



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 60 за 2021 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

А.Е. Куренных, [В.А. Судаков](#)

Комбинированные
рекомендательные
алгоритмы для
научно-производственных
объединений

Рекомендуемая форма библиографической ссылки: Куренных А.Е., Судаков В.А. Комбинированные рекомендательные алгоритмы для научно-производственных объединений // Препринты ИПМ им. М.В.Келдыша. 2021. № 60. 12 с. <https://doi.org/10.20948/prepr-2021-60>
<https://library.keldysh.ru/preprint.asp?id=2021-60>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

А.Е. Куренных, В.А. Судаков

**Комбинированные рекомендательные
алгоритмы для научно-
производственных объединений**

Москва — 2021

Куренных А.Е., Судаков В.А.

Комбинированные рекомендательные алгоритмы для научно-производственных объединений

Рассмотрен подход к реализации комбинированных алгоритмов выработки рекомендаций на примере отдельного модуля информационной системы научно-производственного предприятия. Предложены четыре группы критериев, которые разносторонне с применением различных подходов позволяют исследовать и оценить ранжируемые объекты применительно к решаемой задаче выработки рекомендаций. Разработаны требуемая для реализации алгоритма модель базы данных, а также соответствующее программное обеспечение.

Ключевые слова: выработка рекомендаций, рекомендательные системы, комбинированные алгоритмы выработки рекомендаций.

Alexey Evgenevich Kurennykh, Vladimir Anatolevich Sudakov

Combined recommender algorithms for enterprises

The authors describe an approach to the implementation of combined algorithms for recommender systems considered with the example of a separate module of the information system of an enterprise. We propose four groups of criteria, which, using various approaches, allows one to explore and evaluate objects in relation to the problem of developing recommendations. During a research database schema required for the implementation of the algorithm has been developed, as well as the software.

Key words: development of recommendations, recommender systems, combined recommender algorithms.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект 20-31-90043.

Оглавление

Введение	3
Постановка задачи.....	5
Метод решения задачи.....	6
Программная реализация.....	9
Заключение.....	10
Библиографический список.....	11

Введение

В настоящее время широко прослеживается тренд повсеместной цифровизации различных сфер человеческой деятельности. Происходит внедрение информационных систем различных классов, автоматизация широкого спектра производственных процессов, применение современных вычислительных технологий затронуло все виды бизнес-процессов. Особо отчетливо виден рост степени информатизации на отечественных производственных предприятиях, где активно распространяются системы электронного документооборота, а также внедряются корпоративные информационные системы, системы управления жизненным циклом изделий, системы конструкторской документации, автоматизированные рабочие места и т.д.

Использование корпоративных информационных систем имеет ряд плюсов, которые выходят за рамки простой автоматизации процессов и облегчения труда сотрудников предприятия, пользующихся этими системами. В подавляющем большинстве случаев задачи корпоративной информационной системы сводятся к автоматизации процессов учета на предприятии: складской учет, бухгалтерский учет и прочие, то есть решаются задачи управления ресурсами предприятия, для чего есть отдельный класс систем – ERP системы. Однако потребности некоторых пользователей этих систем выходят за пределы учетных функций. Современные реалии, высокая конкуренция на внутреннем и внешнем рынках, пандемия требуют от предприятий точного оперативного и долгосрочного планирования своей деятельности, закупок, расходов и потребности в персонале. Текущее развитие информационных систем позволяет с высокой точностью строить производственные планы в разрезе номенклатуры производимых изделий, требуемых материальных ресурсов, профессий рабочих, периодов времени и других важных для предприятия аналитических разрезах. Еще одним плюсом использования корпоративных информационных систем является возможность автоматизированного сбора статистических данных, накопленных в учетной системе, для более точного планирования потребности в материально-производственных запасах.

