

# АЛГОРИТМ СИСТЕМЫ РЕКОМЕНДАЦИЙ НА ОСНОВЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Мамедли Рамиль Эльман оглы

К.ф.-м.н.,

Нижевартовский государственный университет

prog-nv@mail.ru

## THE ALGORITHM OF THE RECOMMENDATION SYSTEM BASED ON GENETIC ALGORITHMS

R. Mammadli

*Summary.* In the Internet age, almost all kinds of services and products are available online for selection and use. In addition, there are several different providers for one type of product or service. In this context, to compare and find the right service according to the end customer, a service recommendation system is required. The purpose of the recommendation system is to understand the current requirements of the client and to study the database to restore the most likely services. To demonstrate the problems of this area and their solutions, a real problem is used, namely food and grocery delivery services. According to the design task, this recommender system is treated as a search system for a structured data source. Thus, the quantum genetic method is used to find suitable results from the proposed working model. This method first takes information from the data set and the user's requirement and then encodes the information in binary values. Next, the request sequence is processed as a binary string with all ones. Finally, a genetic algorithm is implemented to find a suitable solution among all available binary sequences. The generated results from the genetic algorithm are considered as the final recommendation of the search engine. In addition, fitness values are used to rank solutions. Implementation and evaluation of results are performed in C#. After that, performance is measured using time and space complexity. The obtained performance parameters demonstrate the acceptability of the work.

*Keywords:* data mining, recommender systems, genetic algorithm, search, databases.

*Аннотация.* В век Интернета почти все виды услуг и продуктов доступны онлайн для выбора и использования. В дополнение к этому для одного вида продукта или услуги существует ряд различных поставщиков. В этом контексте для сравнения и поиска подходящей услуги в соответствии с конечным клиентом требуется система рекомендаций по услугам. Целью системы рекомендаций является понимание текущих требований клиента и изучение базы данных для восстановления наиболее вероятных услуг. Чтобы продемонстрировать проблемы этой области и их решения, используется реальная проблема, а именно службы доставки еды и продуктов. По задаче проектирования эта рекомендательная система трактуется как поисковая система по структурированному источнику данных. Таким образом, для нахождения подходящих результатов из предложенной рабочей модели используется квантово-генетический метод. Этот метод сначала принимает информацию из набора данных и требования пользователя, после чего выполняется кодирование информации в двоичных значениях. Далее последовательность запроса обрабатывается как двоичная строка со всеми единицами. Наконец, реализуется генетический алгоритм для поиска подходящего решения среди всех доступных бинарных последовательностей. Генерируемые результаты из генетического алгоритма рассматриваются как окончательная рекомендация поисковой системы. Кроме того, значения пригодности используются для ранжирования решений. Реализация и оценка результатов выполняются на языке C#. После этого измеряется производительность с использованием временной и пространственной сложности. Полученные параметры производительности демонстрируют приемлемость работы.

*Ключевые слова:* интеллектуальный анализ данных, рекомендательные системы, генетический алгоритм, поиск, базы данных.

## Введение

В настоящее время в Интернете доступно большое количество онлайн услуг. Их качество можно оценить только на основе отзывов пользователей. Но вопрос о том, какие услуги подходят для различных потребностей пользователей, является сложной проблемой. В представленной работе предлагается модель рекомендательной системы для исследования и разработки, позволяющая найти наиболее подходящее решение в соответствии с потребностями конечного пользователя. Системы рекомендаций — это, в основном, методы прогнозирования или поиска релевантной информации по требованию пользователя. Чтобы найти нужную и подходящую услугу через онлайн-сервисы, предлагается система рекомендаций, основанная на методе интеллектуального анализа данных.

В качестве ключевой выбрана реальная проблема, учитывая, что для одного вида услуги имеется несколько поставщиков услуг. Все поставщики считают, что их услуги -наиболее подходящие для пользователей, но потребности клиентов отличаются от предлагаемых услуг. Таким образом, требуется система, которая анализирует услуги поставщиков и требования конечного пользователя, предлагая наиболее подходящую услугу среди доступных. В этом контексте данная работа сосредоточена на поставщиках онлайн-услуг для исследования и разработки решения. Дополнительно, предполагается, что для анализа входных данных, поступающих от обеих сторон, подходящим методом является поисковое решение на основе эвристики. Таким образом, предлагаемая методика использует квантовый генетический алгоритм для предоставления решения для нужд конечного пользователя.

### Обзор системы

Методы гибких вычислений позволяют находить много подходящих и вероятных данных среди большого количества паттернов. Обширна область приложений, где требуются высокоточные результаты при поиске закономерностей и используются подходы к программным вычислениям, например, приложения-регуляторы температуры, выбор признаков, управление запасами и производством и другие. В представленной работе демонстрируется новое применение подхода гибких вычислений для создания рекомендательной системы. Системы рекомендаций — это, по сути, своего рода системы предложений, понимающих требования и находящих наилучшее соответствие среди доступных товаров или услуг. Предлагаемая система рекомендаций предназначена для поиска подходящих услуг, представленных различными поставщиками.

