ECAI 2018 - International Conference – 10th Edition
Electronics, Computers and Artificial Intelligence
28 June -30 June, 2018, Iasi, ROMÂNIA

Similar Cluster recommendation of Product Purchases by Pages liked Analysis

Chantima Buaklee
Department of Computer Engineering
Faculty of Engineering, Chulalongkorn University
Bangkok, Thailand
Chantima.B@student.chula.ac.th

Sukree Sinthupinyo
Department of Computer Engineering
Faculty of Engineering, Chulalongkorn University
Bangkok, Thailand
Sukree.S@Chula.ac.th

Abstract – **In recent years, social network become more popular and influential on the online commerce. Many customers can find the detail of products and trends on the social network using it to make the decision whether to purchase the product or not. In addition, it is also commonly used to collect users' related data to predict their future actions. In this paper, we study the correlation of purchasing histories and the customer's interests on social network (Facebook) to predict the future interests of customers that would show through social network. In addition, we use the data to find more similar target group of customers who are also interested in the same category of product by text analysis with Facebook Pages Liked. We focus on the product categories aligned by the shopping online websites and find the customer cluster which is similar with the customer who purchased. The study results are beneficial to focus groups of customers for communication or suggestion about the products matching with the groups of customers who are interested in the same group of products and have high potential to buy them. This model can adept with many real-world businesses to identify the user's characteristics and find more similar group of customers who interest in the products by changing the user attribute to find the user's interest. In our study, we adapt this model with the prototype system to recommend the similar group of customers.**

Keywords-social network; similar cluster; similar cluster recommendation; interest behavior; page liked; text classification; purchased history

I. INTRODUCTION

Nowadays, method of focusing on customer groups, who are interested in the product or using it, before communicating get more attention from the business. Due to more effectiveness of the attached rate which customers turn to purchase than those of communication with a mass group of customers, the usage patterns and history analysis are popular methods used to focus a group of customers before communication [1]. For example, once customer visits a page to review or buy the product. The system will record the customer pattern and use these data to suggest the other products which have the same category via the online ads. Although, this method is limited by the time due to the changing in the interests of customers through time. However, the ads still show recommendations with the old pattern history. Moreover, in the recent years, social network become more influential on the online commerce [2] and also has evolved into be more efficient in communicating, sharing information [3] and representing personal interests by likes activity. We believe that Facebook's number of likes reveal the mutual influences among the semantically related entities [8].

In our study, we focus on the Facebook Pages Liked which are already separated by the categories. In addition, we found that Facebook page categories still have a narrow meaning, or some pages are not explained about the page appearance then we deeply found more about the characteristics of each page by retrieving the data feeds in Facebook pages, using publicly Facebook Application Programming Interface (APIs) to analyze and identify the real page appearance.

We found that customer page liked activities can identify the interests of customers showed on the online society and had more correlation with purchasing from the online commerce. After we can specify the interest of old customers who have purchased histories then we will find the similar group of customers who have similar interests.

In this paper, we claim that interest's behavior on the social network of old customers purchasing on online commerce can identify the similar group of new customers who also have potential to buy the products. For example, the customers who like Facebook pages of beauty will purchase the cosmetic products then if we can identify other customers who also have the similar interest in beauty pages categories then we can classify the new group of customers who have the potential to purchase the beauty products.

The following contents of the paper are organized as: Section 2 reviewing related previous works to apply in our study, section 3 describing our proposed method step by step, section 4 discussing about the experimental results and analysis, and finally, the paper conclusions and future work idea in section 5.

II. RELATED WORKS

Service recommendations work base on service usage patterns by identifying the personal attribute-based clustering and similarity of rating preference. The method for recommendation uses the

978-1-5386-4901-5/18/$31.00 ©2018 IEEE

collaborative filtering, which has been implemented in many real word businesses. The historical usage patterns still have the limit among the temporal variations. This approach provides the service usage pattern transformation with the multi-period of time [1].

A. Collaborative Filtering

Collaborative Filtering (CF) does the recommendation by learning other customers' previous preferences from the set of service then calculates similarity weight between customers to recommend the potential favorite products for a user [1]. Some studies use Collaborative Filtering method to suggest new products to a customer by finding the set of products that similar with customers' interests while using Facebook categories and likes to be factors to identify the customer characteristics [5].