Задачи планирования ресурсов имеют высокую важность и актуальны для многих научно-производственных объединений, а их решение требует использования математических подходов, задача которых заключается в обеспечении высокой точности прогнозирования потребности в тех или иных видах ресурсов. Таким образом, авторы данного исследования пришли к идее реализации модуля выработки рекомендаций как дополнения к основной информационной системе предприятия. Назначение этого модуля не ограничивается задачами планирования ресурсов, но и находит применение в различных повседневных задачах, например, в задаче выбора поставщика материальных ресурсов.

Стоит отметить, что рекомендательные системы, так же как и корпоративные информационные системы, имеют свою классификацию, основанную на принципе их действия и используемых алгоритмических подходах к выработке рекомендаций. Принцип действия рекомендательных систем основан на различных подходах, которые обладают своими преимуществами и недостатками:

- первый тип рекомендательных систем основан на коллаборативной фильтрации, такие системы обладают достаточно высокой точностью при наличии достаточного количества информации, хорошо применимы при поиске фильмов [1] и музыкальных композиций [2], а также попытки их использования есть в сфере здравоохранения [3];
- второй тип рекомендательных систем основан на контенте – данных, собранных о каждом объекте и других объектах, которыми уже был заинтересован субъект. Такой подход успешно применяется в текстовых рекомендациях [4] и новостных лентах [5]. Эти системы достаточно эффективны, однако сильно зависят от предметной области, в которой применяются;
- третий тип рекомендательных систем основан на знаниях о предметной области, такие рекомендации имеют высокую точность, позволяют учитывать большое количество разнородных факторов, взаимосвязи между объектами. Например, такие системы применяются в сфере финансовых фондов [6], корпораций [7]. Ядром такой системы является база знаний о предметной области, которая может быть эффективно представлена в виде графа знаний [8];
- четвертым типом рекомендательных систем являются гибридные. Они основаны на комбинациях вышеперечисленных подходов, что позволяет достигать качественных результатов, а также нивелировать недостатки каждого типа рекомендательных систем. Основной недостаток таких систем – высокая сложность разработки. Гибридные рекомендательные системы могут быть развитием систем, основанных на знаниях [9], но самым распространенным и используемым примером является комбинация коллаборативных и основанных на контенте систем [10]. В гибридных системах используются разнообразные подходы, например, нечеткие методы и искусственный интеллект [11, 12], широко применяются подходы кластерного анализа данных [13].

Основная проблема для рекомендательных систем – «проблема холодного старта», которую решают с помощью стохастических моделей [14] и альтернативных источников данных [15, 16].

Постановка задачи

Задача авторов данного исследования состояла в разработке алгоритмического и программного обеспечения для выработки рекомендаций с применением комбинированных методов. Основная идея разработанных алгоритмов заключается в инвариантности по отношению к предметной области, что дает возможность применения разработки не только в сфере материально-технического обеспечения, которая рассматривается в данной статье в качестве примера, но и во многих других.

В процессе выполнения алгоритма обрабатываются две сущности: сами сравниваемые объекты (альтернативы), а также набор их характеристик, которые могут быть собственными атрибутами альтернатив, статистическими данными, внешними характеристиками, а также выражать субъективное мнение пользователя, работающего с этими объектами (критерии). В качестве альтернатив в процессе решения одной задачи выработки рекомендаций может выступать конечное множество объектов одного класса, обозначим их множеством $\mathbf{O} = \{o_i\}, i = \overline{1, k}$, где o_i – конкретная альтернатива, рассматриваемая в данный момент, k – количество альтернатив данного класса в системе в момент решения задачи выработки рекомендаций. Для того чтобы разрабатываемый алгоритм мог претендовать на универсальность, а также инвариантность по отношению к предметной области заранее не оговаривается, какой именно класс объектов образует множество \mathbf{O} . Наполнение данного множества конкретными альтернативами должно происходить автоматически в момент решения конкретной задачи без необходимости редактировать исходный код программного обеспечения. Критерии, характеризующие альтернативы, также не подлежат жесткому декларированию в программном коде, а определяются пользователем прикладного программного обеспечения при решении конкретной задачи. Множество критериев альтернатив обозначим $\mathbf{C} = \{c_{ij}\}, i = \overline{1, k}, j = \overline{1, n}$, где c_{ij} – критерий с номером j для альтернативы с номером i , n – количество критериев для альтернатив данного класса. Основной задачей разрабатываемого алгоритма является агрегация векторных критериев отдельно взятых альтернатив с последующим ранжированием объектов в соответствии с критерием оптимизации в конкретной задаче выработки рекомендаций, что может быть формально записано следующим образом:

