

СОЗДАНИЕ РЕКОМЕНДАЦИОННОЙ СИСТЕМЫ УЧЕБНОГО ТИПА С ИСПОЛЬЗОВАНИЕМ ФРЕЙМВОРКА WINDOWS COMMUNICATION FOUNDATION

Глибовец Николай Николаевич, Сидоренко Марина Олеговна

Аннотация: В рамках работы исследуются технологии, которые используются для создания рекомендационных систем. Определены особенности рекомендательных систем, предложена их классификация в соответствии с алгоритмами фильтрации и сферами их применения. Предложена агентная архитектура системы для рекомендации выборочных курсов на базе фреймворка Windows Communication Foundation.

Ключевые слова: рекомендационная система, фильтрация по содержанию, коллаборативная фильтрация, Windows Communication Foundation

ACM Classification Keywords: I.2.11 Artificial Intelligence- Multiagent systems

Введение

На сегодняшний день, наблюдается стремительное совершенствование технологий рекомендационных систем. Впервые опробованные в среде электронной коммерции, на данный момент они используются в различных сферах, в частности в системах электронного образования, благодаря целому ряду преимуществ - от повышения эффективности поиска информации, до содействия дистанционному обучению. Первые разработки рекомендательных систем берут свое начало в 90-х годах [Adomavicius, 2005], однако толчком, который вывел создание рекомендательных систем на новый уровень, является конкурс Netflix Prize [Bennett, 2007], который был организован в 2006 году компанией Netflix.

Рекомендательные системы помогают пользователям, которые столкнулись с проблемой огромного выбора, определить, что именно лучше всего отвечает их потребностям. Студент, перед которым стоит проблема выбора дисциплины обучения, является ярким представителем пользователя, которому не хватает опыта для самостоятельного выбора. С помощью рекомендательных систем, учебный процесс студента будет спланирован более продуктивно и эффективно.

Наиболее развитые системы изучают предпочтения каждого пользователя и предоставляют персонализированные рекомендации. В процессе работы эти системы собирают данные о пользователях, используя сочетание явных и неявных методов. Рекомендательные системы приобретают все большее распространение как перспективная технология. Ее использование позволяет пользователю решить проблему выбора среди большого количества альтернатив.

В работе анализируются технологии, которые используются для создания рекомендационных систем. Предложена агентная архитектура системы для рекомендации выборочных курсов на базе фреймворка Windows Communication Foundation. Описан созданный прототип рекомендательной системы по выбору учебных курсов, построенных на основе рекомендаций, полученных от других студентов. Обоснованы выбранные методы представления профиля студента, алгоритмы рекомендаций, способы оценки сходства студентов.

Обзор рекомендательных систем

Рекомендательная система - система, которая дает рекомендации пользователю [Fasli, 2007]. Такое очень обобщенное определение рекомендательных систем, которое встречается в научной литературе чаще чем другие, говорит о том, что под эту категорию попадает достаточно большое количество различных технологий. Рекомендательная система позволяет пользователю замечать свои предпочтения

и возвращает результаты, которые полезны для него, основываясь на оценках других пользователей и предположениях самой системы. В отличие от поисковых систем, чтобы получить ответ, рекомендательная система не требует четкого запроса. Пользователю предлагается оценить некоторые объекты из коллекции и на основании его оценок и сравнения их с оценками других пользователей строятся предположения и возвращаются наиболее близкие к ним результаты. Более подробное описание и спецификацию рекомендательной системы указывают при описании конкретного способа ее реализации. Рекомендательную систему с коллаборативной фильтрацией определяют как систему, которая «предлагает пользователям определенные предметы, потому что эти предметы были высоко оценены другими пользователями со схожими вкусами» [Uchyigit, 2003].

