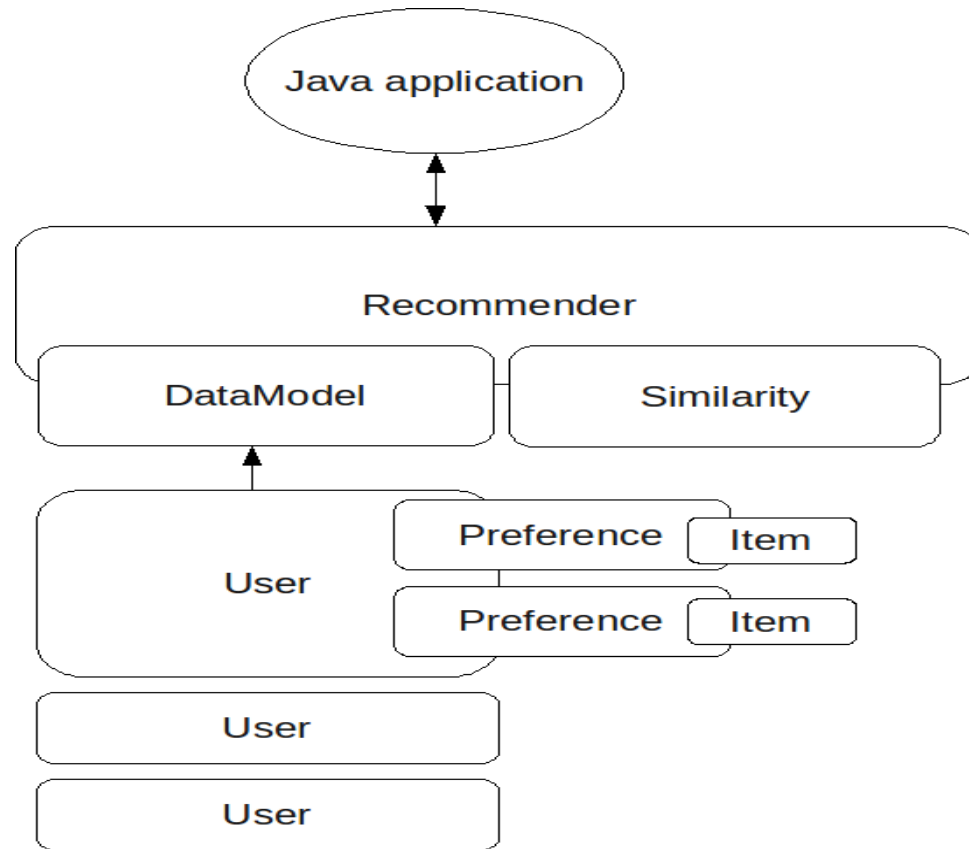


BTP PRESENTATION-PHASE III



Rahul Sihag (200901093)
Devender Goyal (200901072)

Tools Used

- ❖ Cygwin
- ❖ Java SDK 6u23 x64
- ❖ Eclipse 3.6(helios) SR1 x64
- ❖ Maven 3.0.2
- ❖ Hadoop 0.21.0

Working of Recommendation System

5	3	4	4	2	2	1	2	40
3	3	3	2	1	1	0	0	18.5
4	3	4	3	1	2	0	0	24.5
4	2	3	4	2	2	1	4	40
2	1	1	2	2	1	1	4.5	26
2	1	2	2	1	2	0	0	16.5
1	0	0	1	1	0	1	5	15.5

- S is the similarity matrix between items
- U is the user's preferences for items
- R is the predicted recommendations

Injecting Domain specific information

- Take advantage of what we know about the content i.e gender, wealth , North – South, graduation etc.

Example :

Create **custom similarity metric** for Gender. :

- Two Urban profiles assign similarity value 1.
- Two Rural values assign similarity value 1.
- Urban - Rural profile assign value -1.
- If Information not disclosed Assign value 0.
- Similarly create preferences for other parameters like urban – rural ,North- south , graduation etc

Injecting Domain specific information

- Problem : Mahout does not provide content based recommendation system.
- Solution : Mahout has various expansion points and APIs that allow us to do that.

