DEVELOPING AN APPLICATION RECOMMENDATION SYSTEM USING COLLABORATIVE MINING

Project in Data Warehousing and Data Mining (CSE408)

By

Mayank Jaggi (14 BCE 0848)
Akansha Maloo (14 BCE 0621)
Nakin Ahluwalia (14 BCE 0387)

Submitted to:

Prof. Hari Seetha
Associate Professor
Scope, VIT University

22nd March 2017

Introduction

In today's world, where the internet is growing faster than ever, and E-Commerce is expected to reach up to $4 trillion by 2020[1], the number of choices offered to a customer are overwhelming [2]. Hence, there arises a need to develop systems which help the customer to get products based on the customer's needs, interests and preferences. These systems that deal with the problem of information overload are called recommender systems[3] which work by filtering vital information fragments from data based on user's preferences, interest, or observed behaviour about one or more items [4]. Recommender systems predict whether a particular user would prefer an item or not based on the user's profile. One of the techniques used in building recommender systems is collaborative filtering technique. Collaborative filtering recommends items by identifying other users with similar taste; it uses their opinion to recommend items to the active user. This project aims at building a recommender system based on collaborative filtering technique to recommend smartphone applications to users based on what applications the user and other users have.

Literature survey

Collaborative filtering became one of the most researched techniques of recommender systems since this approach was mentioned and described by Paul Resnick and Hal Varian in 1997[5]. The idea of collaborative filtering is in finding users in a community that share appreciations [6]. If two users use same or almost same items in common, then they are likely to have similar tastes. Such users build a group or a so called neighbourhood. A user gets recommendations to those items that he/she isn't using, but that are being used by users in his/her neighbourhood.

Table 1 shows what all applications the three users from a given set of five applications. All three users have most applications that they use in common with each other. This means that they have similar taste and hence build a neighbourhood. The user A is not using 'WhatsApp', while other users in his neighbourhood are, hence, he will get this item recommended. As opposed to simpler recommender systems where recommendations are based on the most rated item or the most popular item methods, collaborative recommender systems care about the taste of user. The taste is considered to be constant or at least change slowly.

	Facebook	**twitter**	**Instagram**	**snapchat**	**WhatsApp**
User A	Yes	Yes	Yes	Yes	-
User B	Yes	Yes	Yes	Yes	Yes
User C	-	Yes	Yes	Yes	Yes

Table 1

Collaborative filtering is widely used in e-commerce. Customers can rate books, songs, movies, applications, books etc. and then get recommendations regarding those in future. Moreover collaborative filtering is utilized in browsing of certain documents (e.g. documents among scientific works, articles, and magazines). [7]

Proposed work/system

Collecting the data set

The data set was collected by conducted a survey through Google forms and asking people to check apps they use on their phone out of the following list of applications:

- Facebook
- Twitter
- LinkedIn
- Google+
- Pinterest
- Tumblr
- Reddit
- Flickr
- Whatsapp
- Hike
- Allo
- Facebook Messenger
- Hangouts
- Telegram
- Skype

Developing an algorithm

We tried to create an analogy in our project with market basket analysis. To see what items are bought together, transactions and items in that transactions are dealt with. We found our problem similar to market basket analysis wherein our problem, the transaction ids could be compared to users (or user ids) and items bought to applications being used.

Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. For example, if you are in an English pub and you buy a pint of beer and don't buy a bar meal, you are more likely to buy crisps (US. chips) at the same time than somebody who didn't buy beer.

Frequent itemsets are groups of items that often appear together in the data. The market-basket model of data is used to describe a common form of a many-to-many relationship between two kinds of objects. On the one hand, we have items, and on the other we have baskets, also called 'transactions'. The set of items is usually represented as set of attributes. Mostly these attributes are binominal. The transactions are usually each represented as examples of the DataSet. When an attribute value is 'true' in an example; it implies that the corresponding item is present in that transaction. Each transaction consists of a set of items (an itemset). The frequent-itemsets problem is that of finding sets of items that appear together in at least a threshold ratio of transactions.

