Title: | Subtitle: | Project: | Author: Hilton Lipschitz Affiliation: Noverse LLC Web: http://www.noverse.com Date: June 18, 2012

# Рекомендательные информационные системы в ывфвыа

Рекомендательные системы (РС) — программные системы, предназначенные для оценки уровня интереса пользователя к определенному продукту или сервиса, на основе имеющейся информации о пользователе и/или продукте и сервисе.

**Оценка? Цель ее нахождения? Интерес? В чем выражен? Какие продукты бывают….? Имеющаяся информация? Откуда берется? Где хранится, как обрабатывается? В каком виде и для чего решение?**

В области рекомендательных систем используется специальная терминология. В частности, *профилем пользователя*, *продукта или услуги*, являются данные, характеризующие пользователя, продукт или услугу. Именно эти данные используются в процессе оценивания *релевантности продукта* или услуги к желаниям пользователя, называемого *фильтрацией*. В результате фильтрации продукты или услуги ранжируются в соответвии с полученной оценкой, а пользователю предоставляется некоторое конечное подмножество, элементы которого имеют максимальную релевантность, т.е. оцениваются как наиболее интересные пользователю. Далее под *интересом* будем понимать именно интерес пользователя к продукту или услуге.

Существует ряд подходов к оценке интереса:

1. на основе *фильтрации содержания*, при этом в информационной системе создаются профили пользователей (продуктов и услуг), включающие социальный статус пользователя, возраст, место проживания, род деятельности, а также характеристики, выражающие интерес пользователя к продукту или услуге; профили продуктов и услуг включают позицию в системе классификации, потребительские характеристики продукта.
2. на основе коллаборативной фильтрации, где используется информация о поведении пользователей в прошлом, например, перечень его покупок или оценок товаров или услуг, сделанных на сайте интернет-магазина в прошлом, при этом аналитическим блоком информационной системы автоматически формируются классификации товаров и услуг, а также ранжируются атрибуты по степени значимости в оценке интереса.
3. смешанные методы, которые базируются на подходах пп. 1 и 2 одновременно.

Например, в Music Genome Project музыкальный аналитик оценивает каждую композицию по сотням различных музыкальных характеристик, при помощи которых выявляются музыкальные предпочтения пользователя. Перечень оценок формирует *профиль музыкального произведения*. Основная проблема первого типа рекомендательных систем — это работоспособность системы на начальном этапе ее эксплуатации, так называемый "*холодный старт*". Например, для новых пользователей в системе нет необходимой информации в профиле для принятия решения о том, какие товары ему следует предлагать. В связи с этим в современных рекомендательных системах реализуется механизм сбора и анализа данных о пользователях с применением *явных* и *неявных методов*.

Явные методы сбора данных выполняют следующие действия:

* запрос у пользователя оценки объекта по некоторой 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, характеризующих предпочтения пользователя.

Примерами неявного сбора данных выступают:

* наблюдение за тем, что просматривает пользователь в интернет-магазине или базе данных;
* **ведение записей о поведении пользователя онлайн**;
* **отслеживание содержимого компьютера пользователя**.

Рекомендательные системы сравнивают однотипные данные, полученные от разных людей и вычисляют список рекомендаций для конкретного пользователя. Примеры коммерческого и некоммерческого использования РС приведены в статье [???]. Для вычисления рекомендаций используется граф интересов [3]. Рекомендательные системы — **удобная** альтернатива поисковым алгоритмам, так как позволяют обнаружить объекты, которые не имеют непосредственного отношения к поисковому запросу. Любопытно, что рекомендательные системы часто используют поисковые машины для индексации необычных данных.

## Методы фильтрации содержимого

В статье [Дьяконов] рассмотрена задача разработки алогоритмов оценивания **(Прочесть статью, понять, что это первый подход, используется (какой?) явный/неявный метод сбора информации о видеолекциях). Как решена задача? Как внедрено, и было ли? Какие недостатки?**

## Виды рекомендательных систем

Можно выделить два основных типа рекомендательных систем. Их, конечно же больше, но мы сегодня будем рассматривать именно эти и в особенности коллаборативную фильтрацию.

Пользователю рекомендуются объекты, похожие на те, которые этот пользователь уже употребил. Похожести оцениваются по признакам содержимого объектов. Сильная зависимость от предметной области, полезность рекомендаций ограничена. Методы коллаборативной фильтрации

Для рекомендации используется история оценок как самого пользователя, так и других пользователей. Более универсальный подход, часто дает лучший результат. Есть свои проблемы (например, холодный старт).

Рекомендательные системы появились в интернете достаточно давно, около 20 лет назад. Однако настоящий подъем в этой области случился примерно 5-10 лет назад, когда произошло соревнование Netflix Prize. Компания Netflix тогда давала в прокат не цифровые копии, а рассылала VHS-кассеты и DVD. Для них было очень важно повысить качество рекомендаций. Чем лучше Netflix рекомендует своим пользователям фильмы, тем больше фильмов они берут в прокат. Соответственно, растет и прибыль компании. В 2006 году они запустили соревнование Netflix Prize. Они выложили в открытый доступ собранные данные: около 100 миллионов оценок по пятибалльной шкале с указанием ID проставивших их пользователей. Участники соревнования должны были как можно лучше предугадывать, какую оценку поставит определенному фильму тот или иной пользователь. Качество предсказания измерялось при помощи метрики RMSE (средне-квадратичное отклонение). У Netflix уже был алгоритм, который предсказывал оценки пользователей с качеством 0.9514 по метрике RMSE. Задача была улучшить предсказание хотя бы на 10% — до 0.8563. Победителю был обещан приз в $ 1 000 000. Соревнование длилось примерно три года. За первый год качество улучшили на 7%, дальше все немного замедлилось. Но в конце две команды с разницей в 20 минут прислали свои решения, каждое из которых проходило порог в 10%, качество у них было одинаковое с точностью до четвертого знака. В задаче, над которой множество команд билось три года, все решили каких-то двадцать минут. Опоздавшая команда (как и многие другие, участвовавшие в конкурсе) остались ни с чем, однако сам конкурс очень сильно подстегнул развитие в этой области.

В настоящее время развиваются методы на основе анализа структуры онтологии предметной области. ….. Подходы на онтологии позволяют решать проблему Холодного старта в той или иной мере. (Черкашин)

Предметная область Например в Иркутске можно зяняться следующими задачами, где разработка Рсистемы позволит автоматизировать….

Литература Melville P., Mooney R., Nagarajan R. Content-Boosted Collaborative Filtering for Improved Recommendations (англ.) // University of Texas, USA : Материалы конф. / AAAI-02, Austin, TX, USA, 2002. — 2002. — P. 187-192. Жернакова О. Системы рекомендаций и поиска видеоконтента // Телемультимедиа, 2012. Nadim Hossain. Why the Interest Graph Is a Marketer’s Best Friend (англ.). Mashable. Проверено 7 декабря 2013. wikipedia