В каждой рекомендательной системе мы имеем дело с пользователем, которому предоставляется много альтернатив, среди которых ему необходимо осуществить свой выбор. Пользователю может не хватать опыта и знаний для того, чтобы самостоятельно отбросить альтернативы, которые не соответствуют его потребностям. Пользователь в определенной форме, эксплицитно или имплицитно, предоставляет системе информацию о своих предпочтениях, при этом о некоторых альтернативах он может даже и не знать. Таким образом, рекомендательная система представляется как система, которая использует определенный алгоритм фильтрации и существующие сведения о потребностях пользователя, для рекомендации ему набора альтернатив, которые считает наиболее полезными для него.

Для разработки рекомендательной системы используют два типа фильтрации - коллаборативную фильтрацию и фильтрацию за содержанием. Также существует отдельный тип, который объединяет черты двух предыдущих - гибридная фильтрация [Fasli, 2007; Uchyigit, 2003; Montaner, 2003].

Для простоты описания РС будем использовать термин «объект», который обозначает то, что рекомендует система.

РС, которые разработаны на основе демографической фильтрации, используют описания пользователей для установления связи между определенным объектом и категорией пользователей, которые заинтересованы данным объектом. На основе стереотипной классификации создаются профили группы пользователей. При использовании этого метода фильтрации пользователь должен заполнить анкету при регистрации, информация из которой будет использована для отнесения пользователя к определенной группе.

Система с таким типом фильтрации основана на том, что обобщает интересы пользователей и рекомендует одинаковые объекты всем пользователям, которые принадлежат к одной группе. Система с демографической фильтрацией требует информацию об интересах пользователя, но дает не персонализированные рекомендации, а лишь обобщенные.

Демографический подход не предполагает никакого механизма адаптации профиля пользователя к изменениям в его интересах. При регистрации пользователя относят к определенной предметной категории со схожими интересами, а последующее их изменение никак не учитывается при предоставлении рекомендаций в будущем. Для решения такой проблемы, должно быть внедрено дополнительное редактирование регистрационной формы, в противном случае пользователь должен будет зарегистрировать новую анкету. Преимуществом этого метода считается его относительная простота реализации. Чаще демографическая фильтрация используется в комплексе с другими видами фильтрации, что приводит к улучшению предоставленных рекомендаций.

Неперсонализированные системы, рекомендуют пользователю определенные объекты, используя при этом общий рейтинг этого объекта, который определяется с помощью голосования всех остальных пользователей системы. Рекомендация такого характера не зависит от пользователя, который в ней нуждается. Результатом использования такой системы, будут одинаковые рекомендации выдаваемые на один и тот же запрос. Преимуществом такой системы является минимальное взаимодействие пользователя и системы.

В рекомендательных системах, в которых используется фильтр за содержанием, пользователи являются независимыми от других участников системы. Для того чтобы система сгенерировала рекомендации, ей нужен профиль пользователя, в котором будут отражены его интересы. В профиле в определенной форме (которая зависит от выбранного способа представления и выбранных алгоритмов) хранится информация об объектах (свойствах в виде атрибутов), которыми интересуется пользователь [Protogeros, 2007]. Также система содержит информацию обо всех предметах, которые она может рекомендовать. Такая система берет за основу описание предметов по профилю пользователя, и находит в собственной базе данных похожие предметы, после чего рекомендует их ему. Подход этого типа фильтрации очень уместен, когда у пользователя есть четко определенные, специфические заинтересованности, и он ищет аналогичные рекомендации. Преимуществом фильтрации за содержанием является то, что для начала предоставления рекомендаций не требуется большого количества зарегистрированных пользователей, то есть рекомендации не зависят от других пользователей системы. Главным ограничением этого метода является неспособность системы с таким видом фильтрации рекомендовать новые объекты, которые не следуют из интересов пользователя.