The support of X with respect to T is defined as the proportion of transactions t in the database which contains itemset X

$$\mathrm{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$

As market basket analysis, mainly uses two algorithms – the apriori and FP growth, we decided to go with either one of them. As the apriori is long and time consuming, we thought of going with FP growth.

Tools

Various tools are available for data mining in the market. We decided to go with RaidMiner because of its GUI, Faster result processing and easy data entering.

The FP Growth operator is a RapidMiner core and it efficiently calculates all frequent itemsets from the given DataSet using the FP-tree data structure. It is compulsory that all attributes of the input DataSet should be binominal. Parameters in FP Growth operator as RapidMiner will find only those item sets which exceed this minimum support value.

$$\mathrm{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$

The dataset was cleaned, entered into the tool and results calculated.

Figure 1 shows how the market basket analysis looks in RapidMiner.

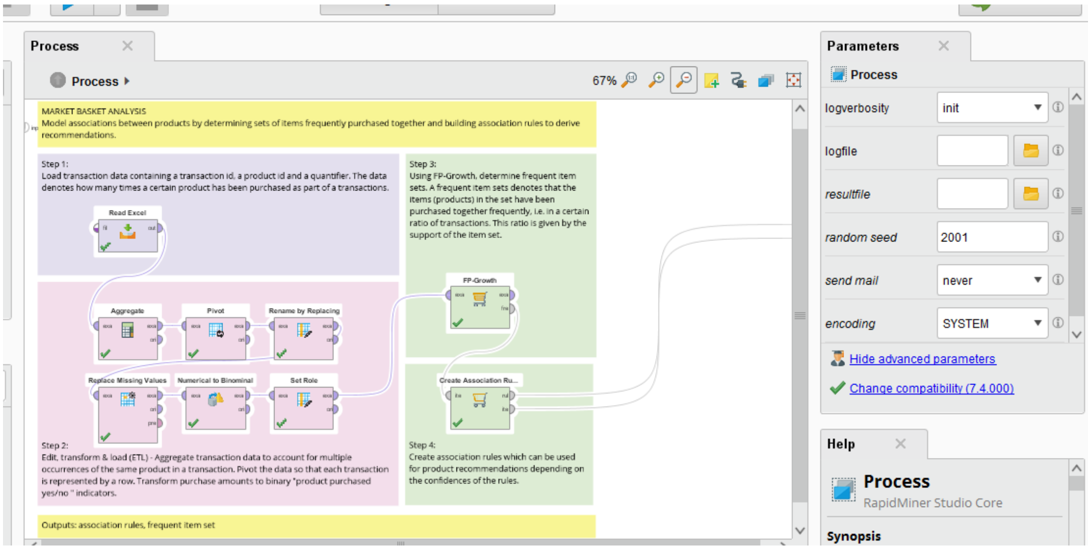

Figure 1: Layout of the Market Basket Analysis on RapidMiner

Results and Discussion

The following results were observed on passing the dataset in the data mining tool.

Support	App1	App2	App3	App4	App5	App6
0.898	Whatsapp	Facebook				
0.857	Whatsapp	Instagram				
0.653	Whatsapp	Messenger				
0.592	Whatsapp	Snapchat				
0.551	Whatsapp	LinkedIn				
0.796	Whatsapp	Facebook	Instagram			
0.653	Whatsapp	Facebook	Messenger			
0.571	Whatsapp	Facebook	Snapchat			
0.49	Whatsapp	Facebook	LinkedIn			
0.408	Whatsapp	Facebook	Skype			
0.347	Whatsapp	Facebook	GooglePlus			
0.633	Whatsapp	Instagram	Messenger			
0.571	Whatsapp	Instagram	Snapchat			
0.633	Whatsapp	Facebook	Instagram	Messenger		
0.551	Whatsapp	Facebook	Instagram	Snapchat		
0.469	Whatsapp	Facebook	Instagram	LinkedIn		
0.469	Whatsapp	Facebook	Instagram	Messenger	Snapchat	
0.388	Whatsapp	Facebook	Instagram	Messenger	LinkedIn	
0.347	Whatsapp	Facebook	Instagram	Messenger	Skype	
0.306	Whatsapp	Facebook	Instagram	Messenger	GooglePlus	
0.347	Whatsapp	Facebook	Instagram	Snapchat	LinkedIn	
0.306	Whatsapp	Facebook	Instagram	Messenger	Snapchat	LinkedIn
0.286	Whatsapp	Facebook	Instagram	Messenger	Snapchat	Skype
0.286	Whatsapp	Facebook	Instagram	Messenger	Snapchat	GooglePlus