B. Naïve Bayes Classification

Due to narrow definitions of Facebook page categories that make the discrepancy when classify the customer characteristics, this study will use the text mining with Facebook feeds to identify the characteristics of each Facebook page by learning model from the set of data using the statistic to classify the item [1]. Naïve Bayes Classification is a well-known and effective method in machine learning [6] that refer to the classification problem to predict the value of a categorical attribute [7]. Bayes rules is the probabilistic models that we focus and can be explained as the equation below.

$$P(H|E) = P(E|H) \times P(H) / P(E) \quad (1)$$

P(H|E) is a conditional probability that situation H happened first then E, H is the hypothesis of any classification values that need to be predicted and E is the evidence to give the values of attributes then we can explain that probability of data which have attribute E will be H class. In our study we used Bayes theorem to do text classification with feed data then classify the interest of customers align by product categories in the shopping online website.

C. Clustering

The customers clustering is for assigning the most similar customers in the same segment by using the similarity matrix to identify the customers in each cluster which contain the most similarity and then recommend products by using purchased history.

These algorithms start with an initial set of our target groups of customers [9] which selected by the corrective results of text classification matching with purchasing results. We apply k-means to cluster customer and calculate the Euclidian distances between the target groups of customers who purchased products and the cluster centers [1].

D. K-fold cross validation

This method split the experimental datasets into two groups, one for data training (the initial target customers who purchased products) and another for the test set (the set of customers who get the recommendation) [1]. In our study, we used 10- folds to validate our model.

III. METHODOLOGY

In this section, we will describe our propose method through the details that are divided in two parts. First part is the data collection, and the second is the frame work of our approach.

A. Data collection

The data set used in our study come from Thai online shopping websites. We randomly selected the purchasing data of any customer who has Facebook account then we used public available Facebook APIs to retrieved Facebook Pages data. The data of our study are as the following information.

- Facebook Pages Liked feeds are used to identify customer characteristic in each product category.
- Purchased history of Thai shopping online websites. We randomly selected 300 customers in the real-world business to learn the user characteristics.

B. Framework of an approach

In this section, we describe our method that show in the Fig. 1 which is divided in 3 steps. First step is gathering customer behavior data. Next is to classify the interest of customer in each product category, and last step is finding the new target group of customers.

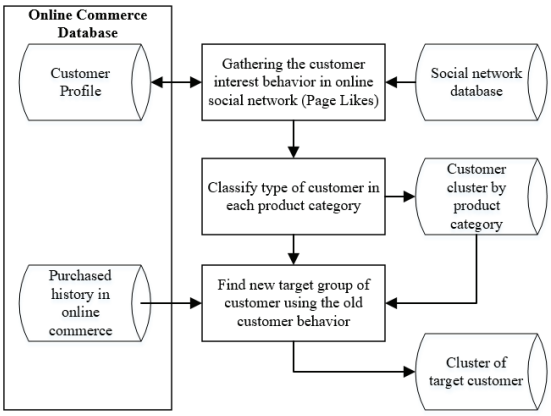

Figure 1. Framework of an approach

Step 1: Gathering customer behavior data

In this step, we gathered the customer behavior data from online social network, Facebook, by using the customer accounts from online commerce website. The behaviors that we retrieved are the customer Pages liked activities and page's feed. Those behaviors will be the attributes to identify the customer interest's behaviors in each product categories.

Step 2: Classifying the interest of old customer in each product category

We provided the ontology of keywords that related with the targeted product category by using the

product description on the online commerce and retrieved Facebook page's feeds which are identified by the experts in each category. In our study, we focus on the Facebook pages electronic, beauty, sport and camera categories that align by the product categories in the shopping online website. After we had identified the ontology of keywords, we retrieved Facebook Page's feeds of each customers and do text mining by Naïve Bayes theory to classify customer liked pages in each category. The prediction results will prioritize by categories.

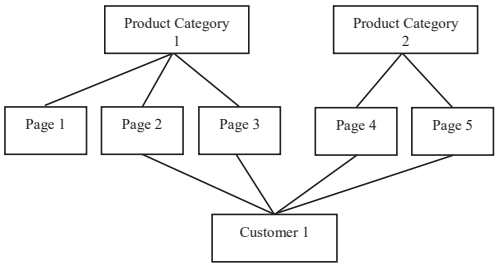

Figure 2. Customer interest in each product categories

Fig. 2 describes that each customer can be classified into more than one category by the interests. For the example, customer who is interested in beauty product might be also interested in the category of sport and others.