Для того, чтобы система имела высокое качество предоставления рекомендаций, она должна постоянно иметь обратную связь с пользователем, и узнавать релевантность предоставленных рекомендаций [Lin, 2007]. Также, эти сведения идут на обновление профиля пользователя. Но возникает другая проблема - пользователь всегда хочет свести к минимуму взаимодействие с системой и неохотно делиться информацией для обратной связи. Это типичная проблема для всех типов рекомендательных систем. А поскольку рекомендации целиком базируются на информации из профиля пользователя, то чем меньше сведений предоставляет пользователь, тем менее полезным будет набор рекомендаций, которые ему предложит система. Чтобы решить эту проблему используются различные техники автоматического сбора информации для поддержки профиля, что затрудняет реализацию такой системы и не всегда является эффективным. Рекомендационные системы с фильтрацией за содержанием чаще всего используются в системах электронной коммерции. Этот метод применяется также в сочетании с коллаборативной фильтрацией для более ценных рекомендаций в других сферах - электронное образование, программирование и прочее.

Рекомендательные системы с использованием коллаборативной фильтрации отыскивают пользователей с общими интересами и рекомендуют те объекты, которые были высоко оценены такими пользователями. Подобный тип фильтрации похож на то, как мы спрашиваем совет у коллег и друзей, поэтому его иногда называют социальным фильтрованием. Интерес каждого пользователя сохраняется в виде системы рейтингов, в которых видно насколько пользователю нравится или наоборот, не нравится тот или иной предмет. Такая система не обязательно должна иметь подробное описание предмета, ей вполне хватает лишь оценки пользователей о нем. Преимущество такой системы, в отличие от фильтрования за содержанием, заключается в том, что система не генерирует рекомендации исходя лишь из интересов одного пользователя.

Рекомендации в системе с коллаборативной фильтрацией в значительной степени зависят от формулировки «сходства» пользователей. Эта мера «сходства» должна быть установлена определенным образом. Для этого, рекомендательные системы используют различные метрики - векторное сходство, коэффициенты Пирсона, байесовскую вероятностную классификацию [Fasli, 2007]. Одним из наиболее используемых алгоритмов для определения сходства является алгоритм «ближайшего соседа» [Bedil, 2006].

Главная проблема, которая возникает при разработке рекомендательной системы с коллаборативной фильтрацией - это проблема «холодного начала» [Montaner, 2003] [Fasli, 2007]. Система не будет способна эффективно генерировать рекомендации, если большинство из пользователей не укажут свои интересы в профиле. Другим важным моментом, который необходимо учитывать при разработке подобных систем является то, что она должна быть масштабируемой.

Гибридные системы обычно комбинируют коллаборативную фильтрацию и фильтрацию за содержанием, что позволяет решить ряд проблем, имеющих место при применении этих методов по отдельности.

Далее будет рассмотрено, каким образом сочетание коллаборативной фильтрации и фильтрации по содержанию может решить эти проблемы. Недостатком систем с фильтрацией по содержанию является их ретроспективный характер, то есть неспособность порекомендовать новые предметы. С этим недостатком способна справиться коллаборативная фильтрация. Рекомендации, сгенерированные с помощью коллаборативной фильтрации, создаются на основе профилей остальных пользователей, поэтому не всегда похожи на объекты, которые активный пользователь указал в своем профиле.

Одновременно, весомым недостатком системы с коллаборативной фильтрацией является неспособность давать рекомендации пользователям с нетипичными и оригинальными вкусами, учитывая нехватку пользователей с похожими интересами, которые стали бы основой для генерации рекомендаций. В таких случаях эффективнее фильтрация по содержанию. Такой тип фильтрации также решает проблему «холодного начала». Пока система не наберет достаточного количества пользователей, используется фильтрация по содержанию, а когда достаточное количество пользователей будет достигнуто, система начинает работать с методами коллаборативной фильтрации.

В гибридной системе информация об интересах пользователей представлена в профиле в двух видах - как набор атрибутов определенного предмета и как его оценка пользователем [McCarey, 2006]. Такая характеристика является одновременно и преимуществом и недостатком системы. Преимущество заключается в том, что больше сведений дают возможность использовать эффективные алгоритмы фильтрации и генерировать полезные рекомендации. Недостаток такой системы заключается в том, что пользователю нужно ввести больше информации, что делается всегда очень неохотно, или вообще не делается.

