Scientific
Research
Publishing

Research on the Personalized Recommendation Algorithm for Hairdressers

Zhixin Liang, Fengying Wang

School of Information, Beijing Wuzi University, Beijing, China
Email: liangzhixin@sina.com

Abstract

In order to retain customers, hairdressers usually persuade customers to be their members by offering membership card. This paper studies how to set up their recommendation system in the hairdressers. According to the membership information and consumer behavior, the hairdresser provides personalized recommendation to different members and lets customers experience personalized choices. A recommendation algorithm based on customer ratings and a customer classification method based on Logistic Regression Model are discussed in this paper. The former is used to recommend hair style and color to a customer. The latter is used to determine whether to recommend some maintenance programs to a customer or not.

Keywords

1. Introduction

Competition in the market is increasingly fierce in hairdressing industry. Customer source becomes the lifeline of the hairdresser's survival and development. In order to retain customers, hairdressers usually persuade customers to be their members by offering membership card.

Membership marketing is a new type of marketing strategy developed in the 1990s. In order to improve customer loyalty and realize the purpose of increasing corporate long-term profits, the company provides differentiated services and precise marketing for their members. For the hairdressing industry, loyal customers are the root of its profits.

Initially, a customer is recognized as a loyal or potential loyal customer only when the amount of consump-

tion reaches a certain number. Then, the enterprise ranks the customer as a member and provides personalized service or corresponding preferential treatment. But now, a customer becomes the hairdresser's member only by purchasing a membership card. Different preferential treatments correspond with different prices of the membership card. In fact, hairdressing is a highly individualized service industry. Usually, price is not the first concern. A customer will continue to consume at the store if the stylist can design the customer's hair according to his/her characteristic or preference. This personalized service makes the customer trust and depend on the hairdresser and then customer loyalty is improved.

Haircut, hair perm, hair dying, hair styling and maintenance are usually included in hairdressing services. So hairdressing services must involve hairstyle choices and color choices. In a hairstylist website, there are more than 1.5 million pieces of hairdo related images. Looking for a hairstyle suiting for your hair length, such as short hair, you will get more than 30,000 pictures. If you want to color your hair, you will have hundreds of color to choose from. Facing massive information, customers will have difficulty making choices. To help customers make choices, the hairdresser should provide a product display according to the characteristics of different customers. This kind of personalized service can make customers to face less and suitable choices and this accurate service will enhance customers' trust and dependence on the hairdresser's and thus become loyal customers.

Recommender system is a software system, a kind of tool which helps customers find the information conformed to their interest quickly in the vast amounts of information. It is now being widely used in our daily life. Shadanand, U. *et al*. [1] discussed the implementation of a networked system called Ringo, which makes personalized recommendations for music albums and artists. Konstan *et al*. [2] discussed a collaborative filtering system for Usenet news—a high-volume, high-turnover discussion list service on the Internet. Mobasher *et al*. [3] described an approach to usage-based Web personalization taking into account the full spectrum of Web mining techniques and activities. Badrul Sarvar *et al*. [4] introduced three similarity formulas and discussed techniques for obtaining recommendations from the similarities. G. Adomavicius and A. Tuzhilin [5] presented an overview of the field of recommender systems and described the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches. Y. Koren [6] introduced some innovations to merge the factor and neighborhood models and built a more accurate combined model. Y. Shi, M. Larson and A. Hanjalic [7] proposed a novel algorithm, Tag-induced Cross-Domain Collaborative Filtering (Tag CDCF), which exploits user-contributed tags that are common to multiple domains in order to establish the cross-domain links necessary for successful cross-domain C. F. M. Deshpande and G. Karypis [8] presented one such class of model-based recommendation algorithms that first determines the similarities between the various items and then uses them to identify the set of items to be recommended.

In recent years, many approaches for building recommender systems have been developed. But few people focus on the hairdressing industry. Hairdressing is a highly personalized industry. A hairdresser's can build its own recommender system that utilizes either the membership information, or consumer behavior. Consumers experience personalized choices from different personalized recommendation. This article discusses recommendation algorithm based on customer ratings which is used to recommend hair style or color and customer classification method based on Logistic Regression Model which is used to determine whether to recommend some maintenance program to a customer or not.