Idrescorer : Allow us to modify recommendations . If a person is viewing all the profiles of college graduates then recommend him the profile of graduate persons only.

Similarly it is used to filter out mystery novels and books etc. Below a very simple implementation

```
public class GenreRescorer implements IDRescorer {  
    private final Genre currentGenre;  
  
    public GenreRescorer(Genre currentGenre) {  
        this.currentGenre = currentGenre;  
    }  
  
    public double rescore(long itemID, double originalScore) {  
        Book book = BookManager.lookupBook(itemID);  
        if (book.getGenre().equals(currentGenre)) {  
            return originalScore * 1.2;  
        }  
        return originalScore;  
    }  
}
```

Assume
BookManager
exists

Boost estimate
by 20%

Don't change
anything else

Libimseti dataset : 17,359,346 anonymous ratings of 168,791

profiles made by 135,359 LibimSeTi users

The screenshot displays the LibimSeTi website interface. At the top, there is a banner for ESET security software with buttons for 'Antivirus', 'Technická podpora', 'Antispam', 'Antispyware', and 'Firewall', along with an 'Objednat nyní' (Order now) button and the ESET logo. Below this, the website's logo 'libim.se ti.cz' is visible, followed by a search bar and a 'vyhledat' (Search) button. A navigation bar lists various sections: 'MOJE LÍBÍMSETI', 'HODNOCENÍ', 'SEZNAMKA', 'UŽIVATELÉ', 'FOTKY', 'LÍBÍMSETI LIFE', 'DISKUZE', 'CHAT', 'MISS HIGH SCHOOL', 'HRY', and 'DALŠÍ'. On the left side, there is a 'Přihlášení' (Login) section with a text input field, a '@libimseti.cz' label, a 'přihlásit' (Login) button, and a 'registrovat se' (Register) link. Below this, a 'Proč být na Libímseti' (Why be on Libimseti) section lists several benefits: 'nejjoblíbenější zábavní portál', 'vlastní profil, fotoalba', 'unikátní chat a e-mail', 'klubové akce, soutěže', and 'oblíbený server celebrit'. The main content area features a grid of user profiles, each with a name, age in parentheses, a profile picture, and a location. The profiles shown are: 'lucy jr (23.6)' from 'Královehradecký kraj', 'J.A.M.E.S.Xxx (19.7)' from 'Olomouc', 'konas92 (19.8)' from 'Olomouc', and 'XvendissX (21.0)' from 'Bruntál'. Below the profiles, there is a 'Nastav si svůj filtr' (Set your filter) section with a 'filtrovat uživatele' (Filter users) button and a search bar with dropdown menus for 'Ženy i muži', 'od 18', 'do 21', and 'region: - okres -', along with a 'vybrat' (Select) button. To the right of the profiles, there is a 'rozšířené vyhledávání' (Advanced search) section with a 'NORREZPRÁV.CZ' logo and several news snippets. On the far right, there is a vertical sidebar with buttons for 'Antivirus', 'Antispyware', 'Technická podpora', 'Firewall', and 'Antispam', along with an ESET logo and a 'SMART SECURITY 5' logo.

Antivirus
Technická podpora
Antispam
Antispyware
Firewall
Objednat nyní
eset

libim.se ti.cz
2356 online, přihlášeno 511 žen, 877 mužů

MOJE LÍBÍMSETI | HODNOCENÍ | SEZNAMKA | UŽIVATELÉ | FOTKY | LÍBÍMSETI LIFE | DISKUZE | CHAT | MISS HIGH SCHOOL | HRY | DALŠÍ ▾

Přihlášení
@libimseti.cz
přihlásit
registrovat se

Proč být na Libímseti

- nejjoblíbenější zábavní portál
- vlastní profil, fotoalba
- unikátní chat a e-mail
- klubové akce, soutěže
- oblíbený server celebrit

lucy jr (23.6)
Královehradecký kraj

J.A.M.E.S.Xxx (19.7)
Olomouc

konas92 (19.8)
Olomouc

XvendissX (21.0)
Bruntál

Nastav si svůj filtr

filtrovat uživatele
Ženy i muži: ▼ od 18 ▼ do 21 ▼ region: - okres - ▼ vybrat

rozšířené vyhledávání »

Chat

Města a místa
právě chatuje lidí 0

Pokec a zábava
právě chatuje lidí 49

Seznámení a flirt
právě chatuje lidí 10

NORREZPRÁV.CZ
Z matematiky by se mělo povinné maturovat, hlásí její milovník Klaus
Vězni si stěžují na malé cely, žalují stát
Drahá vejce? V knihkupectví či elektře je dostanete jako dárek
Viktor Kožený unikl vydání do USA
V Černobylu se začne stavět nový sarkofág nad havarovanou elektrárnou

ESSET SMART SECURITY 5

Average absolute difference between Estimated and actual values based on different similarity metrics and using the nearest-n-classifier.

Similarity	n = 1	n = 2	n = 4	n = 8	n = 16	n = 32	n = 64	n = 128
Euclidean	1.17	1.12	1.23	1.25	1.25	1.33	1.48	1.43
Pearson	1.3	1.19	1.27	1.3	1.26	1.35	1.38	1.47
Log-likelihood	1.33	1.38	1.33	1.35	1.33	1.29	1.33	1.49
Tanimoto	1.32	1.33	1.43	1.32	1.3	1.39	1.37	1.41

Average absolute difference between Estimated and actual values while evaluating the user based recommender based on different similarity metrics and using the threshold based recommender.

Similarity	t = 0.95	t = 0.85	t = 0.9	t = 0.8	t = 0.75	t = 0.7
Euclidean	1.33	1.37	1.39	1.43	1.41	1.47
Pearson	1.47	1.4	1.42	1.4	1.38	1.37
Log-likelihood	1.37	1.46	1.56	1.52	1.51	1.43
Tanimoto	NaN	NaN	NaN	NaN	NaN	NaN

NaN => Not a Number

Anonymous user problem

The problem of making recommendations to anonymous users, starting from no data is called the **Cold Start Problem**.

Approach 1 : Give a general predefined list of products to recommend. It is generally better than nothing and simple way of recommending.

Approach 2 : Promote these users to real users on their first visit and assign ID and track their activities based on a web session.

Problem : Exploded the number of user , users may not return (so computation going to waste)

Solution : Aggregate anonymous users, treat them as if they were a single user.

This simplifies the process of storing , based on assumption that all such users behave similarly.

The set of recommendations is stored and is computed periodically instead upon every request.

Custom Recommender

- ❖ Recommend more effectively
- ❖ Includes various preferences like food type, movies for entertainment compatibility, gender information, liking for pets, and the region from which the person belongs, the regions were divided into east, west, north and south.
- ❖ Takes into account both user based and item based recommendations.
- ❖ Different preferences were given different weights.

Custom Recommender

Sample Data Showing the location of different users

```
101,N  
102,E  
103,E  
104,S  
105,S  
106,W  
107,E
```

Below is shown the final output

```
For userNo - 1 ,RecommendedItem[item:106, value:6.0]  
For userNo - 2 ,RecommendedItem[item:103, value:9.67399]  
For userNo - 3 ,RecommendedItem[item:102, value:7.2378254]  
For userNo - 4 ,RecommendedItem[item:105, value:2.8]  
For userNo - 5 ,RecommendedItem[item:107, value:10.249556]
```

References

- Mahout's website, wiki and mailinglist
 - <http://mahout.apache.org>
 - user@mahout.apache.org
- B. Sarwar et al: „Itembased collaborative filtering recommendation algorithms“, 2001
- Manning's Early Access Program
<http://manning.com/owen>
- Apache mahout Documentation and Mahout wiki



Thank You