$$r(o_i) = f(\mathbf{c}_i). \quad (1)$$

В этом выражении введены следующие обозначения: $r(o_i) \in \mathbb{R}$ – ранг альтернативы с номером i , являющийся ее итоговой оценкой в задаче выработки рекомендаций; $f(\mathbf{c}_i)$ – агрегирующая функция, параметром которой является вектор значений критериев для данной альтернативы.

Критерий оптимизации запишем в следующем виде:

$$r(o_i) \xrightarrow{o_i} \max. \quad (2)$$

Не исключается и минимизация указанной величины, критерий оптимизации задается конечным пользователем и определяется исключительно решаемой задачей.

Отдельное место в постановке задачи к алгоритмическому и программному обеспечению занимает интерфейс пользователя, через который будет осуществляться работа с альтернативами и их критериями, выбор методов агрегации, а также ввод исходных данных. Этот интерфейс должен предусматривать возможность выбора класса объектов из общесистемного справочника каталогов, выбор критериев, которые являются атрибутами рассматриваемого объекта на уровне системы или ее базы данных, а также добавление внешних характеристик, которые не имеют прямой связи с рассматриваемыми объектами в рамках информационной системы предприятия. Также интерфейс должен представлять возможность настройки метода агрегации для каждой задачи выработки рекомендаций и позволять использовать вариативность алгоритма для каждого пользователя, т.к. у каждого из них могут быть свои субъективные взгляды на решаемую проблему.

Метод решения задачи

Для целей решения задачи выработки рекомендаций необходимо выполнить подготовку массивов исходных данных так, чтобы их было возможно обработать автоматическим алгоритмом, в том числе имеется необходимость их приведения к унифицированному и структурированному виду. С учетом того, что объем данных может быть значительным, рациональное решение – это хранить все данные об альтернативах, критериях и их значениях в таблицах реляционной базы данных. Для этого предлагается создать следующие таблицы и связи между ними:

- таблица классов объектов, в которой хранятся данные о ранжируемых объектах (связана с общесистемным каталогом объектов);
- таблица альтернатив, которая связывает сущности информационной системы с классами ранжируемых объектов (связана с классами объектов и сущностями общесистемных каталогов аналитик);
- таблица задач, хранящая информацию о задачах выработки рекомендаций (связана с классами объектов и пользователями системы);
- таблица критериев, которая хранит информацию о критериях, характеризующих конкретный класс сущностей в рамках данной задачи выработки рекомендаций (связана с классами объектов и задачами);
- таблица значений критериев, в которой хранятся рассчитанные значения критериев для каждой альтернативы в рамках определенной задачи (связана с критериями);

- таблица результатов расчетов для каждой альтернативы в рамках данной задачи (связана с альтернативами и задачами).

Вышеописанные таблицы позволяют решать задачи выработки рекомендаций без дублирования данных основных модулей системы. За счет хранения расчетных данных в таблицах обеспечивается быстрое действие. Этот фактор является значимым для любых информационных систем, а особую важность представляет для транзакционных систем, взаимодействующих с пользователем в режиме реального времени. В результате создания вышеописанных таблиц получается следующий фрагмент даталогической модели базы данных (рис. 1), который служит дополнением к стандартным таблицам информационной системы предприятия. Кроме создания указанных таблиц и связей между ними выполняется создание индексов таблиц, а также ряда хранимых процедур и функций, необходимых для функционирования системы.

