

ПОИСКОВАЯ РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА

М. А. Царегородцева, О. Б. Кудряшова

Бийский технологический институт (филиал)

Алтайского государственного технического университета им. И. И. Ползунова
г. Бийск

Современное Интернет-пространство предоставляет пользователю огромное количество разнообразной информации, в которой становится все сложнее ориентироваться, поэтому применение классических средств поиска и систематизации не может полностью удовлетворить потребности пользователя: невозможно просмотреть все материалы, чтобы выбрать для себя подходящие. В связи с этим стало появляться все больше так называемых рекомендательных систем, которые ориентированы на предоставление информации, наиболее полно удовлетворяющую интересы пользователя и наиболее полно отвечающую его запросу.

Рекомендательные системы – это программы, которые пытаются предсказать, какие объекты (книги, фильмы, музыка, веб-сайты) могут понравиться пользователю, имея определенную информацию о его профиле [1].

Такие программы используются, как правило, в коммерческих целях (в первую очередь, в Интернет-магазинах, либо на специализированных сайтах «по интересам» с целью предложения товаров). С другой стороны, актуальной задачей является интеллектуализация самого процесса поиска в Интернете. Многие пользователи Интернет объективно полагают, что современные возможности поисковых систем не позволят им найти необходимые документы или данные. Для такого мнения пользователей всемирной сети имеются следующие предпосылки:

- взрывной рост объемов доступных обществу данных вообще (увеличение числа книг, фильмов, новостей, рекламных сообщений и пр.);

- увеличение объема онлайн-данных;

- реальный объем информации, окружающей человека, значительно выше того, что он может реально пропустить через себя, чтобы обнаружить необходимую и достаточную, а также понравившуюся ему.

Поэтому необходима технология, которая помогла бы пользователю найти то, в чем он нуждается, и избежать того, на что он не желает тратить свое время и внимание.

ПОЛЗУНОВСКИЙ АЛЬМАНАХ №2 2010

Цель данного сообщения – провести анализ существующих рекомендательных коммерческих систем и методов, на которых они основаны. В дальнейшем предполагается создать поисковую систему с использованием технологий рекомендательных систем.

Прогноз в рекомендательных системах строится на основе данных о пользователе, которые обычно получают применением явных и неявных методов.

Примеры явных методов:

- запрос у пользователя оценки объекта по дифференциальной шкале;

- запрос у пользователя ранжировки группы объектов от наилучшего к наихудшему;

- предъявление пользователю двух объектов с вопросом о том, какой из них лучше;

- предложение создать список объектов, любимых пользователем.

– Примеры неявного сбора данных:

- наблюдение за тем, что осматривает пользователь в интернет-магазинах или базах данных другого типа;

- ведение записей о поведении пользователя онлайн;

- отслеживание содержимого компьютера пользователя.

На основе собранных данных рекомендательная система вычисляет список рекомендаций для конкретного пользователя.

Чаще всего рекомендательные системы основаны на механизме коллаборативной, или совместной, фильтрации. Коллаборативная фильтрация – это метод, дающий автоматические прогнозы (фильтрацию) относительно интересов пользователя по собранной информации о вкусах множества пользователей [2, 3]. Основная идея данного метода состоит в том, что те, кто согласился в прошлом, склонны соглашаться и в будущем.

Системы коллаборативной фильтрации обычно применяют двухступенчатую схему.

1. Находят тех, кто разделяет суждения прогнозируемого пользователя.

2. Используют оценки сходно мыслящих людей, найденных на первом шаге, для вычисления прогноза.

Полученные предсказания должны быть отфильтрованы согласно логике бизнеса: бесполезно предлагать пользователю файл, который у него уже есть.

Примеры сайтов, использующих рекомендательные системы:

1. Имхонет (фильмы, музыка, литература, фото)
2. Last.fm (музыка)
3. Ozon (книги, диски)
4. Imdb(фильмы)
5. Rechelper (фильмы)

Из вышеназванных сайтов наиболее интересен Имхонет, поскольку он работает с разными видами контента.

Рекомендации рассчитываются следующим образом: пользователь оценивает по десятибалльной шкале фильмы, книги и так далее. На основе выставленных оценок программа формирует его потребительский профиль. Сравнивает предпочтения разных пользователей и формирует сообщества единомышленников — людей с близкими вкусами. После чего организует обмен мнениями внутри этих сообществ.

Системы коллаборативной фильтрации обычно используют алгоритмы «ближайшего окружения (соседства)», которые устанавливают группу членов онлайн-сообществ, чьи предпочтения близки ко вкусам целевого пользователя, и оценивают объекты как среднюю величину нормализованных рейтингов этих пользователей, взвешенную по схожести. Главным недостатком этих алгоритмов является то, что в рекомендациях могут участвовать лишь те объекты, которые были предварительно оценены членами сообщества. Это не является проблемой, если объектами являются, например, книги или фильмы, для которых в системе предусмотрен механизм выставления рейтингов.

Не так просто реализовать подобный алгоритм для поиска новостной информации в сети Интернет. Ведь новости, как правило, далеко не всегда оцениваются, а если и оцениваются, то неизвестными поисковой системе пользователями. Исключением может послужить организация виртуальной социальной сети [4], в рамках которой создается такой поисковый алгоритм.

Более перспективными для целей реализации интеллектуальной поисковой систе-

мы авторам представляются явные и неявные методы, накапливающие в базе данных системы информацию о поведенческом поведении данного пользователя в сети, без привлечения коллаборативных данных о поведении сообщества. В этом случае поисковую систему можно создать в однопользовательском варианте в виде программы, устанавливаемой на персональный компьютер.

Еще один, пока не практически не используемый в поисковых системах подход — психологические тесты. Этот метод представляется перспективным также для варианта реализации поисковой системы в рамках виртуальной социальной сети, так как предполагает тестирование многих пользователей, сбор статистики «поискового поведения» пользователей и проведение корреляционного анализа данных тестирования и поискового поведения.

Таким образом, для реализации интеллектуальной поисковой системы в однопользовательском варианте применимы лишь такие методы (явные и неявные) сбора информации, которые не требуют привлечения коллаборативных данных. Методы коллаборативной фильтрации, а также психологические тесты можно и нужно применять для создания интеллектуальной поисковой системы в рамках виртуальной социальной сети [5].

СПИСОК ЛИТЕРАТУРЫ

1. Рекомендательная система / Материал из Википедии ru.wikipedia.org
2. Herlocker J.L., Konstan J.A., Terveen L.G., Riedl J.T. Evaluating Collaborative Filtering Recommender Systems // ACM Transactions on Information Systems, Vol. 22, № 1, 2004. P. 5-53.
3. Adomavicius G., Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions // IEEE Transactions on Knowledge and Data Engineering, Vol. 17, № 6, June, 2005
4. Социальная сеть / Материал из Википедии ru.wikipedia.org
5. Чирков Е.А. Сравнительный анализ научно-образовательных социальных сетей // Материалы Всероссийской научной конференции «Информационные технологии в науке, экономике и образовании» 2-3 сентября 2010 г. / под ред. О.Б. Кудряшовой; Алт. гос. техн. ун-т, БТИ. — Бийск: Изд-во Алт. гос. техн. ун-та, 2010. — с. 158-167.