In this step, we can identify personal interests of each customer. The example shows on Table I. Customer 1 was classified by priority text mining results in Electronic, Beauty, Sport and Camera categories.

TABLE I. IDENTIFY PERSONAL INTEREST OF EACH CUSTOMER

Customer	Shopping online website product category			
	electronic	beauty	sport	camera
Customer 1	0.55400	0.31000	0.10000	0.03700

Index	Nominal value	Absolute count	Fraction
1	[redacted]	9508	1

Figure 3. Customer word count

Index	Nominal value	Absolute count	Fraction
1	electronic	5264	0.554
2	beauty	2946	0.310
3	sport	947	0.100
4	camera	351	0.037

Figure 4. Text classification in each category

Fig.3 and Fig.4 show word count of each customer and text classification in each product category to identify the customer's characteristic.

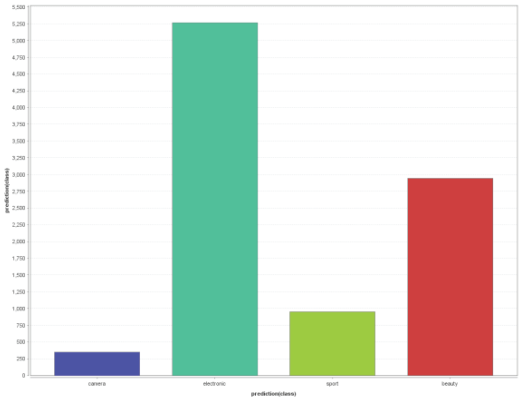

Figure 5. Text classification in each product category

Fig. 5 shows graph of text classification on page's feeds identified in each product category. After we knew the personal interest of customer, we compared it with the product purchased history to cross check the corrective result.

Next step we selected the corrective customers by top 3 categories comparing with product purchased. If the predictive result shows that the personal interest match with product's purchased history, then we keep that customers in the initial target group of old customers. For the assumption of our study, we assume that

"If the customer is interest in or like any page categories, the customer will buy the products which is relevant to that page category".

Steps 3: Find the new target group of customers

After we could identify the initial target group of old customers, then we can find the similar group of other customers who have not bought these products to increase the potential of business to propose or recommend any products related with their interests. In our study, we use K-means to classify the group of customers and Euclidean Distance to find the similarity between old customers who bought the products and new target customers who have the same interest behavior. We could classify customer into 9 clusters and the clusters sum of squared errors shows 4.8022

```
=== Model and evaluation on training set ===

Clustered Instances

0       38 ( 13%)
1       35 ( 12%)
2        8 (  3%)
3       58 ( 19%)
4        9 (  3%)
5       29 ( 10%)
6       53 ( 18%)
7       49 ( 16%)
8       21 (  7%)
```

Figure 6. Customer classification in 9 cluster

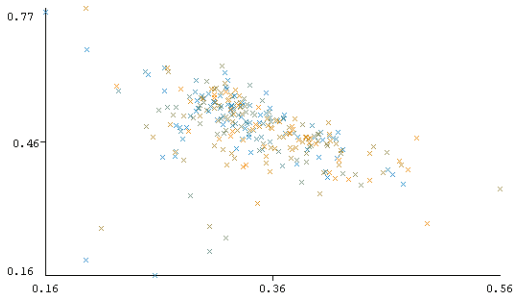

Figure 7. Graph of customer classification in each cluster

Fig.6 and Fig.7 show the result of customer classification in each product category

IV. Experiment Results

In this section, we show the results in each step of our method.

A. Ontology of words aligned by product categories

We define the ontology of words aligned by the online commerce's product categories and retrieved Facebook's pages that relevant in each product type judged by the experts. After that, we did the text analysis by retrieved 100 news feeds from each page, the total is 1578 feeds and cut words to put in the ontology. The ontology also includes 2951 product item descriptions classified in each product category.

B. Identification of a customer's personal interests

From the group of customers randomly selected from the shopping online websites, we retrieved 50 news feeds of each customer's pages likes to do text analysis using Naïve Bayes theorem. We used it to classify the interests of customer that match with the defined ontology. Table II shows the result of Naïve Bayes text classification with accuracy rate is 89.46%.