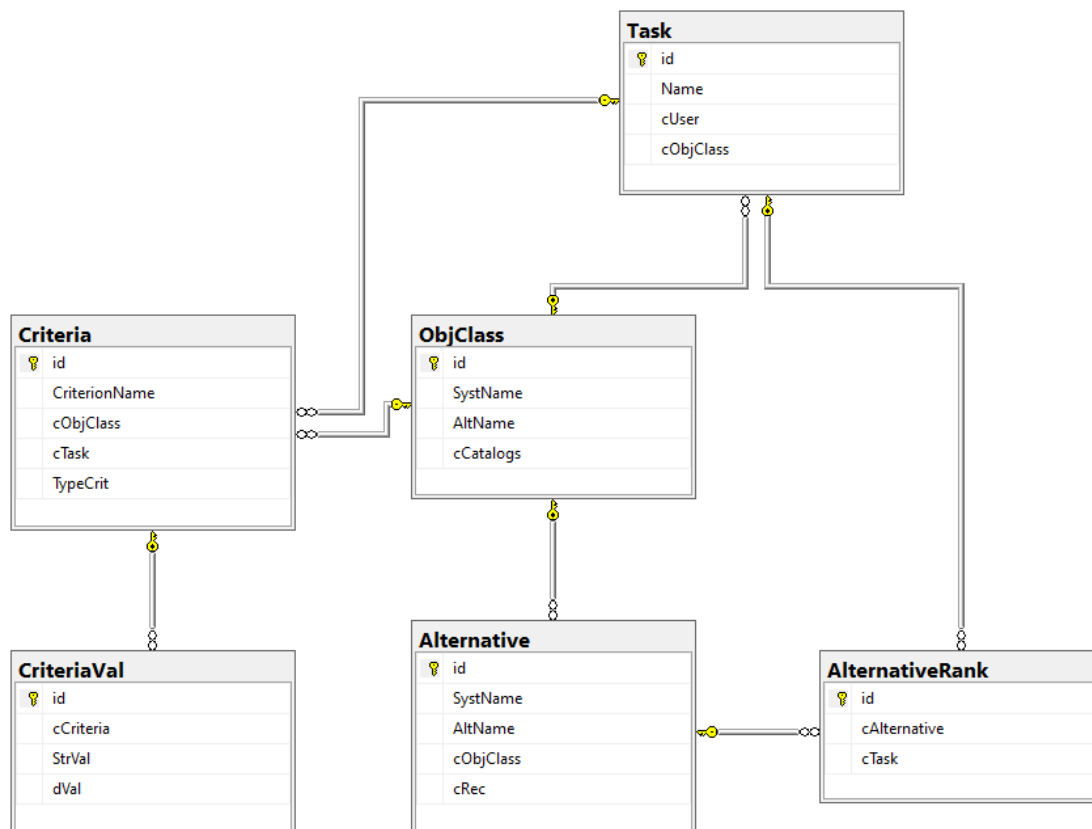


Рис. 1. Фрагмент даталогической модели базы данных

Далее рассмотрим реализованные подходы к вычислению и сбору значений критериев для альтернатив. Все способы вычисления критериев в рассматриваемых подходах к решению задач выработки рекомендаций делятся на четыре типа:

- хранимые непосредственно в базе данных, т.е. такие критерии, для вычисления значений которых не требуется выполнять существенных

математических вычислений. Их значения могут быть взяты из конкретных таблиц базы данных напрямую. Например, таким критерием может являться вхождение какого-либо поставщика в список “черных контрагентов” или, например, цена какой-либо материальной ценности или услуги в действующем прайс-листе предприятия. Обозначим множество этих критериев $SC = \{sc_{ij}\}$, где $SC \subset C$;

