

Практическая работа № 5

«Рекомендательные системы на основе контента и знаний»

по дисциплине «Разработка обеспечивающих подсистем систем поддержки принятия решений»

Цели: приобрести навыки реализации методов рекомендательных систем, основанных на контенте и знаниях о предметной области.

Задачи:

1) создать программную реализацию рекомендательной системы (РС), основанную на поиске сходств в предлагаемом пользователям контенте и связанной с ним предметной областью, включающую:

- актуальную предметную область для применения РС (вроде маркетплейса, медиа ресурсов, соц. сетей, экономической сферы и т.д.) и набор начальных данных для неё: опрос покупателей, статистика за временной период и т.д., включающие набор признаков, связанных с контентом, предлагаемым пользователям в системе (теги, типы, сопутствующие названия, описание и т.д., см. Примечание 1) и описание знаний о предметной области контента (сопутствующие товары, домены, поиск и т.п. см. Примечание 2);

- реализацию поиска сходств, классификации или кластеризации контента на основе коллаборативных методов (см. Примечание 3), а также собственных методов с последующим выводом пользователю рекомендаций;

- реализацию поиска релевантных товаров на основе знаний (например, методами case-based, critique-based, utility-based и т.п.) с выводом рекомендаций пользователю, сравнить результаты разных методов;

- сравнение результатов предложенных в настоящей работе методов с результатами предыдущих практических работ (практические работы 1-4).

2) в качестве дополнительного задания (на доп. баллы) выполнить:

- реализацию гибридной рекомендательной системы, основанной на коллаборативной фильтрации пользователей, контента и знаний о

предметной области рекомендуемых товаров, можно совместить результаты всех работ курса с возможностью переключения режимов рекомендаций (чтобы была возможность переключаться между рекомендациями на основе пользователей, контента и знаний) и поиска на основе синергии всех реализованных за курс методов рекомендательных систем.

ПРИМЕЧАНИЕ:

1. Рекомендательные системы на основе контента создают профиль пользователя, опираясь на интересующий или покупаемый им контент, основой таких систем является матрица полезности, которая описывает отношения между пользователями и интересующими их предметами, векторы предпочтений, образующих матрицу, объединяют в себе информацию о параметрах контента и его оценку для пользователя, такими параметрами могут являться размеры, материалы, стоимость, жанры, год выпуска, рейтинг и т.д., при этом в матрицу, как правило, записывается не информация формата «пользователь/размер», а записывается информация в ячейку формата «пользователь/размер XL» (то есть насколько положительно пользователь относится к товару конкретного размера) в виде числового индекса (как положительного, так и отрицательного), причём чем выше значение индекса, тем более важен для пользователя данный параметр.

2. Рекомендательные системы, основанные на знаниях, получают эти знания о предметной области путём прямого получения обратной связи пользователя (диалоговая рекомендация), когда пользователь проходит определённое тестирование или самостоятельно выстраивает свою область интересов, путём уточняющего поиска (РС на основе ограничений), который представляет собой процесс поиска набора ограничений, составляющего шаблон искомого пользователем предмета (по сути – неявный фильтр), путём навигационных рекомендаций, где пользователь за счёт своей критики относительно рекомендуемого или интересующего его в настоящий момент элемента определяет движение (конкретизацию) рекомендательного

алгоритма в сторону наиболее релевантного контента, а также путём доменной системы, когда для каждого предмета определяется домен (предметная область), в котором он сосуществует с другим предметами, и рекомендация пользователю предоставляется на основе его движения или просто существования в этом домене (например, при покупке монитора пользователь рано или поздно также, скорее всего, купит к нему дополнительные провода, средство для протирки экранов, салфетки или тряпочку, кронштейн, плёнку и т.д.), при этом хорошим тоном является отложенная рекомендация, когда все эти предметы не разом вываливаются на пользователя, а появляются через определённое (статистически посчитанное на основе опыта других пользователей) время, когда они будут наиболее необходимы и у пользователя будут на них деньги.

3. Рекомендации на основе контента с технической точки зрения принято делить на две группы:

а) методы на основе векторного пространства: как и написано в примечании 1, создаётся профиль пользователя в виде векторов каждого интересующего его предмета, полученные векторы являются входными данными для коллаборативных методов, при этом в основном используются именно анамнестические методы (например, расстояние Жаккара или косинусное), т.к. в отличие от пользователей основные данные о контенте известны с момента его публикации, что даёт возможность избежать «холодного старта»;

б) методы на основе кластеризации и классификации: здесь главную роль играет не профиль пользователя, описываемый его контентом, а вектор атрибутов именно самого контента, далее в случае, если известны домены и существует возможность отнести контент к определённому домену по его вектору, производится его классификация либо математическими методами, либо навигационным или поисковым методами (как и в системах, основанных на знаниях), в ином случае необходимо выполнять кластеризацию соответствующими методами.