ISSN:0975-9646

Venkata Ramana Adari et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (6) , 2011, 2857-2861

FaceMovie - A Novel Approach to Post the Mining Classification Data

Venkata Ramana Adari [1] , Dr.Ch.Diwakar [2], Dr.P.Suresh Varma [3]

[1]Professor, *Department of Computer Science & Engineering, Chaitanya Engineering College, Visakhapatnam,AP,India,*
[2] Professor, *Department of Computer Science & Engineering, Visakha Institute of Engineering &Technology,Visakhapatnam,AP,India*
[3] Professor, *Department of Computer Science,Aadikavi Nannayya \ University,Rajahmundry,AP,India,*

Abstract— Social Network sites make it possible to search large amounts of data for characteristic rules and patterns. If applied to monitoring data recorded on a real time data or Business in a network, they can be used to post in the network site database. In this paper, we present "Supervised learning" , a method to cascade the decision tree learning methods to classify into either family oriented, comedy, romantic, horror activities in a social network site. we can be used to build any one of the decision tree such as (ID3, C4.5,CART), here the decision tree on each dataset refines the decision boundaries by learning the subgroups within the database. Our work studies the best algorithm by using classifying movies oriented activities in Social network database with supervised algorithms that have not been used before. We analyse the algorithm that have the best efficiency or the best learning and describes the proposed system of ID3 Decision Tree.

Keywords – Supervised Learning, Social networks, Decision Tree, Facebook, Database.

I. INTRODUCTION

Data mining techniques are used to find essential information from Web documents or Web services (Etzioni, 1996). A successful utilization of the Web data requires the exploit of data mining technologies, giving a rise to the area of Web mining. Realistically, Web mining is the application of data mining in the Web environment and can help find useful patterns and rules of users' behaviors. Since data mining technologies are being applied for a variety of analytical purposes in Web environment, Web mining[1] could be further classified into three major sub-areas: Web content mining, Web structure mining and Web usage mining. Web content mining attempts to discover useful information from Web contents. For example, the classification of web pages is a typical application of content mining techniques (Shen, Cong, Sun, and Lu, 2003). Web structure mining studies the Web linkage structure. Finally, Web usage mining focuses on the Web surfer's sessions and behaviors. Facebook users are almost unthinkable. Since its inception in 2004, this popular social network service has quickly become both a basic tool for and a mirror of social interaction, personal identity, and network building among employers, students, Social network sites deeply penetrate their users' everyday life and as pervasive technology, tend to become invisible once they are widely adopted, ubiquitous, and taken for granted. Pervasive technology often leads to unintended consequences, such as threats to privacy and changes in the relationship between public and private sphere. These issues have been studied with respect to a variety of Internet contexts and applications. Specific privacy concerns of online social networking include inadvertent disclosure of [5], personal information, damaged reputation due to rumors and gossip, unwanted contact and harassment or stalking, surveillance like structures due to backtracking functions, use of personal data by third-parties, and hacking and identity theft. Coupled with a rise in privacy concerns is the call to increase our understanding of the attitudes and behaviors towards ''privacy-affecting systems". Social networks are explicitly exhibit the relationships among individuals and groups called actors. Originally in a social science since the 1930, to date [2] a vast number of studies of social network analysis have been conducted. In the context of the Semantic Web, social networks are useful for trust calculation [3], information sharing and recommendation [1] ontology construction relations and relevance detection, e.g. COI detection [1], and so on.

Social Network Sites (SNSs) such as Orkut, Google+, and MySpace allow individuals to present themselves, articulate their social networks and establish or maintain connections with others. These sites can be oriented towards work-related contexts (e.g., LinkedIn.com), romantic relationship initiation (the original goal of Friendster.com), connecting those with shared interests such as music or politics (e.g., MySpace.com), or the college student population (the original incarnation of Facebook.com). Participants may use the sites to interact with people they already know offline or to meet new people. The online social network application analyzed Facebook, enables its users to present themselves in an online profile, accumulate ''friends'' who can post comments on each other's pages, and view each [4] other's profiles. Facebook members can also join virtual groups based on common interests, see what classes they have in common, and learn each others' hobbies, interests, musical tastes, and romantic relationship status through the profiles.

