Scientific Research

A User Selection Method in Advertising System

Shiyi XIONG, Zhiqing LIN, Bo XIAO

Pattern Recognition & Intelligent System Lab,Beijing University of Posts and Telecommunications, Beijing, China
Email: xiongshiyi@gmail.com, {linzq, xiaobo}@bupt.edu.cn

Abstract

It's important for mobile operators to recommend new services. Traditional method is sending advertising messages to all mobile users. But most of users who are not interested in these services treat the messages as Spam. This paper presents a method to find potential customers who are likely to accept the services. This method searches the maximum frequent itemsets which indicate potential customers' features from a large data set of users' information, then find potential customers from those maximum frequent itemsets by using a bayesian network classifier. Experimental results demonstrate this method can select users with higher accuracy.

Keywords: User Selection, Maximum Frequent Itemsets, Bayesian Network

1. Introduction

Recent years, as the increasing number of mobile operators and new services, the competition between operators becomes more and more furious. The key of the competition is to win more customers. How to introduce new services to proper customers becomes one of the major problems.

Now, advertising system mostly uses these technologies: collaborative filtering technology [1], content based recommendation, knowledge based recommendation, efficiency based recommendation, association rules based recommendation [2], etc. These methods are usually used to select and recommend products to specified users. They analyse the user's historical data and predict what products the user mostly need. For example, an online movie system uses one of above technologies to serve its user whose name is Mike. The system will analyse Mike's history and recommend him some movies in which he may be interested. So we can say these methods solve the problem to select products for users.

The proplem we desire to solve in this paper is a little different. Here we have only one product, namely the new mobile service, and we want to know which users want to buy this service. Those users are potential customers, and we need to pick out them from a large number of users. The problem is to select users for one product. As a result, those traditional methods will not work well on this problem.

This paper presents a new method to select users. This method searchs the maximum frequent itemsets from a large data set, then find potential customers from those maximum frequent itemsets by using a bayesian network classifier.

2. Analysis of Customer Features

Usually, mobile operators keep a large database of consumption data. This data includes the items that represent customers' features, such as the telephone fees of every month, function fees and information fees, etc. This data also shows whether a customer accepts the new service, those customers who accept new service can be seen as positive samples, and others can be seen as negative samples. As a result, the problem to select positive samples can be treated as a two-category classification problem.

Usually, two-category classification problem should have data with features shown in Figure 1(a). However, the mobile users' data are quite different (shown in Figure 1(b). For example, some users who have the same features may belong to different categories. What's more, number of negative samples is much larger than the number of positive sample. Because of this, routine classification methods such as naive bayesian classifier is hard to classfy such data.

3. Frequently Itemsets

3.1. Fomal Model of Frequently Itemsets

Let sample X has n dimensions(Attributes) and the value may be continuous. Divide the value into several seg-
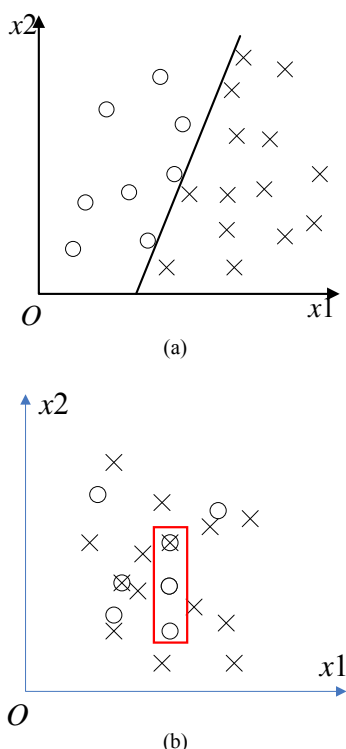
Figure 1. Feature distribution map. (a) Two-category classification; (b) Mobile users' data.

ments and give each segment a unique serial code. Because each sample belongs to an segment, we can use the corresponding segment code to replace this sample. In this way, original samples X are mapped to the segment codes $Y : F : X \rightarrow Y$.

Divide the original samples uses the following guidelines:

1) The number of positive samples in each segment is uniformly distributed.

2) Each Attributes should have the same number of segments.

Make each Attribute divided into k segments so that the segments code should not bigger than k. If the value of a Attributes is a 0-1 type value, the number of segments code should be $\{0,1\}$, not $\{0,1,...,k\text{-}1\}$.

