

# **АЛГОРИТМ АНАЛИЗА АКТИВНОСТИ ПОЛЬЗОВАТЕЛЕЙ ПОИСКОВЫХ СТРАНИЦ ИНТЕРНЕТ-СЕРВИСОВ**

Жиляк Н.А., канд.тех.наук, БГТУ, доцент

*Ключевые слова:* интернет, информация, сервис, поиск, алгоритм.

Существует достаточно большое количество сервисов, которые предоставляют схожий набор функциональных возможностей, связанных с публикацией и распространением виртуальных и не цифровых товаров посредством сети Интернет, тем не менее не смотря на схожесть концепций и идей каждая из таких систем имеет ряд особенностей, в том числе связанных с информацией, которую можно отнести к открытой. Так, например, проводя анализ сервиса Wildberries можно выделить пять основных критериев, а рассматривая Amazon только 4 [1-3].

Не смотря на отсутствие необходимости проведения сложных вычислений, задача вычисления ключевых параметров для определения местоположения результата в поисковой выдаче формирует ряд математических проблем, сложность которых обосновывается непредсказуемым количеством входных параметров, зависящих друг от друга, и от значений одного параметра на временной шкале. Для решения данных математических задач, в рамках разработанного алгоритма, использовался подход весовых коэффициентов для определения уровня значимости того или иного критерия, а также была использована авторегрессионная модель для выявления зависимостей между текущими значениями в рамках одного продукта на временном ряду.

Ввиду специфичности задачи анализа информации на основании открытых данных, а также ввиду наличие неопределенного количества входных параметров, не ограниченных системой их распространения, выявления точных показателей без использования алгоритма платформы, где распространяется продукт, невозможно. Тем не менее, на основании разработанного алгоритмического подхода, можно выявить ряд ключевых критериев, на основании которых можно рассчитать достаточно точное положение заданного поискового результата, а также спрогнозировать его дальнейшее развитие в условиях внутренней конкуренции с другими продуктами.

Анализируя и изучая наиболее известные системы распространения и продажи товаров в Интернет, пренебрегая функциональными особенностями, можно заметить ряд схожих практик и подходов с точки зрения ранжирования результатов, что с одной стороны говорит о том, что часть критериев находится вне всех систем, а с другой позволяет выявить общие. Среди таких критериев можно отметить те, что указывают на бизнес показатели, например количество продаж, а также общие, такие как количество просмотров, комментариев и других видов активности.

Нечеткая логика представляет собой мощный инструмент для анализа данных, особенно в тех случаях, когда данные содержат неопределенность и нечеткость. К основным преимуществам можно отнести:

1. **Обработка неопределенности и неточности.** Нечеткая логика позволяет работать с данными, которые не являются четкими или точными, что часто встречается в реальных приложениях. Например, оценки пользователей или отзывы могут содержать субъективные и размытые мнения.

2. **Интерпретируемость.** Правила нечеткой логики легко интерпретируются человеком. Например, правила типа "Если рейтинг высокий и количество отзывов высокое, тогда товар популярен" понятны и объяснимы.

3. **Адаптивность.** Нечеткая логика позволяет адаптировать систему под конкретные нужды, изменяя функции принадлежности и правила вывода. Это делает ее очень гибкой для различных приложений.

4. **Комплексные системы принятия решений.** Нечеткая логика может быть использована для построения сложных систем принятия решений, где учитываются множество факторов и их взаимодействие.

В сравнении с другими алгоритмами для анализа данных, нечеткая логика позволяет лучше обрабатывать субъективные данные, что делает ее более оптимальным выбором для задач анализа пользовательской активности, где такие данные являются встречаются чаще. Например, алгоритмы классификации требуют четко определенных классов и могут плохо справляться с размытыми границами между классами. В случае анализа пользовательской активности, где границы между "популярным" и "непопулярным" могут быть размыты, нечеткая логика предоставляет более естественный способ обработки таких данных. Аналогично алгоритмы регрессии предполагают наличие точных зависимостей между переменными, а алгоритмы кластеризация, которые используется для группировки данных на основе сходства, не всегда предоставляет интерпретируемые результаты и могут плохо справляться с нечеткими границами между кластерами. Нечеткая логика позволяет более гибко и понятно определить, насколько данные принадлежат к различным группам [4-6].

Алгоритм анализа пользовательской активности на основе нечеткой логики представляет собой последовательность шагов, которые позволяют выявить закономерности и сделать предсказания на основе неопределенной и размытой информации.

