International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

Available at http://inpressco.com/category/ijcet

Research Article

Intelligent System using the concept of Group Profiling by hybrid User Profiling

Kumar Ashish[Å*], Umesh Chandra Jaiswal[Å] and Pankaj Upadhyay[Å]

[Å]Department of Computer Science & Engineering Madan Mohan Malviya University of Technology Gorakhpur(UP), India.

Accepted 10 Sept 2014, Available online 01 Oct 2014, **Vol.4, No.5 (Oct 2014)**

Abstract

Many researchers have been done in this field. We propose a system which uses aggregation-based group profiling methods to form group profile using the users' profile by hybrid method and applying this concept to group profile and gaining the benefits of group profile. The performance of the system is calculated by the user satisfaction level. The system can applied to existing system to improve the performance for providing personalized information. The results shows gain in satisfaction level by comparing with recommendation given by hybrid user profile.

Keywords: *Aggregation, text mining, user profiling, group profiling.*

1. Introduction

Enormous data in the form of text over the world. Several system have been made using different technique to full fill the need for personalized system. We mainly focus on the system using collaborative method and clustered technique

A profile is a description of someone containing the most important or interesting facts about him or her (Josna Jojo *et. al.*, 2013). In the context of users of software applications, a user profile or user model contains essential information about an individual user. The motivation of building user profiles is that users differ in their preferences, interests, background and goals when using software applications. Discovering these differences is vital to providing users with personalized services. The content of a user profile varies from one application domain to another. For example, if we consider an online newspaper domain, the user profile contains the types of news (topics) the user likes to read, the types of news (topics) the user does not like to read, the newspapers he usually reads, and the user's reading habits and patterns. In a calendar management domain the user profile contains information about the dates and times when the user usually schedules each type of activity in which he is involved, the priorities each activity feature has for the user, the relevance of each user contact and the user's scheduling and rescheduling habits. In other domains personal information about the user, such as name, age, job, and hobbies might be important.

Data in the form of text has been stored in the database which is structured and non-structured. Structured data may include like phone number, name etc. Non-structured data may include address, e-mail id where same kind of data is not included. The searching in this kind of database is a challenging task where data type is not same. Text mining processes are applied to ensure the result being searched is as per the need.

2. Related Work

Intelligent user profiling implies the application of intelligent techniques, coming from the areas of Machine Learning, Data Mining or Information Retrieval, for example, to build user profiles. The data these techniques use to automatically build user profiles are obtained mainly from the observation of a user's actions.

2.1 Bayesian Networks

A Bayesian network is a compact, expressive representation of uncertain relationships among variables of interest in a domain. A Bayesian network is a directed acyclic graph where nodes represent random variables and arcs represent probabilistic correlations between variables (Jensen, 2001). The absence of edges in a Bayesian network denotes statements of independence. A Bayesian network also represents a particular probability distribution, the joint distribution over all the variables represented by nodes in the graph. This distribution is specified by a set of conditional probability tables. Each node has an associated conditional probability tables that specifies the probability of each possible state of the node given each possible combination of states of its parents. For nodes without parents, probabilities are not conditioned on other nodes; these are called the prior or marginal probabilities of these variables.

The mathematical model underlying Bayesian network is Bayes' theorem, which is shown in Equation 1. Bayes' theorem relates conditional and marginal probabilities. It yields the conditional probability distribution of a random variable A, assuming we know: information about another

*Corresponding author: **Kumar Ashish**

variable B in terms of the conditional probability distribution of B given A, and the marginal probability distribution of A alone. Equation 1 reads: the probability of A given B equals the probability of B given A times the probability of A, divided by the probability of B.

$$P(A/B) = (P(B/A) P(A)) / P(B) \quad (1)$$

2.2 Association Rules

Association rules are a data mining technique widely used to discover patterns from data. An association rule is a rule which implies certain association relationships among a set of objects in a given domain, such as they occur together or one implies the other. Association rule mining is commonly stated as follows (by Agrawal and Srikant, 1994):

Let *I* be a set of items and *D* be a set of transactions, each consisting of a subset *X* of items in *I*. An association rule is an implication of the form X→Y, where $X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. *X* is the antecedent of the rule and *Y* is the consequent. The rule has support *s* in *D* if *s*% of the transactions in *D* contains $X \cap Y$. The rule X→Y holds in *D* with confidence *c* if *c*% of transactions in *D* that contain *X* also contain *Y*. Given a transaction database *D*, the problem of mining association rules is to find all association rules that satisfy: minimum support (called *minsup*) and minimum confidence (called *minconf*).

