

можно сказать, что рассмотренный метод динамического подключения ресурсов останется эффективным и при значительном увеличении подключаемых вычислительных узлов.

ЛИТЕРАТУРА

1. Балансировка нагрузки: основные алгоритмы и методы // Хабрахабр. URL: <http://habrahabr.ru/company/selectel/blog/250201/> (дата обращения 26. 01. 2016)
2. Э. Таненбаум, М. ван Стеен. Распределенные системы. Принципы и парадигмы — СПб.: Питер, 2003. — 877 с:
3. Исследование стратегий балансировки нагрузки в системах распределенной обработки данных // КиберЛеника. URL: <http://cyberleninka.ru/article/n/issledovanie-strategiy-balansirovki-nagruzki-v-sistemah-raspredelennoy-obrabotki-dannyh> (дата обращения 03.02.2016)

ОБЗОР АЛГОРИТМОВ ДЛЯ СЕРВИСА РЕКОМЕНДАЦИИ КНИГ

К.С. Никитина

(г. Томск, Национальный исследовательский Томский политехнический университет)

e-mail: ksn1@tpu.ru

REVIEW ALGORITHMS FOR SERVICE RECOMMENDATIONS BOOKS

K.S. Nikitina

(Tomsk, National Research Tomsk Polytechnic University)

Abstract. This article describes recommendation systems methods and algorithms. It also it describes advantages and disadvantages of each method, analysis of similar recommendation services like service under development. The article describes the choice of recommendations algorithm is for the developing service.

Keywords: system recommendations, service, collaborative filtering, user-based collaborative filtering, item-based collaborative filtering.

Введение. Каждый человек оказывается в ситуации, когда ему необходимо что-то выбрать. Например, человек хочет посмотреть фильм на выходных, но не знает какой, а в его распоряжении есть только название, краткий сюжет и сопутствующий трейлер – это не простая задача.

Анализ методов рекомендации. Чтобы помочь пользователю выбирать художественные книги разрабатывается сервис рекомендации. Основной задачей любой рекомендательной системы является получение списка наиболее интересных объектов для определенного пользователя. Все системы рекомендаций можно разделить на три класса [1]:

- методы коллаборативной фильтрации;
- методы, анализирующие содержимое объектов;
- методы, основанные на знаниях.

Коллаборативные методы основаны на изучении поведения пользователя, то есть отслеживают оцененные объекты или действия. Процесс сбора информации о пользователе бывает явный и неявный. При явном процессе сбора пользователю предлагается оценить определенные объекты, а при неявном – программа сама отслеживает действия пользователя, например, количество просмотров одного и того же видеоролика. Рассмотрим алгоритм, который состоит из трех шагов: для каждого пользователя вычисляем, насколько его интересы совпадают с интересами конкретного пользователя, затем выбираем ближайших и, наконец, предсказываем оценку на основе оценок соседей с предыдущего шага. Описанный алгоритм

основан на сравнения между собой пользователей. Также есть другой подход: вместо пользователей сравнивать объекты, например, книги. Главный недостаток такой системы – это «холодный старт»: система не знает, кому рекомендовать новый товар и что порекомендовать новому пользователю. Но его можно решить за счет анкетирования пользователя при регистрации или неявно собирать информацию о пользователе [2].

Любой объект имеет название и другую, хотя бы небольшую, текстовую информацию, которая используется в методах второго класса. Например, в книжном интернет-магазине пользователь имеет свой профиль, где указывает любимые жанры и книги. В системе хранится информация о ключевых словах из каждой книги. С помощью схожести ключевых слов подбираются рекомендации определенному пользователю. Основным недостатком методов, анализирующих содержимое объектов, заключается в узконаправленности.

Предыдущие методы оптимально работают при достаточно частой посещаемости сервиса пользователем. Редкие действия определенного пользователя усложняют подбор объекта, которым можно его заинтересовать. Для этого применяются методы, основанные на знаниях, которые решают проблему малой информативности или полного её отсутствия. Методы делятся на два вида: использование жестких ограничений и выбор близких объектов. Алгоритм схож в обоих видах: пользователь перечисляет свои требования к объекту, а система пытается найти нужный объект. Все требования пользователя строго соблюдаются при использовании жестких ограничений. Наиболее близкие объекты предлагаются при использовании второго вида методов, так как в этом случае допускаются близкие характеристики к указанным параметрам, введенным пользователем.