В общем виде алгоритм анализа данных на основе нечеткой логики выглядит следующим образом:

1. **Сбор данных.** На данном этапе осуществляется сбор всех необходимых данных о товарах и пользовательской активности. Данные могут быть получены из различных источников, таких как страницы

товаров на интернет-платформах (например, Amazon, Wildberries, Ozon). Примеры данных могут включать: рейтинг товара, количество отзывов и так далее.

2. Предобработка данных. На этом этапе данные очищаются и подготавливаются для анализа. Это может включать удаление дубликатов, заполнение пропущенных значений, нормализация данных для приведения их к общему масштабу и так далее.

3. Определение функций принадлежности. На данном этапе осуществляется функций принадлежности для каждого параметра. Функции принадлежности определяют степень, с которой каждый параметр принадлежит к различным категориям (низкий, средний, высокий). Наиболее часто используются сигмоидные функции, которые имеют плавный переход между уровнями принадлежности

4. Вычисление степеней принадлежности. Для каждого собранного параметра вычисляются степени принадлежности к категориям низкий, средний и высокий, используя заранее определенные функции принадлежности.

5. Применение правил нечеткого вывода. На данном этапе используются правила нечеткой логики для комбинирования степеней принадлежности различных параметров и получения нечетких выводов. Пример правила нечеткого вывода: "Если рейтинг высокий и количество отзывов высокое и количество заказов высокое и изменение цены высокое и общая оценка высокая и количество вопросов высокое, тогда товар является чрезвычайно популярным."

6. Агрегация нечетких выводов. На этом этапе производится агрегация нечетких выводов для получения обобщенной оценки популярности товара. Это может включать комбинацию различных степеней принадлежности по заданным правилам.

7. Дефаззификация. Процесс дефаззификации заключается в преобразовании нечетких выводов в четкие значения, которые можно интерпретировать и использовать для принятия решений. Один из методов дефаззификации — метод центра тяжести, который позволяет вычислить окончательную оценку популярности товара.

Анализ результатов. Заключительный этап включает анализ полученных результатов, оценку точности и эффективности модели. На этом этапе можно сравнить результаты анализа с реальными данными о продажах и позициях товаров в поисковой выдаче, а также предложить возможные улучшения алгоритма.

В рамках проделанной научной работы был разработан уникальный алгоритм анализа пользовательской активности на поисковых страницах интернет-сервисов, на основании неограниченного количества входных параметров, а также сформировано его математическое описание и

структура, концептуальной идеей которой является подход нечетких вычислений.

Разработанный математический аппарат был успешно внедрён в разработанное программное средство на основе технологии *Node JS*, а также опубликован в сети Интернет на основе платформы *Heroku*, что позволило предоставить открытый программный интерфейс, осуществляющий расчеты результирующих показателей принимая входные параметры клиента. Именно такая структура приложения позволяет использовать его максимально оптимально и широко, ведь интеграция с разработанным программным средством сводится к отправке запросов по протоколу *HTTP*.

В процессе реализации и разработки программного средства был также учтен вопрос программного тестирования разработанного веб-сервиса, что с одной стороны упростило процесс отладки, а с другой позволило переиспользовать составные части теста для проведения оценки адекватности математической модели спроектированного алгоритма.

Среди прочих научных результатов стоит также отметить практическое подтверждение адекватности математической модели, проведённое с использованием разработанного программного средства и выявляющее практическую значимость полученных результатов. Полученные подтверждение, не только демонстрируют определенный уровень точности разработанной системы, но также указывают на возможность практического использования разработки с целью прогнозирования дальнейшего поведения поисковой выдачи на временной шкале.

#### Литература

1. Григорьева Д., Гареева Г. «Основы Нечеткой Логике», 2018. – 43 с.
2. Математические основы нечеткой логики [Электронный ресурс] / электронный документ – Режим доступа: [vc.ru/loginom/201212](http://vc.ru/loginom/201212) – Дата доступа: 18.10.2025.
3. Алгоритмы интеллектуального анализа [Электронный ресурс] / электронный документ – Режим доступа: [docs.microsoft.com/ru-ru](http://docs.microsoft.com/ru-ru) – Дата доступа: 18.10.2025.
4. Регрессионный анализ [Электронный ресурс] / электронный документ – Режим доступа: [machinelearning.ru/wiki](http://machinelearning.ru/wiki) – Дата доступа: 19.10.2025.
5. Шитиков В., Мاستицкий С. «Классификация, регрессия и другие алгоритмы Data Mining», 2017. – 351 с.
6. Хердле В. «Прикладной многомерный статистический анализ», 2019. – 550 с.