К сожалению, сочетание фильтрации по содержанию и коллаборативной фильтрации не решает проблем масштабируемости системы и защиты приватности.

Разработка архитектуры и создание прототипа рекомендательной системы

Windows Communication Foundation - название платформы следующего поколения, ранее известной под кодом Indigo, для коннекционных систем с использованием расширения API.NET. WCF делает возможным построение безопасных, надежных и транзакционных систем через упрощенную унифицированную программную модель межплатформенного взаимодействия. WCF предоставляет единую инфраструктуру для разработки, что повышает производительность и снижает расходы на создание безопасных, надежных и транзакционных Web-служб нового поколения. Заложенные в ней принципы интероперабельности позволяют легко достичь взаимодействия с другими платформами, для чего используются технологии взаимодействия платформ, например WSIT, разрабатываемых на базе источников открытого кода.

Сейчас на рынке существует несколько альтернатив, которые могли бы использоваться как инструмент для построения рекомендательной системы. Основными являются Java RMI, Microsoft WCF, Microsoft Remoting. Для построения рекомендательной системы учебного типа и реализации сетевой коммуникации была выбрана технология .NET Windows Communication Foundation. Этот выбор был сделан, руководствуясь следующими критериями:

- Интероперабельность - дает возможность сотрудничать с платформами, построенными на других операционных системах и средствах коммуникации;
- Удобство использования и настройки необходимых параметров связи и функционирования;
- Поддержка общераспространенного и удобного языка C# и .NET платформы, а также кодов описанных другими языками программирования, входящих в список Common Language Runtime;
- Оптимальное использование ресурсов аппаратной платформы.

Главным преимуществом использования Windows Communication Foundation является ее способность к расширению и изменению системы при переносе в веб, поскольку нет необходимости полностью перерабатывать ее.

Прототип РС состоит из трех основных частей: библиотеки классов RecService, сервера (ConsoleHost), клиента (ConsoleClient). В библиотеке классов RecService описана логика предоставления рекомендаций для студента (RecommendationEngine). Она содержит основные классы для описания поведения студента (Student) и предмета (Course). Здесь реализован класс DB, содержащий логику для работы с базой данных, что может считывать и записывать данные о студентах, предметах и связях между ними. Сервер агрегирует библиотеку RecService и предоставляет клиенту доступ к объекту типа RecommendationEngine. Клиент посылает запрос к серверу и отображает пользователю рекомендованные курсы по заданному имени студента.

Профиль студента реализован как набор прослушанных им курсов. Дальнейшая разработка рекомендательной системы позволит добавить к профилю различные данные о студенте, которые предоставят более точные рекомендации о выборочных курсах.

Для оценки сходства студентов прототип рекомендательной системы использует модель векторного пространства. Профиль студента представлен как вектор состоящий из булевых значений, обозначающих, какие курсы этот студент прослушал. Уровень сходства между двумя студентами определяется как отношение совместно прослушанных курсов к общему количеству прослушанных курсов ими обоими.

Похожими считаются те студенты, уровень сходства для которых больше чем 68.2% (в окрестности δ).

Модель векторного пространства - распространенная техника, которая давно используется для представления документов в различных технологиях информационного поиска [Bedil, 2006]. Выбор векторной схожести для вычисления сходства студентов обусловлен тем, что именно эта техника продемонстрировала лучшие результаты при проведении ряда экспериментов [Microsoft, 2007], т.е. по отзывам пользователей, рекомендательная система производила больший процент полезных рекомендаций, по сравнению с другими техниками.

По описанной выше схеме определяется список студентов схожих с данным студентом. Далее определяется список курсов, которые не были прослушаны этим студентом, но были прослушаны хотя бы одним из списка похожих. Он сортируется по частоте убывания среди подобных студентов. Тогда из них выбирается N первых предметов, и рекомендуются этому студенту.