The Attribute segment i_j is called item. The Attribute segment combination of $m(m < n)$ attributes $\{i_1, i_2, ..., i_m\}$ is called itemsets. m is length of the itemsets I. If a sample's attribute segments in m dimensions are $i_1, i_2, ..., i_m$, we call this sample satisfied the itemsets I.

If an itemsets has positive sample number a while has negative sample number b, $f_i = a/(a+b)$ is the success rate of itemsets I. Given a minimum support threshold α. If $a > \alpha$, we call I frequent itemsets.

3.2. Frequent Close Itemset

Given $I = \{i_1, i_2, ..., i_m\}$ be an itemsets with length m, add an item of attribute $m+1$ into I, so $I' = \{i_1, i_2, ..., i_m, i_{m+1}\}$ has the length $m+1$. If I and I' have the same number of positive samples, then $f_I < f_{I'}$. This can be proofed as below:

Suppose both I and I' have a positive samples. I has b negative samples, while I' has b'. Since I' has one more item i_{m+1}, we can get $b \geq b'$. Thus $\frac{a}{a+b} \leq \frac{a}{a+b'}$, that is $f_I \leq f_{I'}$, and I' has a higher success ratio.

In case of $f_I \leq f_{I'}$, ignore the itemsets I, and keep I'. When $f_I > f_{I'}$, I is called frequent close itemsets(ab. FCI). There are some typical algorithm to mine frequent close itemsets such as Apriori [3] and FP2Tree [4].

As a part of the user selection method, samples in the FCI should be predicted using Bayesian network classifier. Given the number of attributes n and the number of segments k, then the number of itemsets is:

$$\sum_{i=1}^{N} C_N^i \bullet M^i = (M+1)^N - 1$$

The number has an exponential growth, so it's impossible to construct a Bayesian network classifier for all itemsets. A compromised way is colleting all of the samples to be one data set and construct a Bayesian network classifier based on this data set.

4. Bayesian Network Model for User Selection

4.1. Bayesian Network Basics

A Bayesian network is a graphical model that encodes relationships among variables of interest. A Bayesian network consists of a set of nodes and a set of directed edges between nodes [5], which shown in Figure 2. In general, a Bayesian network is expressed as sign $B(G, P)$, which consists of the following two parts [6].
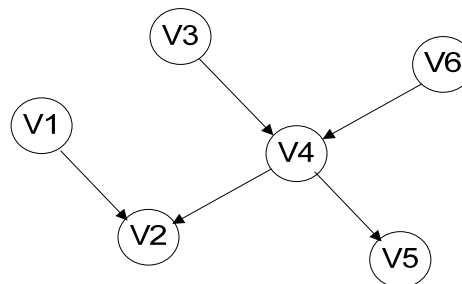

Figure 2. Typical graphic model of Bayesian networks.

1) A directed acyclic graph G with n nodes. The nodes of the graph represent random variables or events. The directed edges between nodes in the graph represent the causal relationships of the nodes. The important concept in Bayesian networks is the conditional independence between variables. For any variable V_i, the parent variables of V_i is $pa(V_i)$, the V_i is independent of the variables set $A(V_i)$, which is the set of the variables that are not child variables of $pa(V_i)$, so the probability of V_i is calculated as:

$$p(V_i \mid A(V_i), pa(V_i)) = p(V_i \mid pa(V_i))$$

2) A conditional probabilities table (ab. CPT) associated with each node P. The CPT is expressed as $P(V_i \mid pa(V_i))$ which pictures the mutual relationship of each node and its parent nodes. Nodes with no parent have a very simple probability table, giving the prior probability distribution of the node. Nodes with parents are much more complicated. These nodes have conditional probability tables, which give a probability distribution for every combination of states of the variable's parents.