2. Recommendation Algorithm Based on Customer Ratings

Based on the membership information, consumer behavior and preferences, hairdressers recommend hairstyle, hair color to their members. So the details of members should be covered in database such as **Table 1**.

In order to record what a customer like, the hairdresser's should ask the customer to score hair styles and colors according to a 5-point scoring model with extreme like, like, in general, dislike, extreme dislike corresponding

Table 1. Information of a member.

Name	Demographic information: gender, age, occupation
	Customer's interest: short hair, long hair, straight hair, curly hair, color
	Customer's propensity to consume: income, level of education, cognitive approach of the hairdresser's (introduced by relatives and friends, from advertising and passing found)
	The stylist professional judgment information: feature, hair quality, skin color

5, 4, 3, 2, 1 respectively. Assume that there are *n* hair styles and *m* colors for customers to choose from in the system and there are *k* members belong to the hairdresser's. A customer rating form will be gotten, which is a $k \times (n+m)$ matrix. Among them, columns 1 to *n* are about hair styles and columns $n+1$ to $n+m$ are about hair colors. Sign s_{ij} is used to record the score Customer *i* applied to hair style or hair color *j*.

Assume that the current customer is Customer *a*. If s_{aj} is null, Customer *a* is thought not to focus on style or color *j*. The hairstyle or color that Customer *a* might like should be find by the recommender system from the hairstyle or color that Customer *a* never gives attention to.

A person's hairstyle or color preference would not change for a certain period. Besides a person's hairstyle or color preferences is usually influenced by his/her feature, skin color, personality, age, and occupation. Therefore if some people had similar preferences in the past, they may have similar preferences in the future. This characteristic is suitable for selection recommendation algorithms based on user's nearest neighbor. It is said that someone whose preference was similar to Customer *a* in the past will be found by using the customer rating form. These customers constitute the similar customers set of Customer *a*. Using these customers' rating, the score of the hairstyle or color which Customer *a* did not focus on should be predicted by the recommender system. If the predicted score reaches a certain value, the corresponding hairstyle or color will be recommended to Customer *a*.

Customers of ten pay more attention to hairstyles which length is similar to his/hers. Hairstyles can be classified by length, for example by short, medium short, medium long, long four sets. Then, respectively, there are $n_1, n_2, n_3, n_4 (n_1 + n_2 + n_3 + n_4 = n)$ hair styles in each set and the rating form is a $k \times (n+m) = k \times \left[(n_1 + n_2 + n_3 + n_4) + m \right]$ matrix. Similar customer set of Customer *a* will be produced under the corresponding class of Customer *a* in hair style recommendation and similar customer set of Customer *a* will be produced in all customers in hair color recommendation.

For example, a female short hair style recommendation will be given to Customer *a* who had left a scoring record for the hair style $j_1, j_2, \cdots, j_q$. Assume that there are total n_1 kinds of hair style to show and the number of short hair members is *t*. The following two recommendations will be provided to Customer *a*. First, the hairstyle she liked before should be found from hairstyle $j_1, j_2, \cdots, j_q$ and be shown. Second, hairstyles she might like should be found from the left $n_1 - q$ kind of hairstyles and shown. Next, the second recommendation will be discussed.

Pearson correlation coefficients is used to calculate the degree of similarity between Customer *a* and Customer *i* $(i = 1, 2, \cdots, a-1, a+1, \cdots, t)$.

Let $N_1 = \{1, 2, \cdots, n_1\}$, $Q = \{j_1, j_2, \cdots, j_q\}$.

$$sim(a,i) = \frac{\sum_{j \in Q} (s_{aj} - \bar{s}_a)(s_{ij} - \bar{s}_i)}{\sqrt{\sum_{j \in Q} (s_{aj} - \bar{s}_a)^2} \sqrt{\sum_{j \in Q} (s_{ij} - \bar{s}_i)^2}}, (i = 1, 2, \cdots, a-1, a+1, \cdots, t) \quad [9] \tag{1}$$

Sort $sim(a,i)(i = 1, 2, \cdots, a-1, a+1, \cdots, t)$. The members who have the top p values to constitute similar customer set of Customer *a* will be selected. Assume that they are Customer $i_1, i_2, \cdots, i_p$ and let $P = i_1, i_2, \cdots, i_p$.