Table2: ItemSets and their support value.

The confidence value of a rule, X=>Y, with respect to a set of transactions T, is the proportion of the transaction that contains X which also contain Y.

Confidence is defined as:

$$\mathrm{conf}(X \Rightarrow Y) = \mathrm{supp}(X \cup Y)/\mathrm{supp}(X) \ .$$

Premises	Conclusions	Support	Confidence
Facebook, Snapchat	Messenger	0.469387755	0.821428571
Instagram, Snapchat	Messenger	0.469387755	0.821428571
Facebook, Snapchat	Whatsapp, Messenger	0.469387755	0.821428571
Whatsapp, Facebook, Snapchat	Messenger	0.469387755	0.821428571
Instagram, Snapchat	Whatsapp, Messenger	0.469387755	0.821428571
Whatsapp, Instagram, Snapchat	Messenger	0.469387755	0.821428571
Facebook, Snapchat	Instagram, Messenger	0.469387755	0.821428571
Instagram, Snapchat	Facebook, Messenger	0.469387755	0.821428571
Facebook, Snapchat	Whatsapp, Instagram, Messenger	0.469387755	0.821428571
Whatsapp, Facebook, Snapchat	Instagram, Messenger	0.469387755	0.821428571
Instagram, Snapchat	Whatsapp, Facebook, Messenger	0.469387755	0.821428571
Whatsapp, Instagram, Snapchat	Facebook, Messenger	0.469387755	0.821428571
Facebook, Instagram, LinkedIn	Messenger	0.387755102	0.826086957
Facebook, Instagram, LinkedIn	Whatsapp, Messenger	0.387755102	0.826086957
Whatsapp, Facebook, Instagram, LinkedIn	Messenger	0.387755102	0.826086957
GooglePlus	Messenger	0.306122449	0.833333333
GooglePlus	Whatsapp, Messenger	0.306122449	0.833333333
Whatsapp, GooglePlus	Messenger	0.306122449	0.833333333
Facebook, LinkedIn	Messenger	0.408163265	0.833333333
GooglePlus	Facebook, Messenger	0.306122449	0.833333333
GooglePlus	Instagram, Messenger	0.306122449	0.833333333
Snapchat, LinkedIn	Messenger	0.306122449	0.833333333
Facebook, LinkedIn	Whatsapp, Messenger	0.408163265	0.833333333
Whatsapp, Facebook, LinkedIn	Messenger	0.408163265	0.833333333

GooglePlus	Whatsapp, Facebook, Messenger	0.306122449	0.833333333
Whatsapp, GooglePlus	Facebook, Messenger	0.306122449	0.833333333
GooglePlus	Whatsapp, Instagram, Messenger	0.306122449	0.833333333
Whatsapp, GooglePlus	Instagram, Messenger	0.306122449	0.833333333
Snapchat, LinkedIn	Whatsapp, Messenger	0.306122449	0.833333333
Whatsapp, Snapchat, LinkedIn	Messenger	0.306122449	0.833333333

Table3: Associations Rule- relation between unrelated apps and their confidence and support.

The results obtained from the data mining tool – RapidMiner can be used to observe what applications are used together. Tis data can be used to suggest to users what else applications they can install on their smart phones based on what applications they already have.