РС начинает работу по запросу пользователя. Взаимодействие пользователей с рекомендационной системой происходит через консольный интерфейс, который в дальнейшей работе, благодаря преимуществам WCF, может быть переделан в веб-интерфейс. Программа способна, имея список курсов прослушанных определенным студентом, предоставить ему рекомендации относительно других курсов, которые он еще не слушал.

Для создания рекомендательной системы было использовано среду Visual Studio 2010 на базе .NET Framework 4.0 с использованием технологии WCF (Windows Communication Foundation).

Версия Visual Studio 2010 была избрана благодаря удобному интерфейсу с использованием Windows Presentation Foundation (WPF), внедрению следующего поколения инструментов ASP.NET, наличию поддержки динамических расширений в языках программирования C # и Visual Basic, использованию новых шаблонов проектов, инструментария для документирования тестовых сценариев и большого количество новых библиотек, поддерживающих Windows 7.

Заключение

Рекомендательные системы - технология, которая стремительно развивается, эффективно используется в различных сферах, в частности в среде электронного обучения. Преимущества WCF, как следующего шага развития технологии .Net Remoting существенные, поэтому этот способ организации коммуникации прекрасно подходит для разработки рекомендательных систем.

В рамках работы исследовались рекомендательные системы в коллаборативных средах. Определены особенности рекомендательных систем, предложена их классификация в соответствии с алгоритмами фильтрации. Проанализированы алгоритмы фильтрации и сферы их применения. Обоснована

целесообразность использования C# вместе с .NET для создания прототипа с использованием фреймворка Windows Communication Foundation.

Дальнейшая работа предполагает усовершенствование алгоритма и интеграцию в систему электронного образования, для предоставления рекомендаций студентам по выборочным курсам. Также, возможным продолжением работы является создание графического интерфейса пользователя (GUI).

Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария www.ithea.org и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина www.aduis.com.ua.

Список литературы

- [Adomavicius,2005] G. Adomavicius, A. Tuzhilin . Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. In: IEEE Transactions of Knowledge and Data Engineering, 2005.
- [Bedil,2006] P. Bedil, H. Kaur. Trust based Personalized Recommender System, 2006.
- [Bennett, 2007] J. Bennett, S. Lanning. The Netflix Prize. In Proc. of KDD Cup Workshop at SIGKDD-07, 13th ACM Int. Conf. on Knowledge Discovery and Data Mining, 2007.
- [Fasli, 2007] M. Fasli. Agent Technology for E-Commerce. In: Recommender Systems. John Wiley & Sons Ltd, 2007.
- [Lin, 2007] H. Lin, Architectural Design of Multi-Agent Systems—Technologies and Techniques. Task Allocation in Case-Based Recommender Systems, 2007.
- [McCarey, 2006] F. McCarey, M. 'O Cinn'eide, N. Kushmerick. RASCAL: A Recommender Agent for Software Components in an Agile Environment, 2006.
- [Microsoft, 2007] Microsoft. Deploying an Internet Information Services-Hosted WCF Service, 2007
- [Montaner, 2003] M. Montaner. Collaborative Recommender Agents Based On Case-Based Reasoning and Trust, 2003.
- [Protogeros, 2007] N. Protogeros. Agent and Web Service Technologies in Virtual Enterprises, A Recommender Agent to Support Knowledge Sharing in Virtual Enterprises, 2008.
- [Uchyigit,2003] G. Uchyigit, K. Clark. A Multi-Agent Architecture for dynamic collaborative filtering, 2003.

Информация об авторах



Глибовец Николай Николаевич – декан факультета информатики, заведующий кафедрой информатики Национального университета «Киево-Могилянская академия», доктор физико-математических наук, профессор; e-mail: glib@ukma.kiev.ua

Поле научных интересов: искусственный интеллект, системы дистанционного обучения



Сидоренко Марина Олеговна – аспирантка 3-го курса аспирантуры Киевского-Национального университета имени Тараса Шевченка, факультет кибернетики; e-mail: tinuriel@gmail.com

Поле научных интересов: байесовские сети, искусственный интеллект, рекомендационные системы