- вычисляемые на прикладном уровне. Для этих критериев значения могут быть вычислены путем сбора статистических данных из базы или путем различных математических вычислений. Например, средняя задержка поставки материала для конкретного поставщика или частота выявления бракованных изделий. Также можно рассмотреть в качестве примера расчетную стоимость закупки с учетом объемов потребности, а также скидок, предоставляемых поставщиком. Обозначим множество этих критериев $CC = \{cc_{ij}\}$, где $CC \subset C$ и $SC \cap CC = \emptyset$;
- внешние критерии. Значения этих критериев попадают в систему выработки рекомендаций исключительно из внешнего мира. Эти критерии не являются значимыми для основного функционала информационной системы предприятия, однако могут положительно сказываться на результате ранжирования альтернатив, повышая точность получаемых результатов. К таким критериям может относиться, например, численность штата сотрудников поставщика продукции или даже такой косвенный показатель, как стоимость его акций, если таковые имеются в открытом доступе. Обозначим множество этих критериев $OC = \{oc_{ij}\}$, где $OC \subset C$ и $SC \cap CC \cap OC = \emptyset$;
- пользовательские критерии. Значения этих критериев могут попадать в систему только посредством ручного ввода и следующих за этим преобразований. Критерии этой группы могут выражать субъективные взгляды пользователя по отношению к решаемой задаче выбора. Например, таким критерием может являться “опыт поставщика”. Обозначим множество этих критериев $UC = \{uc_{ij}\}$, где $UC \subset C$ и $SC \cap CC \cap OC \cap UC = \emptyset$, а также $SC \cup CC \cup OC \cup UC = C$.

Рассмотрение указанных групп критериев в задаче выработки рекомендаций реализует гибридный подход в рекомендательной системе. Действительно, первая, третья и четвертая группы критериев реализуют коллаборативный и контентный подход к выработке рекомендаций, а вторая группа критериев использует в себе знания о предметной области, в которой решается задача рекомендательного характера.

В соответствии с (1) значения критериев каждой из указанных групп подлежат агрегации, не исключено, что агрегация каждой отдельной группы выполняется различным образом, например, методом взвешенной суммы, сверткой Гермейера, мультипликативной сверткой или любыми другим

методом, которые могут настраиваться в системе. Таким образом, итоговый ранг альтернативы может быть определен:

$$r(o_i) = f(f_1(\mathbf{sc}_i), f_2(\mathbf{cc}_i), f_3(\mathbf{oc}_i), f_4(\mathbf{uc}_i)). \quad (3)$$

Здесь f, f_1, f_2, f_3, f_4 – различные агрегирующие функции, которые могут применяться при многокритериальной оценке альтернатив.

Программная реализация

Программная реализация алгоритма выработки рекомендаций и графические интерфейсы пользователя разработаны в среде разработки Viper 5.5 на языке Vip для информационной системы Галактика ERP. Выбранный способ реализации определяется тем, что решение рекомендательных задач происходит в высоконагруженной транзакционной системе, а язык программирования Vip предоставляет все необходимые средства для работы с современными системами управления базами данных, таблицами в памяти и представлениями.

Основные интерфейсы, с которыми работает конечный пользователь, включают в себя окна настройки альтернатив, критериев и способов их агрегации. На рис. 2 показан интерфейс, в котором пользователь создает и редактирует задачи выработки рекомендаций, а также задает критерий оптимизации и способы агрегации векторных критериев.

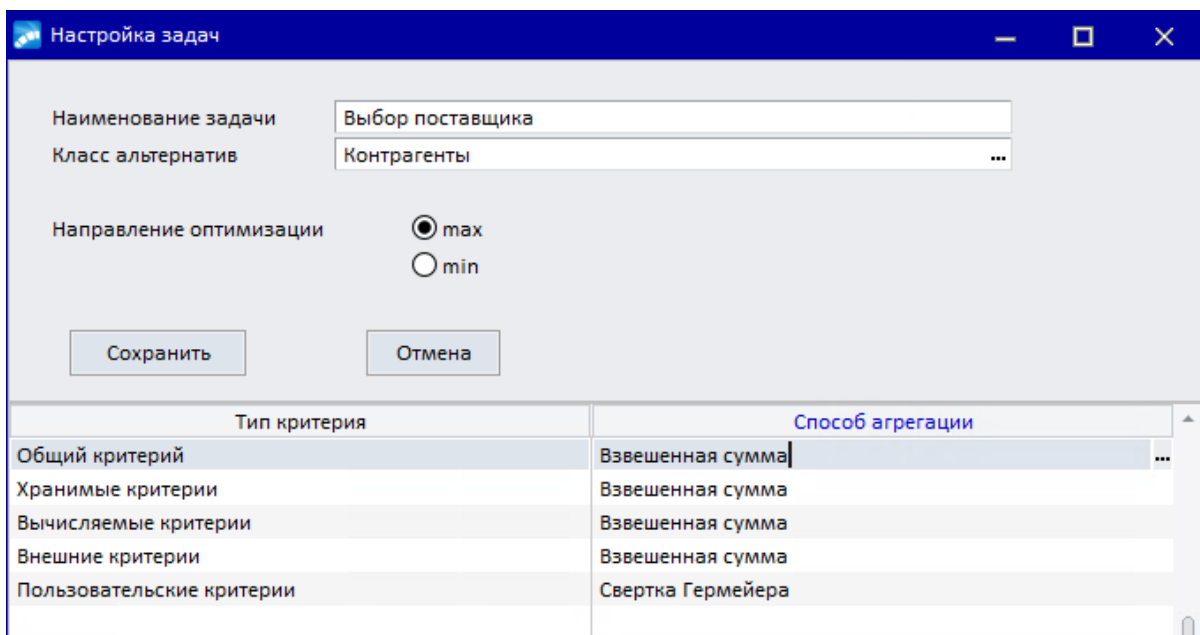


Рис. 2. Интерфейс для общей настройки задач

На рис. 3 показан интерфейс для настройки вычисляемого критерия, которым в качестве примера является количество рекламаций для поставщика материально-производственных запасов.

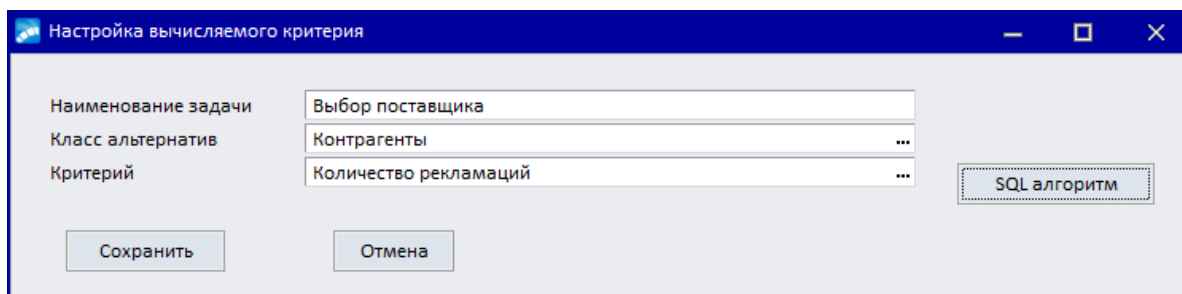


Рис. 3. Интерфейс для настройки вычисляемого критерия

Администратор системы работает только с интерфейсом, в котором описываются способы сбора вычисляемых и внешних критериев. Реализация их обработки требует минимального владения языком SQL и знания базы данных системы. Пример такого интерфейса представлен на рис. 4 и показывает подсчет рекламационных накладных для конкретного поставщика, уникальный идентификатор которого передается в SQL-запрос посредством переменной `FiltAlt`, а само окно настройки открывается по нажатию на кнопку «SQL алгоритм», показанную на рис. 3.

