoidal Form // J. Math. Sci. – 2016. – Vol. 216. – P. 696–701.]
48. *Seresht, N.G., Fayek, A.R.* Fuzzy Arithmetic Operations: Theory and Applications in Construction Engineering and Management: Theory and Applications // In: Fuzzy Hybrid Computing in Construction Engineering and Management. – Bingley: Emerald Publishing Limited, 2018. – P. 111–147. – DOI: 10.1108/978-1-78743-868-220181003.
49. *Mahanta, J., Panda, S.* Distance Measure for Pythagorean Fuzzy Sets with Varied Applications // Neural Computing and Application. – 2021. – No. 33. – P. 17161–17171.

Статья представлена к публикации руководителем РРС В.Ю. Столбовым.

Поступила в редакцию 13.02.2023,
после доработки 23.10.2023.
Принята к публикации 25.10.2023.

Головинский Павел Абрамович – д-р физ.-мат. наук,
✉ golovinski@bk.ru,
ORCID iD: <https://orcid.org/0000-0002-7527-0297>

Шаталова Ангелина Олеговна – старший преподаватель,
✉ angelina.streltsova.93@mail.ru,
ORCID iD: <https://orcid.org/0000-0001-8531-2078>

Воронежский государственный технический университет,
г. Воронеж.

© 2023 г. Головинский П.А., Шаталова А.О.



Эта статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0 Всемирная.

A FUZZY COLD-START RECOMMENDER SYSTEM FOR EDUCATIONAL TRAJECTORY CHOICE

P.A. Golovinskii and A.O. Shatalova

Voronezh State Technical University, Voronezh, Russia

✉ golovinski@bk.ru, ✉ angelina.streltsova.93@mail.ru

Abstract. Several approaches to choosing an educational trajectory are considered, and the advantages of using recommender systems are determined. The cold start problem of recommender systems is formulated and solved by creating a hybrid recommender system that combines a rule-based fuzzy expert system and a recommender system with fuzzy collaborative filtering. As one application, the general approach is implemented for choosing the field of study when entering a higher education institution. A modification of Klimov's career guidance test is used as initial data. The rules for estimating the metrics and similarity of fuzzy triangular data are presented. The algorithms of a fuzzy expert system and a fuzzy recommender system with collaborative filtering are described in terms of the fuzzy representation accepted. The two approaches are combined by generating pseudo data using an expert system. This provides a solution of the cold start problem and yields a recommender system whose quality is gradually improved by substituting the values from real user queries into the database. The programs implementing these algorithms are tested to confirm the effectiveness of the fuzzy recommender system.

Keywords: expert system, recommender system, fuzzy description, fuzzy metric, collaborative filtering, cold start, educational trajectory.