The predictor of hairstyle $j (j \in N_1 - Q)$ of Customer *a* is calculated, which is denoted as s_{aj},

$$s_{aj} = \bar{s}_a + \frac{\sum_{i \in P} sim(a,i)(s_{ij} - \bar{s}_i)}{\sum_{i \in P} sim(a,i)} \tag{2}$$

When $s_{aj} \geq 4$, hairstyle *j* will be shown to Customer *a*.

3. Customer Classification Method Based on Logistic Regression Model

The hairdressers' service is usually classified into two categories: hairdressing and maintenance. Maintenance is usually classified into two categories: hair maintenance and scalp maintenance. The scalp maintenance is a program which only few people choose. Over-recommendation may disgust customers and drive them away. How to decide whether to recommend scalp maintenance program to a new member or not?

A wealth of information of customers has been accumulated in the member database such as gender, occupation, age, monthly average income, personality preferences, address, cognitive approach(how customers get to know the hairdresser's), consumption programs, etc. They are named indexes. Choosing whether to maintain as the classification basis, a variable $Y = \begin{cases} 0 & \text{maintenance} \\ 1 & \text{no maintenance} \end{cases}$ will be produced. Let the above indexes respectively

be $X_1, X_2, \cdots, X_n$. Let Y be the dependent variable and $X_1, X_2, \cdots, X_n$ be the independent variable. Using Logistic Regression Model, the relationship between Y and $X_1, X_2, \cdots, X_n$ will be established.

Let p_i be the probability of Customer i choosing the maintenance program, then

$$p_i = P\left(y_i = 1 \mid (x_{i1}, x_{i2}, \cdots, x_{in})\right) \tag{3}$$

The logistic regression model can be gotten, [10]

$$p_i = \frac{1}{1 + e^{-(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in})}} = \frac{\exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in})}{1 + \exp(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in})} \tag{4}$$

Using the maximum likelihood estimation method, estimate values of the parameters $\alpha, \beta_1, \beta_2, \cdots, \beta_n$ will be offered, denoted as $\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_n$. It is important to note that in order to maintain statistical properties of maximum likelihood estimation, the membership has to reach a certain size, usually above 500 or at least more than $\min(100, 10n)$.

Let $x_1, x_2, \cdots, x_n$ be the value of the index $X_1, X_2, \cdots, X_n$ of a new customer, then the probability the customer may accept maintenance recommended is

$$\hat{p} = \frac{\exp(\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n)}{1 + \exp(\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n)} \tag{5}$$

Let the preset threshold value be θ, when $\hat{p} \geq \theta$, maintenance program will be recommended to the new customer.

4. Simulation on Examples

4.1. The Short Hair Recommendation

Assume that current customer is Customer a, female, short hair. There are a total of 35,000 (10,000 male, 25,000 female) members in a salon chain, and there are 3000 female in the short hair set and there are 300 kinds of short hair styles to recommend. The process is shown by **Figure 1**.

Usually, two problems must be solved by the recommender system when a score of a customer about a hairstyle or color should be predicted. The first is that there is no customer rating data when the recommender system has just been started up or the recommender system is faced with new customers. The second is hundreds or thousands of hair styles and hair colors are provided, but a customer can't score every style or color. A scoring record of hairstyle $j_1, j_2, \cdots, j_q$ Customer a has been left, the similar customer set will be found according to these q column data. But some scores of these q columns of the rest customers may be null. Whether producing similar customer set of Customer a or predicting the score about hairstyle $j(j \in N_1 - Q)$ of Customer a, loss scores of the rest customers must be added.

Using expert recommendation and expert evaluation system is a good solution. That is to say that a stylist rating form should be used. Every stylist scores each style and color from professional angles, according to a customer's age, gender, occupation, feature, skin color and so on. Take the average to form an expert rating form. When the recommender system is started up or faced with new customers, the expert evaluation system is used directly. If a data item is null when the similar customer set of a regular customer should be found, the stylist's score is added.

In this case, current customer has scored 38 kinds of hairstyles and she likes 14 kinds of them. These 14 kinds of hair styles were shown at first and then the top 20 kinds of hair styles she might like were recommended to her.

Besides expert evaluation, there are many other kinds of algorithm about how to add s_{ij}, [11] such as the global average that is the average of all scoring record, the average of all score of Customer i, the average of all score of hairstyle j. It is shown that customers tend to use the stylist scores to add the loss data after surveying 120 customers. It shows the importance of an expert system in the hairdressing industry.

