AN EFFICACIOUS ANALYSIS ON USAGE AND RULE MINING ASSOCIATED WITH SOCIAL NETWORK PORTALS

Tamanna Jain

M.Tech. (CSE) Student

M. M. University, Mullana, Haryana

ABSTRACT

The flourishing prevalence of social networking sites among the Internet users especially teenagers need an introversion of personal and social behaviour.Social networking sites are basically a shift between the private and public information. Today 1.5 billion people across the world have their profiles in social networking sites.The success of Facebook and the rich opportunities offered by social media sites lead to the creation of new web based applications for social networks and open up new frontiers. Thus, discovering the usage patterns of social media sites might be useful in taking decisions about the design and implementation of those applications as well as educational tools. These sites are also one of the best sources on the Web for extracting the content on virtually any topic you can think of. The mining process involves many steps in fetching the appropriate information. In this manuscript, I am going to introduce the concept of mining process together with some approaching dissolution on the usage and rule mining techniques in the manuscript. The factors affecting "Facebook usage time" and "Facebook access frequency" are revealed. At the same time, I would investigate the associations of the students' opinions on the contribution of Facebook in an educational aspect by employing the association rules method.

Keywords: Facebook Usage, Rule Mining, Social Networking, Usage Mining, Web Mining,

1. INTRODUCTION

A social network is a polished, organized and a communal system of connections that keeps an individual in touch with other individual or a group of personage. The term *social* refers to the intercommunication of organisms with other creatures and to their cumulative co-existence. The term *network* refers to a mesh or an organization or any structure. Thus a social networking site or service expedites the framework of social relationships among people sharing interests, activities, backgrounds or real-life connections. This system of connections consists of representation of each

user, his/her links and an assortment accessory of services. Most social network services are web-based and afford platformsfor users to keep in touch over the Internet, such as e-mail and instant messaging. Online community services are sometimes considered as a social network service, though in a widespread sense, social network service usually means an individual-centred service whereas online community services are group-centred. Social networking sites grants users to share ideas, activities, events, and interests within their individual networks.

Thus, it is an effortless deed of embellishing the number of people you know by meeting your friends' friends, their friends' friends and so on. In fact, many of us today use Twitter and Facebook to advertise our extant and imminent businesses. And people looking to connect with other business-associated contacts usually move to sites like LinkedIn, but one need to understand that social media is beyond Twitter, Facebook, LinkedIn and Blogs. After observing and running an analysis on hundreds of Social Networking sites I have listed down 20 most popular social networks across countries.

S.No.	Site	URL	Description
1.	Facebook	https://www.facebook.com/	Facebook is a social utility that connects people with friends and others who work, study and live around them.
2.	MySpace	http://www.myspace.com	**Myspace** is a social networking service with a strong music emphasis owned by Specific Media LLC and pop music singer and actor Justin Timberlake.
3.	Twitter	https://www.twitter.com	Twitter is a very simple service that is rapidly becoming one of the most talked-about social networking service providers.
4.	LinkedIn	http://www.linkedin.com/	LinkedIn is an online social network for business professionals, which is designed specifically for professional networking, to help them find a job, discover sales leads, connect with potential business partners.
5.	Bebo	http://www.bebo.com/	In the United Kingdom, Bebo is the

			second best social network
6.	Friendster	http://www.friendster.com/	Friendster was one of the first Web sites to bring it into mass culture.
7.	Hi5	http://www.hi5.com/	Hi5 shares many similarities with many social network sites; however, it introduces some twists that make it worthwhile for people who love trying out new and interesting online communities.
8.	Habbo	https://www.habbo.com/	The Habbo online community is inhabited by pixelated, cartoon-character alter egos. You can meet others in public rooms (nightclubs, restaurants, shops) and create private rooms for selected friends.
9.	NING	http://www.ning.com/	Ning is the leading online platform for the world's organizers, activists and influencers to create social experiences that inspire action.
10.	Classmates	http://www.classmates.com/	Classmates.com is different from most social networks, in the sense that most of its features are available to premium member.
11.	Tagged	http://www.tagged.com/	Tagged is a blend of social networking features that MySpace and Facebook users will find very familiar.
12.	myYearbook	http://www.meetme.com/	myYearbook, the best place to meet new people and one of the 25 most-trafficked sites in the United States.
13.	Meetup	http://www.meetup.com/	Meetup is an online social networking portal that facilitates offline group meetings in various localities around the world.
14.	MyLife	http://www.mylife.com/	MyLife (formerly Reunion.com) is a social network service.