Анализ аналогов сервиса. Рассмотрим несколько сервисов рекомендаций, которые предлагают своим пользователям не только книги, а ещё фильмы и музыку. «Имхонет» предлагает помощь пользователям в огромном разнообразии игр, книг и фильмов. Информация об интересах пользователя собирается на основе оцененных им объектов. На основе полученных оценок составляется список единомышленников, то есть список людей, поставивших схожие оценки тем же объектам. Затем составляются рекомендации конкретному пользователю исходя из оцененных объектов единомышленников [3]. В данном сервисе используется алгоритм коллаборативной фильтрации основанной на пользователях (*user-based collaborative filtering*). Разработчики сервиса заблаговременно решили проблему «холодного старта» для нового пользователя. Если в системе появляется новый пользователь, то ему необходимо оценить некоторое количество объектов, чтобы система смогла предложить персональные рекомендации. Если новый клиент не оценивает объекты, то ему предлагается список наиболее популярных книг, фильмов и игр.

Рекомендательный сервис Readly направлен на подбор подходящих книг для определенного пользователя. Персональные рекомендации доступны после регистрации пользователя. Система начинает предлагать интересные книги исходя из количества прочитанных книг и поставленных оценок. Следовательно, рекомендательная система основана на коллаборативной фильтрации. При этом оценка пользователя используется для поиска единомышленников, а прочитанная книга для анализа других объектов. За счет перечисления прочитанных книг система отбрасывает их, чтобы не предлагать знакомую книгу пользователю.

Интеллектуальный рекомендательный сервис «Узнай, что почитать» предлагает сориентировать читателя в широком выборе книг. Любой желающий может подобрать книгу в этом сервисе, так как регистрация не требуется. Для получения рекомендации необходимо указать параметры фильтра, например, цель чтения или жанр. Система использует методы, основанные на знаниях с выбором близких объектов. Например, при введении следующих параметров: размышление, переживать, фантастика – сервис предлагает книгу Дэниела Киза «Цветы для Элджернона», что является близким объектом, так как это современная проза.

Выбор оптимального алгоритма. Аналоги используют разные методы для предложения рекомендаций, но наиболее часто используется коллаборативная фильтрация. При

таким подходе система основывается чаще всего на пользователях. Алгоритмы рекомендаций позволяют посоветовать книги, используя ключевые слова или фразы в ней. Можно сделать вывод, что для разрабатываемого сервиса рекомендаций художественной литературы оптимальным вариантом является метод анализа содержимого объекта. Ключевые слова есть в каждой книге, а пользователи не всегда ставят оценки. При использовании коллаборативной фильтрации необходимо учитывать, что некоторые пользователи намеренно занижают или завышают оценку объектам.

ЛИТЕРАТУРА

1. Гомзин А.Г., Коршунов А.В. Системы рекомендаций: обзор современных подходов. – Труды Института системного программирования РАН, 2012 т.22. – С. 401-417.
2. Воронцов К.В. Методы коллаборативной фильтрации и тематического моделирования. URL: <http://www.machinelearning.ru/wiki/images/9/95/Voron-ML-CF.pdf> (дата обращения: 28.03.2016).
3. Имхонет – рекомендательный сервис. URL: <http://imhonet.ru/about/> (дата обращения: 28.03.2016).

ОПРЕДЕЛЕНИЕ АССОЦИАТИВНЫХ ПРАВИЛ МЕЖДУ КОЭФФИЦИЕНТАМИ РАСЧЕТНОЙ СКОРОСТИ

И.А. Петрухина

(г. Томск, Томский политехнический университет)

e-mail: irk-2009@yandex.ru

DETERMINATION OF ASSOCIATION RULES BETWEEN CRITERIA OF DESIGN SPEED

I.A. Petrukina

(Tomsk, Tomsk Polytechnic University)

The following article represents method of Data Mining such as market basket analysis. This method is used for finding association rules between criteria of design speed on a highway. As the result were obtained useful, trivial and incomprehensible rules.

Key words: data mining, market basket analysis, association rules, design speed.

Ассоциативные правила. Одной из наиболее распространенных задач Data Mining является исследование взаимосвязей между различными событиями для того, чтобы обнаружить ассоциации, зависимости между ними. Такой анализ называют анализом рыночной корзины, а полученные результаты – ассоциативными правилами [1]. Их основным достоинством является доступное восприятие человеком в форме логических конструкций «если..., то...», но не все правила представляют интерес. Выделяют три вида правил [2]:

- полезные – содержат действительную, ранее неизвестную информацию;
- тривиальные – содержат действительную, уже известную информацию;
- непонятные – содержат информацию, которая не может быть объяснена.

Структура данных. Целью данной работы является определение ассоциативных правил между коэффициентами обеспеченности расчетной скорости на автомобильной дороге. Входными данными являются участки автомобильной дороги, для каждого из которых определены 10 частных коэффициентов расчетной скорости K_{pcj} и нормативный показатель транспортно-эксплуатационного состояния $KП_n$. Данные K_{pcj} описывают различные характеристики автомобильной дороги [3]: