

АНАЛИЗ ПОДХОДОВ К ФОРМИРОВАНИЮ РЕКОМЕНДАТЕЛЬНЫХ СИСТЕМ

© 2019 Ю. П. Преображенский, В. М. Коновалов

*Воронежский институт высоких технологий (Воронеж, Россия)
ЗАО «Лаборатория Касперского»*

В статье рассматриваются ключевые особенности работы рекомендательных систем. Рассмотрены алгоритмы рекомендательных систем, которые используются для получения корректных результатов.

Ключевые слова: рекомендательная система, управление организацией, обработка данных.

Рекомендательные системы дают возможности для того, чтобы сделать быстрее выбор по правильному варианту при управлении разными объектами.

Теперь необходимо рассмотреть, какое значение они имеют. Эти программы усовершенствовали способы взаимодействия между сайтом и посетителем, потому что вместо того, чтобы предоставлять статическую информацию, пользователь получает интерактивные возможности.

Рекомендации формируются отдельно для каждого человека, опираясь на его предыдущие действия на конкретном веб-ресурсе или на основе прошлой активности. Кроме того, значение имеет и поведение предыдущих участников процесса.

Для интернет-магазинов это в принципе важная функция, а для таких крупных каталогов типа Amazon – один из немногих способов качественно работать.

Способ рекомендации в данном случае не является обычной дополнительной опцией, она обеспечивает удобство навигации пользователя по веб-ресурсу. Если электронный каталог содержит более 20 000 наименований продукции, ориентация уже представляется непомерно тяжелой, что говорить, если товаров миллионы?

Насколько утомляет потенциального покупателя взаимодействие с подобным сайтом? Ответ очевиден [1, 2]. И на помощь приходит виджет поиска товаров, визуально похожих на искомый, или принадлежащий одной группе изделий, или комплиментарная продукция (когда к паре туфель предла-

гают выбрать сумочку, например). Такое решение увеличивает не только число просмотров, это положительно влияет на конверсию [3].

Как показывает практика, не только онлайн-магазины используют подобный прием. Социальные сети также не отстают.

Также подобные приемы легко можно увидеть на разных социальных платформах, порталах, посвященных литературе, путешествиям, на новостных ресурсах, в интернет-магазинах, словом – почти везде. Эта методика действительно очень популярна.

Рекомендательные сервисы собирают различную информацию о человеке, используя несколько методов, по которым и разделяют все системы.

Первый тип – явный сбор данных. Как можно было догадаться из названия, пользователь сам предоставляет необходимые для работы материалы. Например, когда рекомендательные системы Яндекс или других поисковиков просят человека дать оценки разным элементам, составить список фаворитов определенной сферы или же ответить на несколько вопросов [4, 5]. Если же человек отказывается дать информацию самостоятельно, актуальной будет следующая методика.

Второй тип – неявный сбор данных. Условно говоря, это шпионская миссия, согласно которой действия участника процесса фиксируются программой для дальнейшей обработки и применения. Что нужно для этого? Программа распознает покупки, оценки на сайтах, собирает информацию по просмотрам, комментариям. Конечно, выбор такой методики ведет за собой некоторые этические проблемы, ведь защита персональных данных – одно из главных требований, предъявляемых пользователем поисковику. Но пока факт остается фак-

Преображенский Юрий Петрович – Воронежский институт высоких технологий, к. т. н., профессор, petrovichyur@yandex.ru.

Коновалов Виталий Михайлович – ЗАО «Лаборатория Касперского», специалист, dtenis0t62vrwet@yandex.ru.

том – своеобразная слежка возможна, и рядовые посетители сайтов проверить, действительно ли ведутся подобные мероприятия, не могут.

Существуют также виды рекомендательных систем, определяемых по подходам, которые они применяют.

Первая базовая методика называется коллаборативной фильтрацией (collaborative filtering). Рекомендации с использованием данной методики выдаются, основываясь на поведенческих характеристиках одного человека или группы людей, последнее даже является более эффективным. В группы собираются люди, которые похожи между собой по поведению и характеристикам [6, 7].

Приведем пример, чтобы информация воспринималась проще.

Создается сайт, где аудитории будут рекомендоваться музыкальные произведения. Как в данном случае будут работать сервисы рекомендации на основе коллаборативной методики? По такому принципу: за основу возьмут одно сообщество, где участники добавляют в плей-лист одинаковые по жанровой принадлежности треки. Далее, определяются самые популярные из всех музыкальных произведений и рекомендуются одному пользователю из группы, который пока еще не слушал эту мелодию.

Второй подход называется контентной фильтрацией (content-based filtering). Здесь рекомендация формируется исходя из поведения человека. При использовании этого подхода также может браться за основу история просмотров конкретного участника.

На этот раз приведем пример с тематическими онлайн-журналами. Итак, в случае, когда ранее человек читал материалы о горных велосипедах и регулярно комментировал статьи в блогах такого содержания, то методом фильтрации содержимого будет использоваться эта прошлая информация для определения схожих ресурсов и предложения их в качестве рекомендации этого пользователя.

Выделяют также смешанные подходы [8, 9], в соответствии с которыми осуществляется разработка рекомендательной системы.

Смешанный подход – это сочетание коллаборативной и контентной фильтрации. Как известно, больше – лучше, поэтому смешение этих двух методик увеличивают эффективность систем рекомендации, а именно значительно повышают точность прогнозов для конкретных людей.

Ниже будут рассмотрены алгоритмы рекомендательных систем, которые используются для получения корректных результатов.

Корреляция Пирсона.

Этот алгоритм позволяет выделить общие характеристики между несколькими пользователями. Каким образом? С помощью простой математики, а именно определением линейной зависимости между двумя элементами. Важный момент — такая методика не подходит для сообщества людей.

Кластеризация.

Этот принцип работы рекомендательных систем основывается на выделении сходства между элементами (пользователями) путем вычисления их близости друг другу в так называемом пространстве признаков. Признаками выступают те элементы, по которым сходятся интересы определенных участников процесса (для музыкальных ресурсов это треки, для кинопорталов — фильмы). Схожие по характеристикам пользователи объединяются в так называемые кластеры.

Алгоритм совместной фильтрации.

Жесткую кластеризацию можно заменить и другим алгоритмом, который работает по довольно сложной формуле, и также как и все предыдущие, основывается на поведении пользователей из его группы. Однако в этой методике есть несколько довольно существенных минусов. Во-первых, новым или нетипичным пользователям (которые не объединяются в группы) сложно найти рекомендации. Во-вторых, так называемый «холодный старт», когда новые объекты не попадают в рекомендательные системы [10].

Алгоритм фильтрации содержимого.

Алгоритм, симметричный предыдущему, но если в первом случае мы отталкивались от предположения, что объект понравится пользователю, потому что он нравится его «однотруппникам», то тут мы будем рекомендовать на основе похожих объектов, которые он уже отметил для себя. И здесь уже традиционно можно выделить несколько проблем. Тот же «холодный старт» и то, что рекомендации часто обыденны.

ЛИТЕРАТУРА

1. Завьялов, Д. В. О применении информационных технологий / Д. В. Завьялов

// Современные наукоемкие технологии. – 2013. – № 8-1. – С. 71-72.

2. Черников, С. Ю. Использование системного анализа при управлении организациями / С. Ю. Черников, Р. В. Корольков // Моделирование, оптимизация и информационные технологии. – 2014. – № 2 (5). – С. 16.

3. Преображенский, Ю. П. О повышении эффективности работы промышленных предприятий // Исследование инновационного потенциала общества и формирование направлений его стратегического развития. Сборник научных статей 8-й Всероссийской научно-практической конференции с международным участием. – 2018. – С. 45-48.

4. Будко, Н. А. Применение ИНС в интерфейсах человек-машина / Н. А. Будко, Р. Ю. Будко, А. Ю. Будко // Моделирование, оптимизация и информационные технологии. – 2019. – Т. 7. – № 1 (24). – С. 328-340.

5. Свиридов, В. И. Лингвистическое обеспечение автоматизированных систем управления и взаимодействие пользователя с компьютером / В. И. Свиридов, Е. И. Чопорова, Е. В. Свиридова // Моделирование, оптимизация и информационные технологии. – 2019. – Т. 7. – № 1 (24). – С. 430-438.

6. Кизим, А. В. Программный комплекс поддержки модернизации технических систем / А. В. Кизим, А. В. Матохина, А. Г. Кравец, И. П. Мединцева // Моделирование, оптимизация и информационные технологии. – 2019. – Т. 7. – № 2 (25). – С. 311-324.

7. Гуськова, Л. Б. О построении автоматизированного рабочего места менеджера / Л. Б. Гуськова // Успехи современного естествознания. – 2012. – № 6. – С. 106.

8. Преображенский, А. П. Возможности обеспечения развития предприятий / А. П. Преображенский // В мире научных открытий. – 2015. – № 10 (70). – С. 196-201.

9. Преображенский, Ю. П. Проблемы управления в производственных организациях / Ю. П. Преображенский // Актуальные проблемы развития хозяйствующих субъектов, территорий и систем регионального и муниципального управления. Материалы XIII международной научно-практической конференции. Под редакцией Ю. В. Вертаковой. – 2018. – С. 208-211.

10. Петрашук, Г. И. Маркетинг в прикладном менеджменте / Г. И. Петрашук // В мире научных открытий. – 2010. – № 4-7 (10). – С. 35-36.

ANALYSIS OF APPROACHES TO FORMATION RECOMMENDED SYSTEMS

© 2019 Yu. P. Preobrazhenskiy, V.M. Konovalov

Voronezh institute of high technologies (Voronezh, Russia)
Kaspersky Lab CJSC

The paper discusses the key features of the recommendation systems. The algorithms of recommendation systems that are used to obtain correct results are considered.

Key words: recommendation system, organization management, data processing.