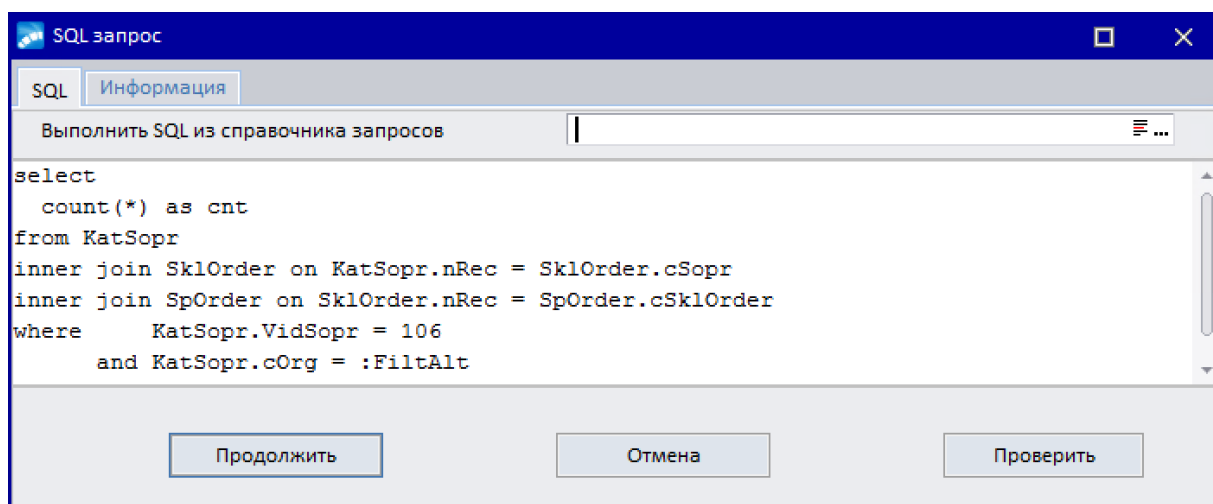


Рис. 4. Интерфейс для настройки вычисляемого критерия

Заключение

В данном репринте была рассмотрена актуальная научно-техническая задача разработки комбинированных алгоритмов по выработке рекомендаций для научно-производственных предприятий. С учетом применения различных подходов к оценке альтернатив, относительно которых происходит выработка рекомендаций, а также возможности изменять список рассматриваемых

критериев и способов их обработки в зависимости от конкретной задачи на прикладном уровне разработанные алгоритмы можно считать универсальными и инвариантными по отношению к предметной области.

Приоритетным направлением для продолжения данного исследования является выделение пятой группы критериев – критериев оценки результатов моделирования, которые будут являться результатом прогона мультиагентных моделей, позволяющих получить больше сведений о предметной области. Такое расширение подхода к выработке рекомендаций не потребует существенной доработки алгоритмов с точки зрения ранжирования альтернатив, однако необходима разработка механизмов интеграции с системами, позволяющими проводить имитационные эксперименты.

Потенциально применение моделей и автоматизированная обработка их результатов существенно расширяет границы применимости описанных алгоритмов, например, на сферу планирования производства, а также на область здравоохранения.