The Bayesian network can represent all of the nodes' joint probability due to the node relationship and the conditional probability table. Applying the conditional independence into the chain rule, we get the following expression:

$$p(V_1, V_2, ... V_n) = \prod_{i=1}^{n} p(V_i \mid pa(V_i))$$

4.2. Bayesian Network Structure Learning

Consider a finite set $U = \{V_1, V_2, ... V_n\}$ of discrete random variables where each variable V_i may take on values from a finite set, denoted by $Val(V_i)$. Formally, a Bayesian network for U is a pair $B =< G, \Theta >$, and defines a unique joint probability distribution over U. The problem of learning a Bayesian network can be informally stated as: Given a training set $D = \{u_1, u_2, ... u_n\}$ of instances of U, find a network B that best matches D. The common approach to this pro- blem is to introduce a scoring function that evaluates each network with respect to the training data, and then to search for the best network according to this function. The two main scoring functions commonly used to learn Bayesian networks are the Bayesian scoring function [8], and the function based on the principle of minimal description length (ab. MDL) [7]. We only introduce MDL scoring function. The MDL scoring function of a network B given a training data set D, written $MDL(B \mid D)$, is given by following expresion:

$$MDL(B \mid D) = \frac{\log N}{2} \mid B \mid -LL(B \mid D)$$

where $\mid B \mid$ is the number of parameters in the network. The second term is the negation of the log likelihood of B given D:

$$LL(B \mid D) = \sum_{i=1}^{N} \log(P_{B(u_i)})$$

Bayesian network structure learning is NP problem, now typical method has K2 [8] developed by Cooper and Hers2kovits.

5. Model of Selection Method

The method searchs the maximum frequent itemsets from a large data set, then find potential customers from those maximum frequent itemsets by using a bayesian network classifier. In detail, we can get result by taking following steps:

1) All users share the same attributes. For each attribute, users have different values from each other. We get all sample's values of each attribute from the data set, and divide these values into k segments using the method presented in Subsection 2.2.1.

2) Based on the data processing in step 1, set a minimum support number and mine all of the FCI.

3) Every FCI contain a certain number of samples, collect these samples and make them to be a new data set D.

4) Learn bayesian network from data set D which is formed in step 3, using K2 algorithm.

5) For samples without knowing is positive or is negative, calculate its positive probability by bayesian network constructed in 4. Once we get the probability, we can get result by comparing it with given threshold.

If a user is predicted to be positive, then the recommendation system will send the advertisement message to this user. On the other hand, user who are predicted to be negative will not receive advertisement messages.

6. Experiment

In experiments we try the new method as well as method based on naive bayesian claasifier so that we can compare the results between both methods.

6.1. Classify Using Naive Bayesian Classifier

1) Discretize the values and train naive bayesian classifier.

2) Give a decision threshold and classify the test data set using the trained naive bayesian classifer. Results are shown in Table 1.
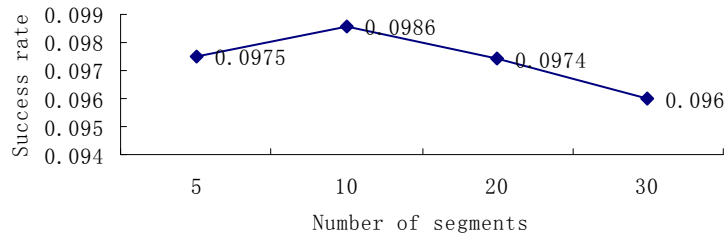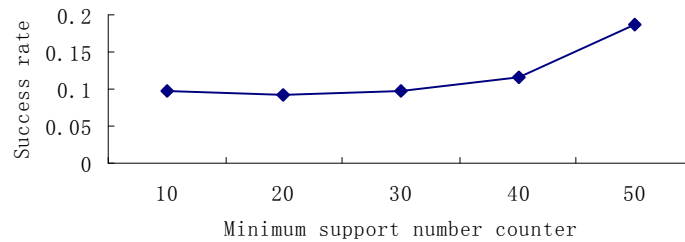
Figure 3. Success rate for different segments number.

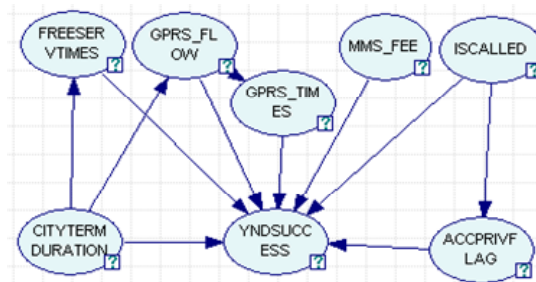**Figure 4. Success rate for different Minimum support number.**

Figure 5. Bayesian network learned by GENIE.

6.2. Steps of New Method

1) We divide the values of each attribute into k segments. While k is 10, 20, 30 respectivly, the results are shown in Figure 3. It shows that the best result occurs while k is 10.