15.	Flixster	http://ww5.fixter.com/	Flixster is a social networking site for movie fans.
16.	myHeritage	http://www.myheritage.com/	MyHeritage is a family-oriented social network service and genealogy website.
17.	Multiply	http://www.multiply.com/	Multiply is a vibrant social shopping destination, but faster and more convenient, where sellers and buyers interact.
18.	Orkut	http://www.orkut.com/	Orkut is a free social networking website where you can create a profile, connect with friends, maintain an online scrapbook and use site features and applications to share your interests and meet others.
19.	Badoo	http://badoo.com/	Badoo is a multi-lingual social networking website. It is gaining popularity in emerging markets like Russia and Brazil.
20.	Gaia Online	http://www.gaiaonline.com/	Gaia Online is a mix of social networking and massive multiplayer online role-playing games.

The research methodology of social network analysis is developed to understand the relationship between "actors", and the term actor can be a person, an organization, an event or an object. In a social network, each actor is presented as a node and each pair of nodes can be connected by lines to show the relationships. The social network structure graph is a graph that formed by those lines and nodes, and social network analysis is therefore a methodology that used to understand the graph and the relationships and actors in the social network. There are three important elements that included in a social network: actors, ties, and relationships. Actors are the essential elements in the social network to define the people, events or objects. Ties are used to construct the relationship between actors by using a mean of path to establish the relationship directly or indirectly. Ties can also be divided into strong and weak tie according to the strength of the relationships; they are also useful for discovering subgroups of the social network. Relationships are used to illustrate the interactions and relationships between two actors. Furthermore, different relationships may cause the network to reflect different characteristics.

2. USAGE MINING

Web usage mining, an application of the web mining process, is the process of fetching the useful information from server logs, i.e, user's history. It differs from collaborative filtering in the fact that we are not interested in explicitly discovering user profiles but rather usage profiles. When preprocessing a log file we do not concentrate on efficient identification of unique users but rather try to identify separate user sessions. These sessions are then used to form the transactions. Web usage mining techniques are applied to identify frequent item-sets, sequential patterns, clusters of related pages and association rules. It can be used to support dynamic structural changes of a web site in order to suit the active user, and to make recommendations to the active user that help him in further navigation through the site he is currently visiting. In the ease of implementing web usage mining system in the form of a proxy server, predictions about which pages are likely to be visited in near future can be made, based on the active users' behaviour. Such pages can be pre-fetched to reduce access times.

2.1 Web Log Data Preparation

The data from web logs, in its raw form, is not suitable for the application of usage mining algorithms. The data needs to be cleaned and pre-processed. The overall data preparation process is described as:

1) *Data Cleaning:* implies to remove the accesses to irrelevant items, accesses by web crawlers and failed requests.

2) *Efficient User Identification*: the user's IP address is poor identification information. Many users can be assigned the same IP address and on the other hand one user can have several different IP addresses even in the same session. The first problem lies with the side- effect of intermediary proxy devices and local network gateways. The second problem occurs when the ISP is performing load balancing over several proxies. All this prevents us from easily identifying and tracking the user. Another means of good user identification is assigning users username and passwords.

3) *Session identification and Path Completion:*session identification is carried out using the assumption that if a certain predefined period of time between two accesses is exceeded, a new session starts at that point. Sessions can have some missing parts. This is due to the browser's own caching mechanism and also because of the intermediate proxy-caches. The missing parts can be inferred from the site's structure.

4) *Transaction Identification:* Some authors propose dividing or joining the sessions into meaningful clusters, i.e., transactions. Pages visited within a session can be categorized as

auxiliary or content pages. Auxiliary pages are used for navigation. Content pages provide some useful contents to the user.

3. RULE MINING

In data mining, association rules are useful for analysing and predicting customer behaviour. They play an important part in shopping basket data analysis, product clustering, catalogue design and store layout. Programmers use association rules to build programs capable of machine learning. Machine learning is a type of artificial intelligence (AI) that seeks to build programs with the ability to become more efficient without being explicitly programmed. **Association rule learning** is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness.

3.1 APRIORI ALGORITHM

Apriori is the best-known algorithm to mine association rules. It was proposed by Agrawal and Srikant in1994. The algorithm finds the frequent set L in thedatabase D. It makes use of the downward closure property. The algorithm is a bottom search, moving upward level; it prunes many of the sets which are unlikely to be frequent sets, thus saving any extra efforts. It is used for finding hidden patterns in the data.

It is based on a simple observation based on a real life application: if very few people go to Baskin Robin and McDonald's on the same day, then there can't be a lot of people going to Baskin Robin, McDonald's, and Taco Bell on the same day. So if you want to find combinations of three stores that lots of people go to on the same day, you don't have to look at combinations that include two stores that very few people go to on the same day. This tremendously reduces the number of combinations you need to look at. As for where it is used best, in proves of concept or toy applications.