Facebook constitutes a rich site for interested in the affordances of social networks due to its heavy usage patterns and technological capacities that bridge online and offline connections. Believe [6] that Facebook represents an understudied offline to online trend in that it originally primarily served a geographically-bound community. When data were collected for this study, membership was restricted to people with a specific host institution email address, further tying offline networks to online membership. In this sense, the original incarnation of Facebook was similar to the wired Toronto neighborhood studied by Hampton and Wellman who suggest that information technology may enhance place-based community and facilitate the generation of social capital. Previous research suggests that Facebook users engage in ‘‘searching’’ for people with whom they have an offline connection more than they ‘‘browse’’ for complete strangers to meet.

II. SECTION

2.1. Survey on Uses of Face Book: Facebook was reported to have more than 21 million registered users generating billion page views each day by 2007. The site integrated into media daily practices typical user spends about 20 minutes a day on the site and other users log in at least once a day Capitalizing on its success among college students and employees, Facebook launched a high school version in past September 2005. In 2008, the company introduced communities for commercial organizations; as of November 2008, almost 32,000 organizations had Facebook [8] directories (Smith, 2008). In 2008, Facebook was used at over 4,000 United States colleges and was the seventh most popular site on the World Wide Web with respect to total page views . Most of the existing academic research on Facebook has focused on identity presentation and privacy concerns looking at the amount of information Facebook participants provide about themselves, the relatively open nature of the information, and the lack of privacy controls enacted by the users, Gross and Acquisti (2005) argue that users may be putting themselves at risk both offline (e.g., stalking) and online (e.g., identify theft). In contrast to popular press coverage which has primarily focused on negative outcomes of Facebook use stemming from users' misconceptions about the nature of their online audience, we are interested in situations in which the intended audience for the profile such as well-meaning peers and friends and the actual audience are aligned. We use Facebook as a research context in order to determine whether offline social capital can be generated by online tools.

2.2. Related Issues on social networks:

First recognizable social network site launched in 1997 introducing the Orkut.com users to create profiles, list their Friends and, beginning in 1998, surf the Friends lists. Each of these features existed in some form before orkut, of course. Profiles existed on most major dating sites and many community sites. AIM and ICQ buddy lists supported lists of Friends, although those Friends were not visible to others. Classmates.com allowed people to affiliate with their high school or college and [9] surf the network for others who were also affiliated, but users could not create profiles or list of Friends until years later. Orkut was the first to combine these features.

Orkut promoted itself as a tool to help people connect with and send messages to others. While Orkut attracted millions of users, it failed to become a sustainable business and, in 2000, the service closed. Looking back, its founder believes that Orkut was simply ahead of its time, personal communication by July 2007. While people were already flocking to the Internet, most did not have extended networks of friends who were online. Early adopters complained that there was little to do after accepting Friend requests, and most users were not interested in meeting strangers. From 1997 to 2001, a number of community tools began supporting various combinations of profiles and publicly articulated Friends. Many users allowed to create personal, professional, and business profiles—users could identify Friends on their personal profiles without seeking approval for those connections. Likewise, shortly [9] after its launch in. In particular, the people behind Ryze, Tribe.net, LinkedIn, and Friendster were tightly entwined personally, business and professionally. They believed that they could support each other without competing. In the end, Ryze never acquired mass popularity, Tribe.net grew to attract a passionate niche user base, LinkedIn became a powerful business service, and Friendster became the most significant, if only as "one of the biggest disappointments in Internet history"