2) Given a minimum support number α and search FCI, then compare the experimental results when α is 10, 20, 30, 40, 50 respectivly. The results are shown in Figure 4. It shows that we get the highest success rate while α is 50. We don't consider a minimum support number larger than 50 because the coverage of potential customer will decrease too much, see the data in Table 2.

3) Learn bayesian network from *D* using a open source bayesian tool called GENIE (http://genie.sis.pitt.edu), which uses K2 algorithm to learn Bayesian network structure. We get a Bayesian network shown in Figure 5, where GPRS_FLOW, MMS_FEE, etc are names of attributes. YND-SUCCESS is the target attribute to be predicted.

4) Last step is predicting by Bayesian network classifier. Since the number of negative samples is much more large than number of positive samples, we set the threshold to be 0.1. That means if a sample has a probability to be positive sample larger than 0.1, then it is determined to be positive.

6.3. Experimental Results

In this experiment, we use a data set from a mobile service provider. We divide the users into two parts. One part with the user number of 12264 is training data set. The other part with 12263 users including 577 positive users is test data set.

For traditional method which sends advertising message to all users. It sends 12264 advertisements and gets 577 customers, the success rate is 4.7%. It sends advertisement to 100% of the users, and gets 100% of the potential customers.

For method based on naïve Bayesian classifier whose results shown in Table 1, the results are almost the same as the results of traditional method. Obviously, naïve Bayesian classifier does not work well with mobile user's data set.

For new method, the results are shown in Table 2. Take mininum support number 50 for example, it sends 1621 advertisements and gets 303 customers, the success rate is 18.69%. It sends advertisement to 13.22% of all users, and gets 49.78% of the potential customers.

The results show that this user selection method increases success rate efficiently. Even the lowest success

Table 1. Results of naïve Bayesian classifier.

Threshold	Positive	Correct	Success rate
0.01	12172	596	4.89%
0.03	12135	595	4.90%
0.05	2704	164	6.07%
0.1	17	1	/

Table 2 Experiment result.

Min support Number	Positive	Correct	Success rate	Cost	Coverage of potential customer
10	5709	563	9.86%	46.55%	92.73%
20	5341	495	9.27%	43.55%	84.95%
30	4311	425	9.86%	35.15%	71.71%
40	3348	386	11.53%	27.30%	63.86%
50	1621	303	18.69%	13.22%	49.78%

rate 9.85%, is dramatically higher than the success rate of traditional method and of naïve Bayesian classifier. It also saves advertising cost and cuts down Spam remarkably.

7. Conclusions

Recomendation system becomes more and more important to service providers now. User selection is one of the difficulty problems. This paper presents a method to select users, the method searchs the maximum frequent itemsets from a large data set, then find potential customers from those maximum frequent itemsets by using a bayesian network classifier. The success rate is improved dramatically after using the method. This method also cuts down advertisement cost for mobile operators and avoids making large number of Spam messages.

There are a lot of data sets which have similar features with mobile user's data set, so this method can be used in many similar advertising recommendation systems. It has a good universality.

8. Acknowledgement

9. References

[1] E. Rich, "User modeling via stereotypes [D]," Cognitive Science, Vol. 3, No. 4, pp. 329–354, 1979.

[2] U. M. Fayyad, G. Piatecsky-Shapiro, and P. Smyth, "Advances in knowledge discovery and data mining [M]," California, AAAI/MIT Press, 1996.

[3] R. Agrawal and M. Srikant, "Fast algorithms for mining association rules in large databases [R]," IBM Almaden Research Center, Tech Rep: RJ9839, 1994.

[4] J. W. Han, J. Pei, Y. W. Yin, *et al.*, "Mining frequent patterns without candidate generation: A frequent-pattern tree approach [J]," Data Mining and Knowledge Discovery, Vol. 8, No. 1, pp. 53–87, 2004.

[5] F. V. Jensen, "An introduction to Bayesian networks," Springer, New York, 1996.

[6] M. desJarins, "Representing and reasoning with probabilistic knowledge: A Bayesian approach," Uncertainty in Artificial Intelligence, pp. 227–234.

[7] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," Machine Learning, pp.131–163, 1997.

[8] G. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data [D]," Machine Learning, Vol. 9, No. 4, pp. 309–347, 1992.