Hes·so
Haute Ecole Spécialisée
de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts
Western Switzerland

Master of Science HES-SO in Engineering

Product Recommender System improved with social network information

Simone Cogno

Professors: **Ghorbel Hatem, Punceva Magdalena**

HE-Arc group Analyse de données

Client: **HE-ARC group Analyse de données**

DESCRIPTION

Recommender systems are obtaining a lot of importance in e-commerce and a variety of other world wide web applications. Some of the most common are probably movies, books, music, news, documents and products in general.

During the last years, we also observe and huge increment of application that uses social network for improving the user experience. Some of them are also publicly available.

We can cite the famous Twitter website that allows retrieving a lot of social network data through their public API. There are also several datasets available online provides these data for research propose, for example, the Epinions.com dataset.

The most common techniques to create a Recommender System (RS) are the Content-Based (CB) and Collaborative Filtering (CF).
CF approach build a model from the past rating history of users as well as other similar decisions made by other users. This model is used for retrieve a list of items that the users my be interested in. A Content-based RS use the past user reviews to describe an user profile using several discrete characteristics. The items having similar properties are then recommended to the users.

The aim of this project is to analyze several techniques used for creating an RS including some new approaches that take advantages of the social network information for enhance the recommendation quality.

After this research, an implementation of the most interesting algorithm will be created and tested for demonstrating the correctness and the performance of the system.

OBJECTIVES

The main objective of this thesis is to create a general propose Recommender System. A state of the art of the most common technique will be established as well as a research of the public datasets available.

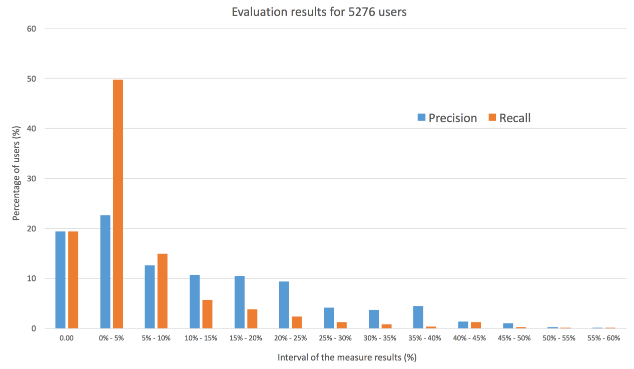
An algorithm has to be chosen for implementing a prototype of a Recommender Systems using the dataset founded. Ideally the algorithm has to be implemented with Hadoop or Spark to allowing us to analyses a big dataset by taking advantage of the computation power of several machines in parallel.

The algorithm implemented has to be tested with the several common metrics for providing an estimation of the quality of the Recommender System.

As a secondary objectives, we want also to create a stand-alone web application that demonstrate the recommendation to users. Furthermore we wish to adapt the algorithm for the RecSys challenge 2016 to retrieve more reliable and comparable results among the other Recommender Systems.

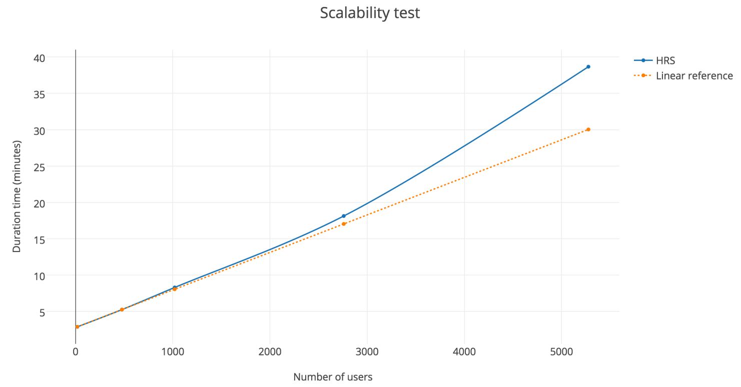
RESULTS

To evaluate the Hybrid algorithm, we used the Goodreads and the Epinions datasets. We took 75\% of the ratings made by users as a train set and we hide the remaining 25% for the evaluation phase. We expect that the items recommended by the hybrid algorithm are most likely to be the same as the items contained in the test set.
The graph showed in the figure below we can see in the X axis the percentage of users where their recommendation have a Precision and Recall in a particular interval.
For example, we can see that the 10.7% of users (fourth group of the column from left) have a Precision between 10 and 15% and only the 4.4\% have precision between 35-40%.
We calculated that the 45.4% of users have a precision greater than 10% and the 7% of users have a Precision bigger that 35%.



Percentage of users having a Precision and Recall in a given interval

We performed a scalability test for the Hybrid Recommender System by computing the recommendation to 477 up to 5276 users. We can see for the graph in the figure below that the performance of the algorithm is almost linear for up to 3000 users. We observe, then, an increment in the duration time when performing the recommendation for 5276 users. The difference between the measured time and the linear reference for 5276 users is around 8.6 minutes.



Scalability test of the Hybrid RS

CONCLUSION

The results of the Hybrid recommender System showed that for the 45% of the users the recommendation has a precision between 10% and 60% by using the Goodreads dataset. We observed, instead, more limited performance when tested using the Epinions.com dataset.

To improve the performance of the FWUM algorithm is recommended to retrieve more users in the Goodreads dataset. In fact, with a higher number of users, there is more probability of finding similar users. Another possibility is to cluster the similar items to reducing the sparsity of the utility matrix and decreasing the effect of the cold start problem.

By regarding the initial purpose, we can see that the main aims are widely achieved. A Recommender System has been implemented and tested. The little web application for demonstrating the work is also established. The last objective, the one concerning the RecSys challenge 2016, will be an excellent continuation of this thesis as this challenge begins just after the deadline for this project.

SWISS UNIVERSITIES
APPLIED SCIENCES
AND ARTS

MSE
MASTER OF SCIENCE
IN ENGINEERING