Библиографический список

1. Yutian Hu, Fei Xiong, Dongyuan Lu, Ximeng Wang, Xi Xiong, Hongshu Chen, Movie collaborative filtering with multiplex implicit feedbacks, *Neurocomputing*, Volume 398, 2020, Pages 485-494, <https://doi.org/10.1016/j.neucom.2019.03.098>.
2. Diego Sánchez-Moreno, Ana B. Gil González, M. Dolores Muñoz Vicente, Vivian F. López Batista, María N. Moreno García, A collaborative filtering method for music recommendation using playing coefficients for artists and users, *Expert Systems with Applications*, Volume 66, 2016, Pages 234-244, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2016.09.019>.
3. Anam Mustaqeem, Syed Muhammad Anwar, Muhammad Majid, A modular cluster based collaborative recommender system for cardiac patients, *Artificial Intelligence in Medicine*, Volume 102, 2020, 101761, ISSN 0933-3657, <https://doi.org/10.1016/j.artmed.2019.101761>.
4. Donghui Wang, Yanchun Liang, Dong Xu, Xiaoyue Feng, Renchu Guan, A content-based recommender system for computer science publications, *Knowledge-Based Systems*, Volume 157, 2018, Pages 1-9, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2018.05.001>.
5. Jia Li, Hua Xu, Suggest what to tag: Recommending more precise hashtags based on users' dynamic interests and streaming tweet content, *Knowledge-Based Systems*, Volume 106, 2016, Pages 196-205, ISSN 0950-7051, <https://doi.org/10.1016/j.knosys.2016.05.047>.
6. Álvaro Tejada-Lorente, Juan Bernabé-Moreno, Julio Herce-Zelaya, Carlos Porcel, Enrique Herrera-Viedma, A risk-aware fuzzy linguistic knowledge-based recommender system for hedge funds, *Procedia Computer Science*, Volume 162, 2019, Pages 916-923, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2019.12.068>.

7. Lu Zhen, George Q. Huang, Zuhua Jiang, An inner-enterprise knowledge recommender system, *Expert Systems with Applications*, Volume 37, Issue 2, 2010, Pages 1703-1712, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2009.06.057>.
8. Vincent Lully, Philippe Laublet, Milan Stankovic, Filip Radulovic, Enhancing explanations in recommender systems with knowledge graphs, *Procedia Computer Science*, Volume 137, 2018, Pages 211-222, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2018.09.020>.
9. John K. Tarus, Zhendong Niu, Abdallah Yousif, A hybrid knowledge-based recommender system for e-learning based on ontology and equential pattern mining, *Future Generation Computer Systems*, Volume 72, 2017, Pages 37-48, ISSN 0167-739X, <https://doi.org/10.1016/j.future.2017.02.049>.
10. Bogdan Walek, Vladimir Fojtik, A hybrid recommender system for recommending relevant movies using an expert system, *Expert Systems with Applications*, Volume 158, 2020, 113452, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.113452>.
11. Naime Ranjbar Kermany, Sasan H. Alizadeh, A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques, *Electronic Commerce Research and Applications*, Volume 21, 2017, Pages 50-64, ISSN 1567-4223, <https://doi.org/10.1016/j.elerap.2016.12.005>.
12. Tulasi K. Paradarami, Nathaniel D. Bastian, Jennifer L. Wightman, A hybrid recommender system using artificial neural networks, *Expert Systems with Applications*, Volume 83, 2017, Pages 300-313, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2017.04.046>.
13. Touraj Mohammadpour, Amir Massoud Bidgoli, Rasul Enayatifar, Hamid Haj Seyyed Javadi, Efficient clustering in collaborative filtering recommender system: Hybrid method based on genetic algorithm and gravitational emulation local search algorithm, *Genomics*, Volume 111, Issue 6, 2019, Pages 1902-1912, ISSN 0888-7543, <https://doi.org/10.1016/j.ygeno.2019.01.001>.
14. Antonio Hernando, Jesús Bobadilla, Fernando Ortega, Abraham Gutiérrez, A probabilistic model for recommending to new cold-start non-registered users, *Information Sciences*, Volume 376, 2017, Pages 216-232, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2016.10.009>.
15. Senthilselvan Natarajan, Subramaniaswamy Vairavasundaram, Sivaramakrishnan Natarajan, Amir H. Gandomi, Resolving data sparsity and cold start problem in collaborative filtering recommender system using Linked Open Data, *Expert Systems with Applications*, Volume 149, 2020, 113248, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2020.113248>.
16. J. Herce-Zelaya, C. Porcel, J. Bernabé-Moreno, A. Tejeda-Lorente, E. Herrera-Viedma, New technique to alleviate the cold start problem in recommender systems using information from social media and random decision forests, *Information Sciences*, Volume 536, 2020, Pages 156-170, ISSN 0020-0255, <https://doi.org/10.1016/j.ins.2020.05.071>.