2.3 Collaborative filtering

Collaborative filtering is the process of filtering or information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. collaborative filtering is a method of making automatic predictions (filtering) about the interests of a user by collecting preferences or taste information from many users (collaborating). The underlying assumption of the collaborative filtering approach is that if a person *A* has the same opinion as a person *B* on an issue, A is more likely to have B's opinion on a different issue *x* than to have the opinion on x of a person chosen randomly.

2.4 Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

In centroid-based clustering, clusters are represented by a central vector, which may not necessarily be a member of the data set. When the number of clusters is fixed to k, *k*-means clustering gives a formal definition as an optimization problem: find the K cluster center and assign the objects to nearest cluster center, such that the squared distances from the cluster are minimized.

2.4 Hybrid Approach

A "hybrid" approach, such as in (Josna Jojo *et. al.*, 2013) – authors mention three categories: content-based, only the content of an item is relevant to any recommendation; collaborative, users' interaction is what builds recommendations; hybrid, the recommendations are mixed from the analysis of items and users' actions – because the content of items being recommended has the same relevance as the historical preferences shown by users.

3. Proposed Work

Personalized system is a conceptual model that describe and specifies user background knowledge. From our daily life observation we found that Web user have different expectation from their search result.

As in the paper (Josna Jojo *et. al.*, 2013) author explained how the level of satisfaction increases using the concept of Hybrid approach. The author user concept of classification and clustering for generating a hybrid user profile. We are introducing concept of Group profile which further increases the satisfaction level can achieve. Group profiles are generated using the user profiles and methods of generating group profiles. Aggregation-based group profiling method is used for generating group profile.

Group profiling aims to find features that are shared by the whole group, a natural and straightforward approach is to find attributes that are most likely to occur within the group.

Aggregation-based group profiling (AGP):

1. Take n node in a network G.
2. d be the attributes in set $\{A_1, A_2, \ldots, A_d\}$
3. tp: number of positive instances containing feature A.
4. tn: number of negative instances not containing feature A.
5. fp: number of negative instances containing feature A.
6. fn: number of positive instances not containing feature A.
7. tpr: conditional probability of a feature occurring in a group.
$$tpr = P(A|+) = \frac{tp}{tp + fn}$$
8. fpr: conditional probability that a feature associated with the nodes that are not of the group.
$$fpr = P(A|-) = \frac{fp}{fp + tn}$$
9. AGP finds features shared by whole group as:
$$\max_{\{A_i\}_{i=1}^{k}} \sum_{i=1}^{k} P(A_i|+)$$
10. Aggregates all the attributes and pick *top-k* most-frequent feature to form a group profile.

4. Prerequisites and Motivation

There exist two main approaches for providing recommendations to a group of users when the "real" group profile is not available. The first combines individual recommendations to generate a list of group recommendations (L. Ardissono *et. al.*,2003), while the second computes group recommendations using a group profile derived from individual profiles (e.g. (J. F. McCarthy *et. al.*,2001)). In the last decade, several

strategies allowing the aggregation of individual user preferences for building a group profile have been proposed (J. Masthoff *et. al.*, 2006).

4.1 Group Profiling Strategies

Several strategies allowing the aggregation of individual user preferences for building a group profile have been proposed (J. Masthoff *et. al.*, 2006). We classified them into three categories (C. Bernier *et. al.*, 2010): majority-based, consensus-based, and borderline strategies.

The majority-based strategies use the most popular items (or item categories) among group members. For example, with the Plurality Voting strategy, each member votes for his preferred item (or item category) and the one with the highest votes is selected.

The consensus-based strategies consider the preferences of all group members. Examples include the Utilitarian strategy which averages the preferences of all the group members, the Fairness strategy, or the Alternated Satisfaction strategy.

The borderline strategies consider only a subset of items (item categories) in individual profiles, based on user roles or any other relevant criteria. For example, the Dictatorship strategy uses the preferences of only one member, who imposes his tastes to the group.

4.2 The Profiling Approach

The user profile is represented by a set of <concept, value> pairs, where each value is taken from the interval [0,1] and reflects the level of interest in the given concept (item category). More generally, the profiling engine manipulates three important types of information:

− *Quantity of Affiliation (QoA)* characterizes the degree of affiliation of a content item to a given concept. Each content item is characterized by a set of QoA.

− *Quantity of Consumption (QoC)* characterizes the degree of intensity of a consumption act with respect to a given concept. For example, the larger part of a movie is viewed by the user, the higher is his interest in the respective concepts.

− *Quantity of Interest (QoI)* characterizes the degree of interest of the user in a given concept. The user profile is composed of a set of QoI.

The profiling algorithm consists of first estimating the QoC values for each user consumption trace, and then updating iteratively the QoI values.