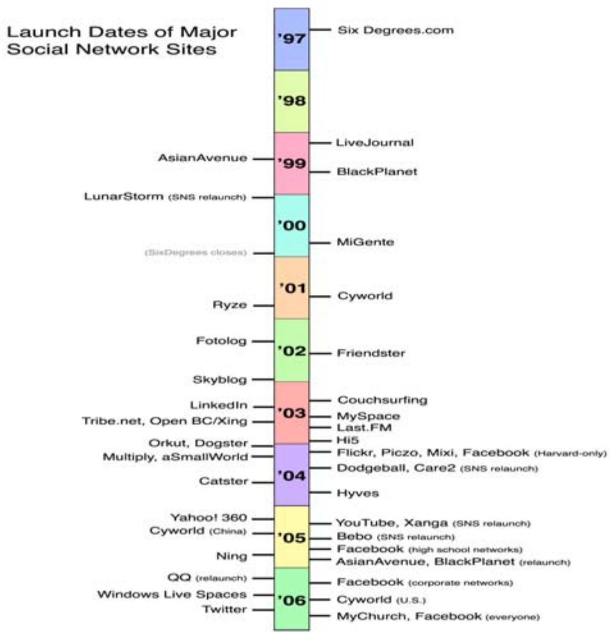

Figure 1 shows the timeline of the social network sites when it was created

In the Friendster, MySpace, and Facebook, three key SNSs that shaped the business, cultural, and research landscape. Don't talk to invisible strangers. New York Times. Retrieved

III. SECTION

3. Problem Definition: People populate a lot more information on social network sites as their profile data and etc. Now, almost every Internet user has an account in social network sites and they spend a significant part of time in browsing in social networks. Our research is an attempt to take the above facts as benefits and provides correct information to users to reach their needs at right time. One way to collect information from people is making survey. But, its difficult to make a survey for millions of people and it has many constraints. Most of the time people will not provide right information. But it mostly true that they upload right information in their profiles. We will make use the profile data and other information to achieve the objective of business needs.

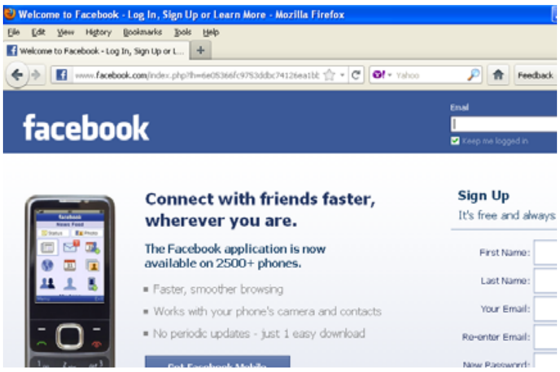

Figure 2 Example screen of facebook

Our proposed solution analysis aims to capture the user interest in facebook and post the data into categories such as Horror, Romantic, Comedy, Family oriented movies. To implement this solution we consider that a data mining technique supervised learning.

- Reviews the user interest from facebook
- Collecting the movie related data from news paper, internet etc.
- We use the classification technique to classify into different contexts
- Uploading our classification results in facebook database

This work not only uploads the movie related data but it also concentrates on news updates which increase the growth of business needs.

3.1. Why we need to use facebook: social networks like facebook, Orkut, Google+ increases the growth of day to day and business applications.

1) Facebook helps us to find new friends and also helps to find our old friends. I have heard stories of people who found their old classmates with the help of Facebook.

2) Accessibility to chosen universities having a high level of security: Facebook provides a safe environment for university related data transferring.

3) It is easy and secure to use – Facebook has a clear and a simple template or theme which makes the user feel comfortable in using facebook.

4) It helps you to share your idea with your friends – You can update your status, share pictures, videos, music and more.

5) Facebook helps in eliminating the effect of distance between you and your friends.

6) It helps in business promotion through Ads and Fans page. Web masters create fans page at facebook and people who like their website can become a fan of their page and get updates from their site. This helps web masters in promoting their sites.

7) Facebook entertains people through Applications and Games. Some people comes to facebook only to play games. Funny right they have lot to know about facebook!

8)You can discuss with your friends about your classes, if you didn't attend them. So, you need not ring your friends up to ask what happened in class when you were on leave.

9) Even some people use facebook as a dating system.

10) People, mainly students of age 15-20 use to gather as a group and starts to chat. Using Facebook's group chat feature, you can chat in group without meeting your friends. You can just create a group and add our friends who use to join your during chat, and then you can just chat in groups.

3.2. Mining the data using Supervised Learning: To post the results of classification details in the facebook database we consider to apply the data mining classification techniques that are Association rule mining and Classification

3.2.1. Classification: A classification task begins with build data (also known as training data) for which the target values (or class assignments) are known. Different classification algorithms use different techniques for finding relations between the predictor attributes' values and the target attribute's values in the build data.

In Classification, training examples are used to learn a model that can classify the data samples into known classes. The Classification process involves following steps:

a. Create training data set.
b. Identify class attribute and classes.
c. Identify useful attributes for classification (relevance analysis).
d. Learn a model using training examples in training set.
e. Use the model to classify the unknown data samples.

3.2.2. Association Rule: Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases .

Piatetsky-Shapiro[11] describes analyzing and presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawa[12]l et al. introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-scale (POS) systems in supermarkets. For example, the rule{ onions, potatoes}=>{beef} found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy beef.

Association model is often used for market basket analysis, which attempts to discover relationships or correlations in a set of items. Market basket analysis is widely used in data analysis for direct marketing, catalog design, and other business decision-making processes. Traditionally, association models are used to discover business trends by analyzing customer transactions. However, they can also be used effectively to predict Web page accesses for personalization. For example, assume that after mining the Web access log, Company X discovered an association rule "A and B implies C," with 80% confidence, where A, B, and C are Web page accesses. If a user has visited pages A and B, there is an 80% chance that he/she will visit page C in the same session. Page C may or may not have a direct link from A or B. This information can be used to create a dynamic link to page C from pages A or B so that the user can "click-through" to page C directly. This kind of information is particularly valuable for a Web server supporting an e-commerce site to link the different product pages dynamically, based on the customer interaction.

IV. SECTION

4. Usage of Mining Algorithms to post the data in facebook:

Machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to learn based on data, such as from sensor data or databases. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. Hence, machine learning is closely related to fields such as statistics, probability theory, data mining, pattern recognition, artificial intelligence, adaptive control, and theoretical computer science.

Implementation of classification real datasets, we consider the classification mining Decision tree

4.1.Decision Tree: Decision tree rules provide model transparency so that a business user, marketing analyst, or business analyst can understand the basis of the model's predictions, and therefore, be comfortable acting on them and explaining them to others Decision Tree does not support nested tables. Decision Tree Models can be converted to XML. Several algorithms in Decision trees are mentioned below.

4.2.ID3 Decision Tree Iterative Dichotomiser is an algorithm to generate a decision tree invented by Ross Quinlan, based on Occam's razor. It prefers smaller decision trees(simpler theories) over larger ones. However it does not always produce smallest tree, and therefore heuristic. The decision tree is used by the concept of Information Entropy. The ID3 Algorithm steps are:

1)Take all unused attributes and count their entropy concerning test samples

2) Choose attribute for which entropy is maximum

3)Make node containing that attribute

ID3 (Examples, Target _ Attribute, Attributes) Create a root node for the tree

If all examples are positive, Return the single-node tree Root, with label = +.

If all examples are negative, Return the single-node tree Root, with label = -.

If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.

Otherwise Begin

A = The Attribute that best classifies examples.

Decision Tree attribute for Root = A.

For each possible value, v_i, of A,

Add a new tree branch below Root, corresponding to the test A = v_i.

Let Examples(v_i), be the subset of examples that have the value v_i for A

If Examples(v_i) is empty common target value in the examples

Else below this new branch add the sub tree ID3 (Examples(v_i), Target_ Attribute, Attributes – {A}

End

Return Root

4.3. CART*:* It is a (Classification and regression trees) was introduced by Breiman, (1984). It builds both classifications and regressions trees. The classification tree construction by CART is based on binary splitting of the attributes. It is also based on Hunt's model of decision tree construction and can be implemented serially (Breiman, 1984). It uses gini index splitting measure in selecting the splitting attribute. Pruning is done in CART by using a portion of the training data set (Podgorelec et al, 2002). CART uses both numeric and categorical attributes for building the decision tree and has in-built features that deal with missing attributes (Lewis, 200). CART is unique from other Hunt's based algorithm as it is also use for regression analysis with the help of the regression trees. The regression analysis feature is used in forecasting a dependent variable (result) given a set of predictor variables over a given period of time. It uses many single variable splitting criteria like gini index, symgini etc and one multi-variable (linear combinations) in determining the best split point and data is sorted at every node to determine the best splitting point. The linear combination splitting criteria is used during regression analysis. SALFORD SYSTEMS implemented a version of CART called CART® using the original code of Breiman, (1984). CART® has enhanced features and capabilities that address the short-comings of CART giving rise to a modern decision tree classifier with high classification and prediction accuracy.

V. CONCLUSION

In this paper a general supervised for classifying instance in large and complex movie datasets and an explanation mechanism to explain which movie belongs to either horror, romantic, comedy, family oriented results was described. The specific approaches of the implementation of movie datasets posting system learning are characterized, we developed the ID3 decision tree method is based on cascading machine learning techniques the ID3 decision trees. The ID3 decision tree build on each dataset learns the sub classifies within the data and partitions the decision space into classification regions; there by improving the system classification performance.

Our future direction is to utilize dependency measure and classification results to post the data in social network database for other purpose such as cascading the classifiers developed using different Back propagation, fuzzy set related data, decision trees like C4.5 and classification and Regression trees(CART)which increases the growth of business application. Motivated by issues of verification from false positives. To end this we designed and implemented a novel explanation mechanism for the problem of having user interest can also be resolved by one such approach that uses ID3 is to achieve a high rate of accuracy in human-understandable can also improve the efficiency when analyzing the complex dataset.

REFERENCES

[1]. S. Ghita, W. Nejdl, and R. Paiu. Semantically rich recommendations in social
networks for sharing, exchanging and ranking semantic context. In Proc. ISWC05, 2005.

[2]. J. Golbeck and B. Parsia. Trust network-based filtering of aggregated claims. International Journal of Metadata, Semantics and Ontologies, 2006.

[3]. P. Massa and P. Avesani. Controversial users demand local trust metrics: an experimental study on epinions.com community. In *Proc. AAAI-05*, 2005.

[4] Y. Matsuo, M. Hamasaki, H. Takeda, J. Mori, D. Bollegala, Y. Nakamura
T. Nishimura, K. Hasida, and M. Ishizuka. Spinning multiple social networks for semantic web. In *Proc. AAAI-06*, 2006.

[5] . P. Mika. Ontologies are us: A unified model of social networks and semantics. In Proc. ISWC2005, 2005.

[6] . J. Mori, M. Ishizuka, T. Sugiyama, and Y. Matsuo. Real-world oriented information sharing using social networks. In Proc. ACM GROUP'05, 2005.

[7] July 21, 2007 from http://www.nytimes.com/2006/03/09/fashion/thursdaystyles/09parents.html

[8] Barnes, S. (2006). A privacy paradox: Social networking in the United States. First Monday, 11 (9). Retrieved September 8, 2007 from http://www.firstmonday.org/issues/issue11_9/barnes/index.htm

[9] Romero, C., Ventura, S. Educational Data Mining: a Survey from 1995 to 2005. Expert Systems with Applications, 2007, 33(1), pp.135-146.

[10] Yudelson, M.V., Medvedeva, O., Legowski, E., Castine, M., Jukic, D., Rebecca, C. Mining Student Learning Data to Develop High Level Pedagogic Strategy in a Medical ITS. AAAI Workshop on Educational Data Mining, 2006. pp.1-8.

[11]Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, in G. Piatetsky-Shapiro & W. J. Frawley, eds, 'Knowledge Discovery in Databases', AAAI/MIT Press, Cambridge, MA.

[12] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Databases. Santiago, Chile, 487-499.