Apriori algorithm is an algorithm of association rule mining. It is an important data mining model studied extensively by the database and data mining community. It Assume all data are categorical. It is initially used for Market Basket Analysis to find how items purchased by customers are related.

The problem of finding association rules can be stated as :

Given a database of sales transactions, it is desirable to discover the important associations among different items such the presence of some items in a transaction will imply the presence of other items in the same transaction. As example of an association rule is:

Contains (T, "milk") $\rightarrow$ Contains (T, "bread") [Support= 4%, Confidence=40%]

The interpretation of such rule is as follows:

☐ 40% of transactions that contains milk also contains bread;

☐ 4% of all transactions contain both of these items.

The calculations of the Support(S) and

Confidence(C) are very simple:

- CONF (A $\rightarrow$ B) = SUPP(AUB)
- SUPP(A)
- S (A) = (Number of transactions containing item A)/(Total number of transactions in the database)
- S (A $\rightarrow$ B) =(Number of transactions containing items A and B) / (Total number of transactions in the database)

The above association rule is called singledimension because it involves a single attribute or predicate (Contains). The main problem is to find all association rules that satisfy minimum support and minimum confidence thresholds, which are provided by user and/or domain experts. A rule is frequent if its support is greater than the minimum support threshold and strong if its confidence is more than the minimum confidence threshold. Discovering all association rules is considered as two phase process where we find all frequent item sets having minimum support. The search space to enumeration all frequent item sets is on the magnitude of 2 * n. In second step, we generate strong rules. Any association that satisfies the threshold will be used to generate an association rule. The first phase in discovering all association rules is considered to be the most important one because it is time consuming due tothe huge search space (the power set of the set of all items) and the second phase can be accomplished in a straightforward manner.

4. LIMITATION OF RULE MINING ALGORITHM

4.1 Limitation of Association Rule Mining

- End users of ARM encounter the problem as the algorithm do not return result in a reasonable time.
- It only tells the presence and absence of an item in transactional database.
- It is not efficient in case of large dataset.
- It treats all items in database equally by considering only the presence and absence of an item within the transaction. It does not take into account the significance of item to user or business.
- It fails to associate user objective and business value with outcome of ARM analysis.

Association Rule Mining has a lot of disadvantages. These can be removed by using attributes like weight and quantity, weight attribute will give user an estimate of how much quantity of item has been purchased by the customer, profit attribute will calculate the profit ratio and tell total amount of profit an item is given to the customer.

4.2 Limitation of Apriori Algorithm

- Needs several iterations of data.
- Uses a uniform minimum support threshold.
- Difficulties to find rarely occurring events.
- Alternative methods can address this by using a non-uniform minimum support threshold.
- Some competing alternative approaches focus on partition and sampling.

5. CONCLUSION

In the real world entity social networking sites is the most flourishing means of communicating and fetching the relevant information. Social sites are thus a vast collection of the knowledgeable information as well as the information of the thousands and lakhs of people who have theirprofies on these sites. Physical analysis of these huge amount of information stored in modern databases is very difficult. Data mining provides tools to reveal unknown information in large databases which are stored already. Awell-known data mining technique is association rule mining in which it provides an effective algorithm is apriori algorithm. The large quantity of information collected through the set of association rules using algorithm can be used not only for illustrating the relationships in the database, but also used for differentiating between different kinds of classes in a database. This paper provides some of the analysis of apriori algorithm.

REFERENCES

1. Romero, C., Ventura, S.: Educational Data Mining: A survey from 1995 to 2005. Expert Systems with Applications 33, 135–146 (2007)

2. Liu, J., Wang, Z., Xiao, X.: A Hybrid SVM/DDBHMM Decision Fusion Modeling for Robust Continuous Digital Speech Recognition. Pattern Recognition Letters 28, 912–920(2007)

3. Fuller, C.M., Piros, D.P., Wilson, R.L.: Decision Support for Determining Veracity via Linguistic-Based Cues. Decision Support Systems 46, 697–703 (2009)

4. Saunders, S.: The Role of Social Networking Sites in Teacher Education Programs: A Qualitative Exploration. In: McFerrin, K., et al. (eds.) Proceedings of Society for Information Technology and Teacher Education International Conference, pp. 2223–2228. AACE, Chesapeake (2008)

5. Mazman, S.G., Usluel, Y.K.: Adoption Process of Social Network and Their Usage in Educational Context. Master Thesis. The Institute for Graduate Studies in Science and Engineering. Hacettepe University, Ankara (2009)