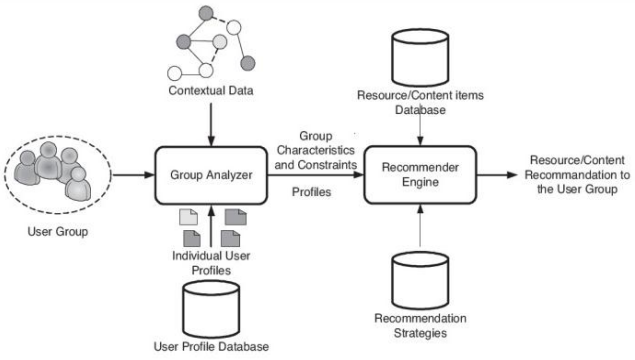

Figure1.0 Group Selection using users' group

5. Results and Analysis

The system which we proposed gives on average 17% more level of satisfaction then the system uses only user profiling concept for providing personalized search results to the user.

The following graph is generated using the vote by the user which shows the difference between the system using hybrid approach of user profiling and the system using those profiles and generating group profiles.

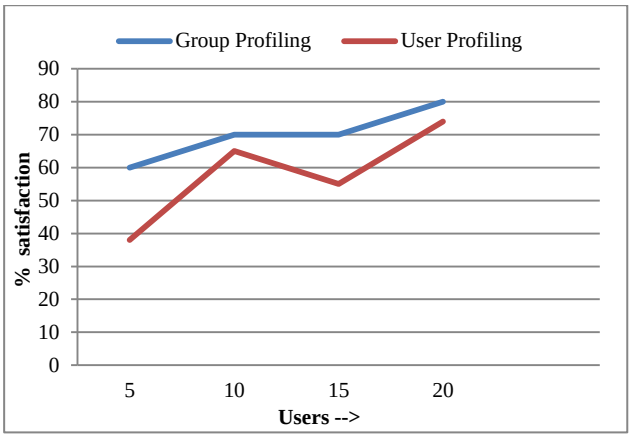

Figure 2.0 Graph showing comparison between group profiling system over user profiling system.

The x-axis indicates the number of user who voted equal i.e. 5 user voted 60% as there level of satisfaction. The y-axis indicates percentage level of satisfaction. The users' votes are taken according to their level of satisfaction in the manner in which they get the personalized search results. This system indicates how the level of satisfaction gets increased while applying the concept of Group profiling.

Conclusion and Future work

The result in the graph shows that the concept of Group profiling is better than the user profiling and it provides better level of satisfaction. The author in (Josna Jojo *et. al.*, 2013) proposed a hybrid system of user profiling in which he user concept of clustering and collaboration methods for creating a user profile. They show that the time required for searching the results get lesser as compared to normal search. Our system uses such concept in generating the group profile and with the use of that the level of satisfaction gets increased.

The future work may include the other group profiling strategies to follow and find out the desired results which can be much more personalized.

References

Josna Jojo, Dr.N.Sugana (2013). *User Profile Creation Based On Navigation Pattern for Modeling User Behaviour With Personalised Search* ICCTET'13, page no.371-374. IEEE-32107.

Christophe Senot, Dimitre Kostadinov *et. al.* (2011). *Evaluation of Group Profiling Strategies* International Joint Conference on Artificial Intelligence, page no. 2728-2733.

C. Bernier, A. Brun, *et. al.* (2010). *Topology of communities for the collaborative recommendations to groups*. 3rd International Conference on Information Systems and Economic Intelligence (SIIE'2010).

Silvia Schiaffino, Analía Amandi, (2009). *Intelligent User Profiling* M. Bramer (Ed.): Artificial Intelligence, LNAI-5640, page no. 193 – 216, 2009. Springer-Verlag Berlin Heidelberg.

L. Ardissono, A. Goy, *et. al.* (2003). *INTRIGUE: Personalized Recommendation of Tourist Attractions for Desktop and Handset Devices*, Applied Artificial Intellingence: Special Issue on Artificial Intelligence for Cultural Heritage and Digital, Libraries 17 (8-9), page no. 687-714.

J. Masthoff and A.Gatt, (2006). *In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems*, User Modeling and User- Adapted Interaction, Vol. 16, No. 3-4, page no. 281-319, September.

J. F. McCarthy, T. J. Costa, E. S. Liongosari,(2001). *UniCast, OutCast & GroupCast: Three Steps Toward Ubiquitous, Peripheral Displays*, UBICOMP.

D. Godoy and A. Amandi, (2005). *User Profiling in Personal Information Agents: A Survey*, Knowledge Eng. Rev., vol. 20, no. 4, page no. 329-361.