В этом контексте в качестве исходной проблемы рассматривается одна из популярных в наше время отраслей, обслуживающих своих клиентов в Интернете, а именно службы доставки еды и продуктов. Большинство пользователей ищут эту услугу через Интернет, и иногда их не удовлетворяют фактические характеристики. Поэтому представленная работа предназначена для разработки системы рекомендаций для онлайн-сервиса доставки. Система состоит из двух модулей: первое — пространство поиска пользователя; второе — пространство решений, которое доступно при вводе набора данных. Чтобы получить решение, наиболее соответствующее потребностям пользователя, в качестве алгоритма поиска используется квантовый генетический алгоритм для генерации решения в соответствии с потребностями пользователя.

Генетический алгоритм является одним из популярных и эффективных методов поиска в программных вычислениях, который должен обеспечить решение задачи.

### Методология

Предлагаемая системная архитектура для поиска рекомендаций по требованиям пользователя описана с использованием рис. 1. На рисунке показаны используемые шаги для обработки данных и получения требуемого решения.

**Загрузка данных:** интеллектуальный анализ данных — это метод анализа данных, поэтому в любом приложении набор данных является неотъемлемой частью интеллектуального анализа. В представленной работе используется набор данных провайдера. Этот набор данных содержит различные виды услуг, предлагаемые поставщиками. Кроме того, также включен список значений услуг на основе обратной связи с пользователем. Это первая фаза системы, когда пользователь выбира-

ет набор данных из каталога локального компьютера. Предлагаемая система считывает набор данных и хранит его в структуре данных для выполнения анализа.

**Предварительная обработка данных:** предварительная обработка данных применяется для улучшения качества набора. Для этого выполняется очистка и преобразование данных. В предложенной работе для улучшения качества данных нулевые значения из набора данных удалены.

**Ввод пользовательских требований:** это дополнительный этап, созданный для поиска пользовательских требований. Таким образом, все отдельные значения атрибутов рассматриваются как отдельные услуги, и для каждой услуги оценивается весь набор данных. При этом для каждой службы вычисляются доступные уникальные значения из набора данных и добавляются в раскрывающийся список. Делая выбор из этих доступных услуг и значений, пользователь выполняет поиск. Следовательно, проблема заключается в поиске выбранных пользователем значений в содержимом набора данных (см. рис. 1).

**Кодирование данных:** на этом этапе значения обеих частей системы сначала принимаются из пользовательского интерфейса ввода, а затем — из доступного набора данных. После этого весь набор данных сравнивается с входным запросом пользователя и выполняется кодирование каждой последовательности решений. В этом контексте, когда значение экземпляра набора данных совпадает с требуемыми пользователем значениями, результат отмечается как 1, в противном случае он сохраняется как 0. Процесс кодирования набора данных приведен с использованием листинга 1.

Входные данные: Датасет  $D$ , пользовательские данные  $U$   
Выходные данные: бинарная кодированная строка  $B$

Процесс:

1.  $[row, col] = readDataset(D)$
2.  $[col, values] = readUserInput(U)$
3. for  $i=1$  to  $row$  do
  - a.  $temp = null$
  - b. for  $j=1$  to  $col$  do
    - i. if  $(D_{ij} = U_j)$   
 $temp.add(1)$
    - ii. else  
 $temp.add(0)$
    - iii. end if
  - c. end for
  - d.  $B.add(temp)$
4. end for
5. return  $B$

Листинг 1. Кодирование данных

**Генетический алгоритм:** после генерации двоичных закодированных строк к двоичному набору данных применяется генетический алгоритм. Весь набор данных



Рис. 1. Предлагаемая архитектура системы

рассматривается как популяция для генетического алгоритма. После выполнения поиска система генерирует подходящие решения в соответствии с их значениями пригодности.

Генетический алгоритм — это эвристический алгоритм поиска, который находит оптимальное решение в огромном пространстве данных. Доступные ресурсы обрабатываются с использованием методов естественной эволюции, таких как наследование, мутации, отбор и кроссинговер, чтобы найти наиболее подходящий ответ среди ряда решений, что в основном представляет собой итеративный процесс поиска более подходящего решения. Этот метод поиска гарантирует нахождение наилучшего решения, но промежуточное решение также вырабатывается на каждом последовательном шаге. Поэтому перед использованием этого алгоритма необходимо изучить функционирование генетического алгоритма. Генетические алгоритмы используют три ключевые идеи для поиска ответа: размножение, естественный отбор и разнообразие генов.

Генетические алгоритмы обрабатывают пару существ, эти существа представляют собой систему символов, которые вносят свой вклад в пространство решений. Новое поколение формируется с помощью процесса селекции и генетически поощряемых операторов. Краткое описание всего процесса расследования дается как последовательность следующих шагов:

- **Генерация начальной популяции** — в первую очередь генетические алгоритмы запускаются с произвольно сгенерированными порядками, с разрешенными алфавитами для генов. Для сокращения вычислительной процедуры все сгенерированные порядки популяции имеют одинаковое количество символов в отдельном порядке.
- **Проверка завершения алгоритма** — для остановки генетического алгоритма принцип остановки обязателен для обнаружения наилучшего решения. Можно остановить процедуру генетической оптимизации с помощью фиксирования
  - значения фитнес-функции,
  - наибольшего количества итераций,
  - количества поколений.
- **Селекция** — это процесс выбора лучших символов среди всех объектов, в этой ситуации для определения новой популяции используются два оператора, а именно кроссовер и мутация. В этом состоянии осуществляется восхождение последовательностей и использование этих лучших существностей передается новому поколению. Эксклюзивность гарантирует, что значение функции оптимизации не может дать самые неподходящие результаты.
- **Кроссовер** — это, по сути, процесс рекомбинации: объекты выбираются путем отбора и рекомбинируются друг с другом. Цель состоит в том, чтобы

получить новые объекты популяции, которые получают наиболее вероятные гены своих родителей.

- **Мутация** — произвольная вариация некоторых генов гарантирует, что даже если ни одна из существ не содержит необходимых генов-решений, все равно существует вероятность их генерирования посредством процесса мутации путем рандомизации исследования.
- **Новое поколение** — объекты, выбранные в результате селекции, в сочетании с теми генами, которые управляются скрещиванием и мутацией для развития последующего поколения.

Согласно описанию в [3] традиционный генетический алгоритм может быть определен с использованием приведенного ниже генетического псевдокода (листинг 2).

Входные данные: Популяция  $P$ , размер популяции  $\alpha$ , коэффициент элитизма  $\beta$ , коэффициент мутации  $\gamma$ , количество итераций  $\delta$

Выходные данные: решение  $X$

Процесс:

```
// инициализация
1. сгенерировать  $\alpha$  возможных решений случайным образом
2. сохранить их в популяции  $P$ 
// повторить указанное количество раз
3. for  $i=1$  to  $\delta$  do
    // селекция на основе элитизма
    a. число элитизма  $ne = \alpha \cdot \beta$ 
    b. выбор лучших  $ne$  решений из  $P$  и сохранение их в  $P_1$ 
    // кроссовер
    c. число кроссовера  $nc = (\alpha - ne) / 2$ 
    d. for  $j=1$  to  $nc$  do
        i из  $P_1$  случайно выбрать 2 решений ( $X_A$  и  $X_B$ )
        ii сгенерировать  $X_C$  и  $X_D$  путем одноточечного кроссовера с  $X_A$  и  $X_B$ 
        iii сохранить  $X_C$  и  $X_D$  в  $P_2$ 
    e. end for
    // мутация
    f. for  $j=1$  to  $nc$  do
        i из  $P_2$  случайно выбрать решение  $X_j$ 
        ii мутировать каждый бит  $X_j$  с коэффициентом  $\gamma$  и сгенерировать новое решение  $X'_j$ 
        iii if  $X'_j$  не подходящий
            заменить  $X_j$  на подходящее решение
        iv d. end if
        v e. в  $P_2$  заменить  $X_j$  на  $X'_j$ 
    g. end for
    // обновление
    h. обновить  $P = P_1 + P_2$ 
4. end for
// возвращение наиболее подходящего решения
5. возврат  $X_{best}$  из  $P$ 
```

Листинг 2. Генетический алгоритм

**Рекомендация:** сгенерированные решения с использованием генетического алгоритма используются в качестве рекомендационной системы. В качестве основы ранжирования здесь используются значения приспособленности процесса генетического поиска.

**Производительность:** производительность предлагаемой поисковой системы также рассчитывается с точки зрения требований времени и использования памяти.

#### Предлагаемый алгоритм

В этом разделе представлены этапы работы системы, поэтому все процессы системы, описанные выше, суммированы в виде небольших шагов в листинге 3.

Входные данные: Датасет  $D$ , пользовательские данные  $R$

Выходные данные: подходящее решение  $S$

Процесс:

```
1.  $T = \text{readDataset}(D)$ 
2.  $P_N = \text{preprocessData}(T)$ 
3.  $U = \text{readUserInput}(R)$ 
4. for  $i=1$  to  $N$  do
    a.  $B = \text{encodeData}(P_i, U)$ 
5. end for
6.  $S = \text{geneticAlgorithm.Search}(U, B)$ 
7. return  $S$ 
```

Листинг 3. Алгоритм решение

#### Заключение и будущая работа

Целью предлагаемой работы является разработка рекомендательной системы на основе интернет-сервисов, предлагающих услуг в соответствии с их качеством обслуживания. Ключевое проектирование и реализация предложенной методики завершены, и в этой главе представлены итоги выполненной работы с возможностью дальнейшего расширения.