Conclusions and Scope for future work

Recommender systems are assisting users in the process of identifying items that fulfil their wishes and needs. These systems are successfully applied in different e-commerce settings, for example, to the recommendation of news, movies, music, books, applications etc.

Recommender systems are being used by an ever-increasing number of E-commerce sites to help consumers find products to purchase. What started as a novelty has turned into a serious business tool. Recommender systems use product knowledge – either hand-coded knowledge provided by experts or "mined" knowledge learned from the behaviour of consumers – to guide consumers through the often-overwhelming task of locating products they will like.

Still there remain many challenges yet to be tackled with respect to recommendation systems. Some of them are-

- Focusing on the User Perspective- There are many other scenarios quite similar to the above mentioned amazon.com one where the recommender system is clearly focused on increasing business revenues. An approach which is in the line of the idea of a stronger focus on the quality of user support is required.
- Sharing Recommendation Knowledge- Besides commercial interests, one of the major reasons for the low level of customer orientation of today's recommender solutions is the lack of the needed recommendation knowledge. In order to recommend books already read by friends the recommender would need the information of the social network of the customer. The global availability of CPG goods information seems to be theoretically possible but is definitely in the need of a corresponding cloud and mobile computing infrastructure.
- Unobtrusive Preference Identification- Knowledge about user preferences is a key preliminary for determining recommendations of relevance for the user. A major issue in this context is the development of new technologies which allow the elicitation of preferences in an unobtrusive fashion- The three major modalities to support such a

type of preference elicitation are the detection of facial expressions, the interpretation of recorded speech, and the analysis of physiological signals.

- Psychological Aspects of Recommender Systems- Building efficient recommendation algorithms and the corresponding user interfaces requires a deep understanding of human decision processes. This goal can be achieved by analysing existing psychological theories of human decision making and their impact on the construction of recommender systems.

References

[1] “Worldwide Retail Ecommerce Sales: The eMarketer Forecast for 2016”

[2] Recommendation systems: Principles, methods and evaluation. F.O. Isinkayea, Y.O. Folajimib, B.A. Ojokohc

[3] J.A. Konstan, J. Riedl Recommender systems: from algorithms to user experience User Model User-Adapt Interact, 22 (2012), pp. 101–123

[4] C. Pan, W. Li Research paper recommendation with topic analysis In Computer Design and Applications IEEE, 4 (2010) pp. V4-264

[5] P. Resnick and H. R. Varian, “Recommender systems,” Commun. ACM, vol. 40, pp. 56–58, March 1997.

[6] G. Linden, B. Smith, and J. York, “Amazon.com recommendations: Item-to-item collaborative filtering,” IEEE Internet Computing, vol. 7, pp. 76–80, January 2003. [Online]. Available: http://portal.acm.org/citation.cfm?id=642462.642471

[7] S. Jung, J. Kim, and J. L. Herlocker, “Applying collaborative filtering for efficient document search,” in Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence, ser. WI ’04. Washington, DC, USA: IEEE Computer Society, 2004, pp. 640–643

[8] Evaluating Collaborative Filtering Recommender Systems Jonathan L. Herlocker Oregon State University and Joseph A. Konstan, Loren G. Terveen, and John T. Riedl University of Minnesota

[9] Toward the Next Generation of Recommender Systems: Applications and Research Challenges Alexander Felfernig, Michael Jeran, Gerald Ninaus, Florian Reinfrank, and Stefan Reiterer

[10] Basic Approaches in Recommendation Systems Alexander Felfernig, Michael Jeran, Gerald Ninaus, Florian Reinfrank, Stefan Reiterer, and Martin Stettinger

[11] A Survey Paper on Recommender Systems. Dhoha Almazro, Ghadeer Shahatah, Lamia Albdulkarim

[12] Algorithms and Methods in Recommender Systems. Daniar Asanov, Berlin Institute of Technology

[13] Data Mining Methods for Recommender Systems. Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol

[14] Recommendation systems: Principles, methods and evaluation. F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh