Рекомендательная система с настраиваемым блоком анализа данных

Автор: Гарманов Владимир Владимирович Студент кафедры ЭВМ, ВШ ЭКН, ЮУрГУ

Научный руководитель: Надточий Ирина Львовна К.т.н., доцент

Актуальность темы

Recommended for you

The Silence of the Lambs

Learn more



★★★★★★★★★ 8.7/10

A young FBI cadet must confide in an incarcerated and manipulative killer to receive his help on catching another serial killer who skins his victims.

Director: Jonathan Demme Stars: Jodie Foster and Anthony Ho...

Recommended because of your interest in Snatch., Pulp Fiction and Confidence.















Customers Who Bought This Item Also Bought





Apple iPad MC705LL/A (16GB, Wi-Fi, Black) NEWEST MODEL

***** (360) \$509.95



3 Pack of Premium Crystal Clear Screen Protectors for Apple iPad

★★☆☆☆ (1,221) \$1.69

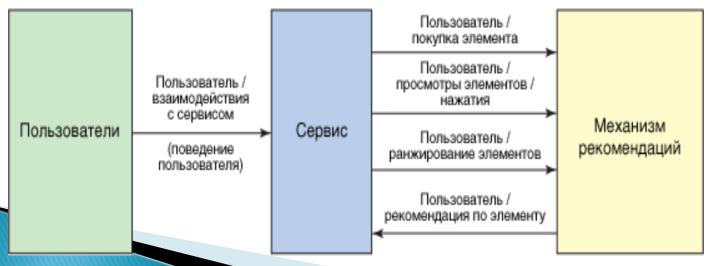


3 Pack of Universal Touch Screen Stylus Pen (Red + Black + Silver)

★★☆☆ (2,082) \$1.10

Введение в предметную область

▶ Рекомендательная система – это программа, имеющая свой логику работы, алгоритм, который осуществляет анализ поступающей на него информации пользователя и на её основании выводящий какой – либо продукт, который предположительно может быть интересен данному пользователю.



Цели и задачи

Цель:

- Разработка рекомендательной системы с настраиваемым блоком анализа данных.
 Задачи:
- Провести анализ существующих подходов к проектированию рекомендательных систем;
- Провести анализ существующих рекомендательных систем;
- Выбрать средства для разработки;
- Спроектировать и разработать основную конструкцию рекомендательной системы;
- Спроектировать и разработать тестовый алгоритм для спроектированной рекомендательной системы;
- Протестировать работоспособность и эффективность разработанного алгоритма.

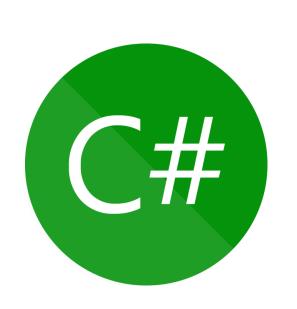
Основные подходы к реализации рекомендательных систем

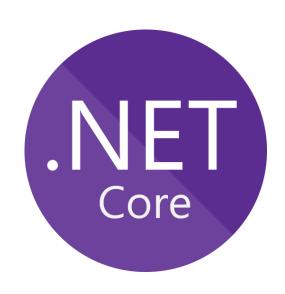
Коллаборативная фильтрация	Контентная фильтрация	Гибридный подход
+ высокая точность;	+ низкое потребление программных	+ высокая точность; + отсутствие
+ простота реализации.	ресурсов; + быстрота работы; + отсутствие «холодного старта».	«холодного старта»; + масштабируемость;
плохо поддаются масштабируемости;«холодный старт».	– низкая точность;	

Анализ аналогичных проектов

	retailrocket	REES46	Lenskit Python Tecls for Recommender Experiments	My _{Media Lite}
Поддержка модификации	+	-	+	+
Коллаборативная фильтрация	?	?	+	+
Контентная фильтрация	?	?	-	-
Смешанный тип	?	?	-	-
Открытость кода	гкрытость кода -		+	+
Стоимость	Есть бесплатный тариф	Есть триал- версия	Бесплатно	Бесплатно
Простота	+	+	-	+

Используемые технологии

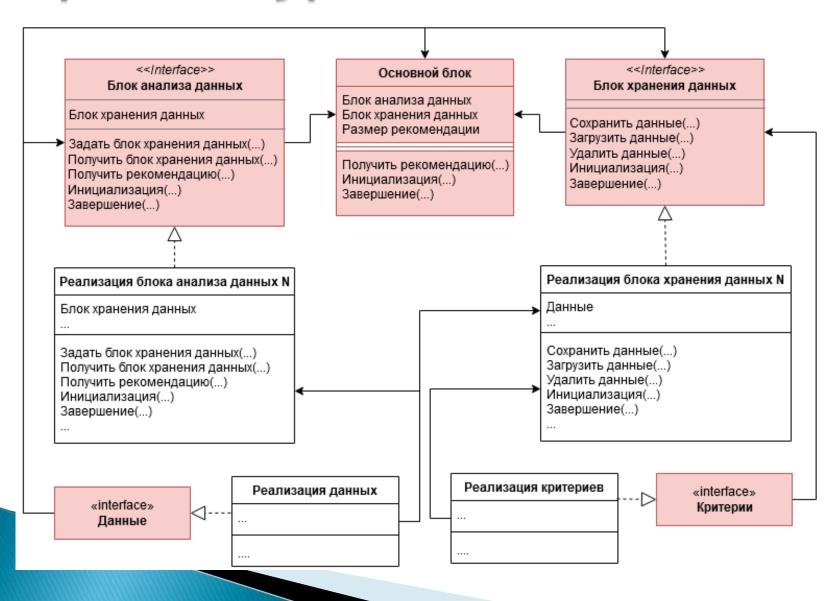








Архитектура системы



Основные классы

Название класса	Тип класс	Описание
RecCore	Класс	Является основным классом, с которым у пользователя происходит взаимодействие для получения рекомендации.
RecommendAlgorithm	Абстрактный класс	Содержит абстрактные методы, которые могут понадобиться для работы с анализом данных и выдачей результатов
DataContainer	Абстрактный класс	Содержит абстрактные методы, которые могут понадобиться для работы с хранилищами данных.
IModifiedData	Пустой интерфейс	Реализации интерфейса используются для установки на вход рекомендательной системы данных любого вида.
ICriteriaData	Пустой интерфейс	Реализации интерфейса используются для задания каких-либо пользовательских условий в некоторых методах.

Описание методов основных классов

Имя метода	Класс	Описание
GetRecommendation	RecCore	Возвращает пользователю рекомендацию на основе данных пользователя и ранее проанализированных данных.
End	RecCore	Выполняет методы End из реализаций абстрактных классов DataContainer и RecommendAlgorithm.
RecCore	RecCore	Конструктор. Принимает на вход реализации абстрактных классов DataContainer и RecommendAlgorithm и размер требуемой рекомендации.
SetDataContainer	RecommendAlgorithm	Устанавливает значение переменной dataContainer.
GetDataContainer	RecommendAlgorithm	Получает содержимое переменной dataContainer.

Имя метода	Класс	Описание
GetRecommendation	RecommendAlgorithm	Возвращает рекомендацию пользователю. Выполняет анализ данных.
Init	RecommendAlgorithm	Метод инициализации.
End	RecommendAlgorithm	Метод завершения работы. Выполняется через метод End класса RecCore.
SaveData	DataContainer	Выполняет сохранение данных в хранилище данных.
RemoveData	DataContainer	Удаляет данные из хранилища данных на основании каких-либо условий.
LoadData	DataContainer	Возвращает данные из хранилища данных на основании каких-либо условий.
Init	DataContainer	Метод инициализации.
End	DataContainer	Метод завершения работы. Выполняется через метод End класса RecCore.

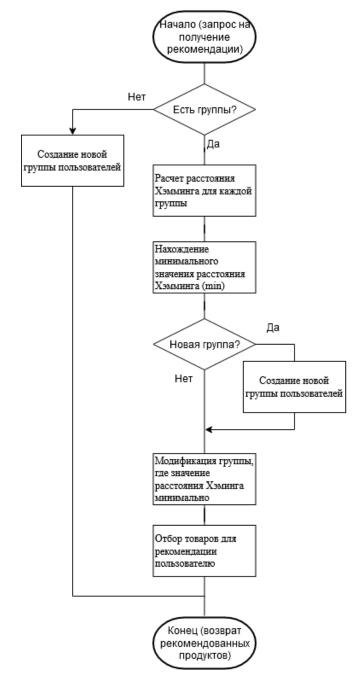
Алгоритм получения рекомендации

Расстояние Хэмминга:

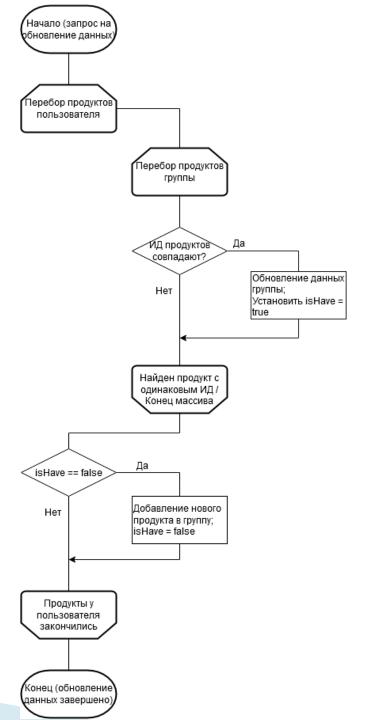
$$d_{ij} = \sum_{k=1}^{p} \left| x_{ik} - x_{jk} \right|;$$

Условие создание новой группы:

 $d_{min} > N_{\text{прод.пользователя}};$



Алгоритм модификации данных



Пример работы программы

```
C:\Users\vovag\source\repos\RecommendSystem\RecommendSystem\bin\Debug\RecommendSystem.exe
      20
      19
      20
      20
                96 | 118
      11
           316 | 483 | 871
344
      21
               23
                    28
      21
           23
                53
                     57
346
      21
           23
                53
                     57
      21
                53
           23
                     56
348
      19
           173
               429 | 434 |
349
      20
           12
                     96
350
      21
           53
                56
                     57
351
      21
           23
                50
                     53
352
      2
          202
                318 | 169
353
               12
                    69
354
      21
           23
                53
                     56
      18
           11
                32
                     50
           173 | 429 | 434 |
      18
           11
                32
                     50
          32
               86
                    134
359
      16
           5
               12
                    69
      11
           316 | 871 | 371
361
      21
           57
                72 | 116
      12
           79
                228 | 318
363
      20
           64
                96 | 122
      22
               8 | 11
      22
      15
           89 | 96 | 143
367
```

Класс управления базой данных

```
static class WorkDatabase
154
155
                 ссылок: 0
                 static public void ClearDB()...
156
                 static public void AddGroup(GroupInfo group)
164
165
                     using (var db = new LiteDatabase(@"GroupInfo.db"))
166
167
                         var col = db.GetCollection<GroupInfo>("groupInfo");
168
                         col.Insert(group);
169
170
171
172
                 static public void UpdateGroup(GroupInfo group)
173
174
                     using (var db = new LiteDatabase(@"GroupInfo.db"))
175
176
                         var col = db.GetCollection<GroupInfo>("groupInfo"):
177
                         col.Update(group);
178
179
180
                 ссылка: 1
                 static public List<GroupInfo> LoadGroupData()...
181
199
                 ссылок: 2
                 static public GroupInfo LoadGroupByID(int grID)...
200
218
```

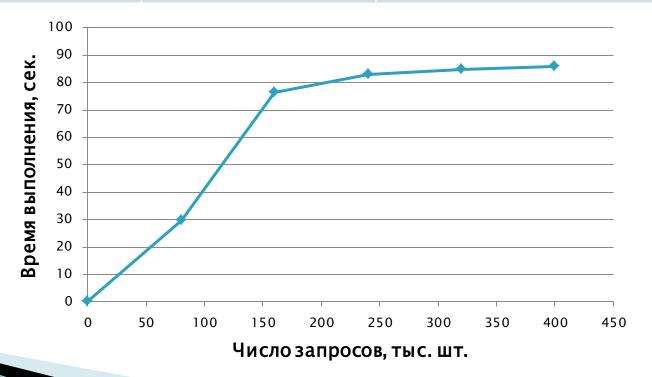
Класс Tools

Имя метода	Описание	
EqualCoef	Рассчитывает значимость группы пользователей на	
	конкретного пользователя, с помощью формулы расстояния	
	Хэмминга.	
RMSE	Рассчитывает значение RMSE (Root Mean Square Error) для	
	оценки эффективности алгоритма.	
GetUserInfo	Осуществляет парсинг тестового документа.	
ResultRating	Рассчитывает итоговый рейтинг продуктов в группах.	

```
static class Tools
 70
                 public static double EqualCoef(GroupInfo group, DataExample user)
 71
 72
 73
                     int n = 0;
                     double score = 0;
 74
                     for (int i = 0; i < user.prodID.Length; i++)</pre>
 75
 76
                         for (int j = 0; j < group.product.Length; j++)</pre>
                             if (group.product[j].Id == user.prodID[i])
 77
 78
                                 score += Math.Abs(user.score[i] - group.product[j].rating);
 79
 80
 81
                     score /= (double)n / user.score.Length;
 82
                     return score;
 84
 85
                 public static double RMSE(double[] yu, double[] yg)...
 86
                 public static double ResultRating(double score, int iterations)...
 99
                ссылка: 1
                public static List<DataExample> GetUserInfo(string path)...
100
150
```

Результаты тестирования скорости работы

№ файла	Общее время	Время на файл, сек.	Время на ед. товара, мс.
1	00:00:29.75	29,75	0,371875
2	00:01:46.07	76,32	0,954
3	00:03:09.00	82,93	1,036625
4	00:04:33.99	84,99	1,062375
5	00:05:59.75	85,76	1,072



Тестирование эффективности

№ теста	Значение RMSE
1	0.94327323
2	0.96181821
3	0.94569236
4	0.96917385
5	0.94019237

Рекомендации, выданные испытуемому:

(ИД – Продукт)

23 - Таксист (1976);

39 - Странные дни (1995);

79 - Беглец (1993).

Заключение

- В рамках данной работы была спроектирована и разработана рекомендательная система с настраиваемым блоком анализа данных.
- В процессе реализации были разработаны тестовый алгоритм и тестовое хранилище данных, которые показали относительно неплохие результаты на этапе тестирования.
- Для дальнейшего развития требуется провести отладку кода для повышения эффективности алгоритма.