#### Заключение

Термин «система рекомендаций» относится к электронной коммерции, где эта система реализована для предложения хороших продуктов. Кроме того, в таких рекомендательных системах учитывается поведение пользователей. Представленная работа представляет новую модель рекомендательной системы для выбора веб-услуг. В настоящее время поставщики продвигают множество схожих продуктов и услуг, из-за чего онлайн-пользователи путаются при выборе наиболее подходящих для них. Поэтому для поиска оптимального выбора онлайн-сервисов предлагается новая рекомендательная модель.

Предлагаемая система рекомендаций работает на основе доступных или предлагаемых услуг постав-

щиками услуг и их соответствующих значений обратной связи. В соответствии с потребностями или требованиями конечных клиентов услуги изучаются, оптимальные услуги выбираются и предлагаются клиентам. Чтобы продемонстрировать ключевые вопросы разработки модели рекомендации услуг для предлагаемых онлайн-услуг, выбрана реальная проблема службы доставки еды и продуктов. Большую часть времени конечный клиент беспокоится о предложениях и реальности, в этом контексте требуется решение. Предлагаемая модель использует концепцию квантовой генетики для поиска подходящих рекомендаций в соответствии с требованиями клиента. Модель работает в два этапа: сначала она обрабатывает набор данных для уточнения и кодирования данных с использованием квантовой концепции. Здесь для кодирования также используются требования и запросы пользователя. После кодирования пространство решений преобразуется в двоичные строки. В следующем модуле закодированные данные используются в качестве пространства решений для поиска требований пользователя. Окончательный результат алгоритма поиска принимается как рекомендуемая услуга через предложенную систему. Реализация предложенной ме-

тодики осуществляется с использованием технологии C# и структур данных на основе C#.

#### Будущая работа

Основная цель работы по внедрению веб-модели рекомендации услуг для повышения приемлемости для пользователей успешно завершена. В ближайшем будущем возможны следующие расширения для работы.

1. В настоящее время система реализована только с генетическим алгоритмом, в ближайшем будущем будут исследоваться более мягкие вычислительные методы, и оптимальная техника будет реализована с помощью системы.
2. Текущая система не оценивается на реальных данных, она разработана и протестирована только для предопределенного набора данных и их атрибутов. В ближайшем будущем будут предприняты усилия по привлечению атрибутов реального мира.
3. Текущая система не включает обзоры и отзывы в социальных сетях; предлагается задействовать оба метода анализа настроений, чтобы еще больше улучшить текущий механизм рекомендаций.

#### ЛИТЕРАТУРА

1. Демидова Л.А., Коняева Е.И., Кортаев А.Н. Генетический алгоритм поиска оптимальной комбинации фаззификаторов для FCM-алгоритма на основе нечетких множеств второго типа. Информационные технологии моделирования и управления. 2008. № 6 (49). С. 657–664.
2. Мамедли Р.Э. Системы управления базами данных. — Нижневартовск: Изд-во Нижневартовского государственного университета, 2021.
3. Ульянов С.В. и др. Интеллектуальные системы управления: в 5 т. Т. 4. Оптимизатор баз знаний на квантовых вычислениях: в 2 ч. Ч. 2. Самоорганизующиеся интеллектуальные системы управления. Дубна: Изд-во Гос. ун-та Дубна, 2014. 182 с.
4. Huaixiao Wang, Jianyong Liu, Jun Zhi. The improvement of quantum genetic algorithm and its application on function optimization. Mathematical Problems in Engineering, 2013, vol. 2013, art. 730749, 10 p. DOI: 10.1155/2013/730749
5. Lahoz-Beltra R. Quantum genetic algorithms for computer scientists. Computers. 2016, vol. 5, iss. 4, 24 p. DOI: 10.3390/computers5040024
6. Malossini A., Blanzieri E., Calarco T. QGA: A quantum genetic algorithm. Tech. Report, Univ. of Toronto, 2004, pp. 1–19. URL: [http://eprints.biblio.unitn.it/711/1/qga\\_techrep.pdf/](http://eprints.biblio.unitn.it/711/1/qga_techrep.pdf/) (дата обращения: 01.12.2022)
7. Pintu Chandra Shill, Faijul Amin, Kazuyuki Murase. Parameter optimization based on quantum genetic algorithms for fuzzy logic controller. Proc. 27th Fuzzy System Sympos. Japan, 2011, pp. 1065–1068.

© Мамедли Рамиль Эльман оглы (prog-nv@mail.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»