TABLE II. Number of Selected Pages by Categories

accuracy: 89.46% +/- 2.21% (mikro: 89.46%)					
	true electronic	true camera	true beauty	true sport	class precision
pred. electronic	580	21	38	6	89.92%
pred. camera	10	352	4	0	96.17%
pred. beauty	42	16	353	2	85.47%
pred. sport	8	6	11	107	81.06%
class recall	90.62%	89.11%	86.95%	93.04%	

C. Selection of the initial target group of customer who purchased the product

We compared the predictive results of customer interests in each product category with purchased history to find the customer who shares similarity between interest showed on social network and the online purchasing to be the initial target groups.

D. Finding the similar customer behavior

After we can identify the initial group of customers who show relevant between the social media interest behavior and the online purchased history, we find more customers who are closely similar with our targeted groups. We found that our new user being closely to cluster 0 which is classified in Electronic group. By checking the result with the purchased history of this user, the result is also showed correctly.

E. Validation

We do 10-Folds cross validation to validate our model. The error and accuracy rate are shown in Fig.8

```
=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.95
Mean absolute error                      0.0013
Root mean squared error                  0.0056
Relative absolute error                  9.8604 %
Root relative squared error             31.3246 %
Total Number of Instances              300
```

Figure 8. 10-folds cross validation accuracy results

In Addition, for the algorithms, we use Weka and Rapid Miner (Student license) and the execution environment is Intel® Core i5-7200U2.5 GHz with Turbo Boost up to 3.1 GHz, 4.0 GB RAM.

V. Conclusion and Feature Work

In our study, we focus on the correlation between the customer behaviors on online social network (Facebook) and the purchased histories in shopping online website by retrieved the interest using Facebook Pages Liked and then retrieve feeds from each page to do the text analysis to classify the characteristic of page aligned by the product category of shopping online websites. After we knew the customer's characteristic, we cluster the customers and find the initial target group of customers by matching the result of product purchased then calculating the Euclidean distance to find the similarity between our target customer and the others.

After we created the model, we develop the prototype system to recommend the similar cluster of customer who has more potential to buy a product in each product categories. By putting in user's Facebook accounts, the system will run the model step by step. The output would show the results of user cluster in each product categories and can be exported as an Excel file, the advantage is that it is easy to use the list of user's accounts to recommend the products.

However, this study still has the limitation of the data set due to the publicly Facebook API. For the future work, we can adapt our model with another dataset and other data attributes to find more customers who has the similarity in the product purchased history.

REFERENCES

[1] Ruilin Liu; Xiaofei Xu; Zhongjie Wang, "Service Recommendation Using Customer Similarity and Service Usage Pattern," in Web Services (ICWS), 2015 IEEE International Conference on , vol., no., pp.408-415, June 27 2015-July 2 2015.

[2] K. Y. Kwahk and X. Ge, "The Effects of Social Media on E-Commerce: A Perspective of Social Impact Theory," System Science (HICSS), 2012 45th Hawaii International Conference on, Maui, HI, 2012, pp. 1814-1823.

[3] Y. Tsuboi, A. Jatowt and K. Tanaka, "Product Purchase Prediction Based on Time Series Data Analysis in Social Media," 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Singapore, 2015, pp. 219-224.

[4] Nathan N. Liu, Qiangs Yang. Eigenrank: a ranking-oriented approach to collaborative filtering. the Proceedings of ACM SIGIR '08. SIGIR '08, 2008: 83-90.

[5] Yongzheng Zhang and Marco Pennacchiotti. 2013. Predicting purchase behaviors from social media. In Proceedings of the 22nd international conference on World Wide Web (WWW '13). ACM, New York, NY, USA, 1521-1532.

[6] Z. Qin, "Naive Bayes Classification Given Probability Estimation Trees," 2006 5th International Conference on Machine Learning and Applications (ICMLA'06), Orlando, FL, 2006, pp. 34-42.

[7] F. Klawonn and P. Angelov, "Evolving Extended Naive Bayes Classifiers," Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), Hong Kong, 2006, pp. 643-647.

[8] Ohsawa, Shohei and Yutaka Matsuo. “Like prediction: modeling like counts by bridging facebook pages with linked data.” WWW (2013).

[9] G. Linden, B. Smith and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," in IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan/Feb 2003.