4.2. Numerical Example of Classification about Members

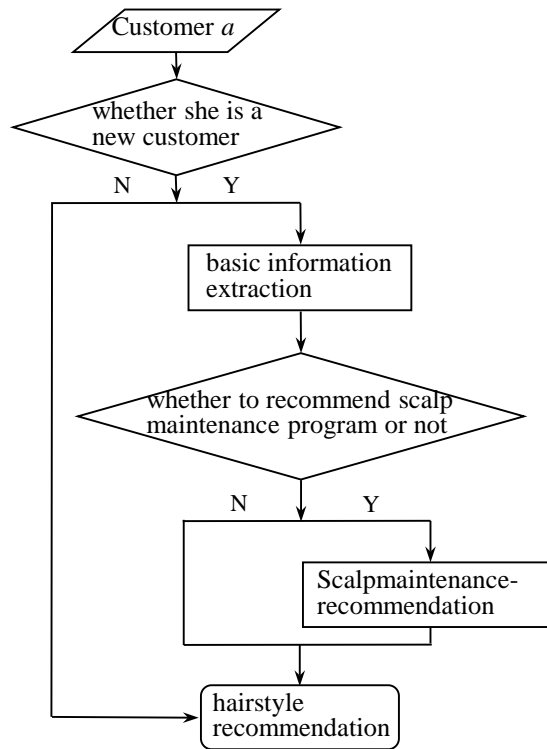In this case, samples were extracted which sample size is 500. Six indexes are selected to constitute independent

Figure 1. Flow chart.

variables: gender, occupation, age, monthly salary, education years and the cognitive approach, among which gender, occupation and cognitive approach are classification variables, their possible values are

$$\text{gender} = \begin{cases} 0 & \text{male} \\ 1 & \text{female} \end{cases}$$

$$\text{professional} = \begin{cases} 1 & \text{Professional and technical personnel} \\ 2 & \text{Administrative personnel agencies enterprises and institutions} \\ 3 & \text{student} \\ 4 & \text{other} \end{cases}$$

$$\text{The cognitive approach} = \begin{cases} 1 & \text{Relatives and friends introduce} \\ 2 & \text{From advertising} \\ 3 & \text{Their discovery} \end{cases}$$

Selecting the stepwise regression, Logistic regression analysis has been gotten as the following **Table 2**.

As can be seen from the **Table 2**, there are four variables in the regression model: gender, monthly average income, education years and cognitive approach. The cognitive approach is originally a three classification variable. In the process of Logistic regression it is transformed into a set of design variables, they are cognitive approach (1), cognitive approach (2) and cognitive approach (3). Cognitive approach (3) is considered as a reference, the corresponding values are

$$\text{cognitive approach}(1) = \begin{cases} \text{relatives and friends introduce} \\ \text{other} \end{cases}$$

$$\text{cognitive approach}(2) = \begin{cases} \text{from advertising} \\ \text{other} \end{cases}$$

Table 2. Logistic regression output (variables into the model).

Variables into the model	Regression coefficient B	Standard error	Wald value	Degrees of freedom	Sig.	Exp (B)
Gender	−2.730	0.940	8.442	1	0.004	0.065
Monthly average income	0.001	0.000	44.339	1	0.000	1.001
Education years	−0.150	0.060	6.238	1	0.013	0.861
Cognitive approach			46.061	2	0.000	
Cognitive approach (1)	4.474	0.971	21.220	1	0.000	87.747
Cognitive approach (2)	2.761	0.424	42.457	1	0.000	15.821
Constant	−2.650	0.981	7.297	1	0.007	0.071

$$\text{cognitive approach} = \begin{cases} \text{their discovery} \\ \text{other} \end{cases}$$

Being a reference, cognitive approach is not gotten into the regression model. Its estimated coefficient is 0. Set gender, monthly salary, education years, cognitive approach (1) and cognitive way (2) as X_1, X_2, X_3, X_4, X_5,

$$\hat{p} = \frac{\exp\left(-2.56 - 2.73x_1 + 0.001x_2 - 0.15x_3 + 4.474x_4 + 2.761x_5\right)}{1 + \exp\left(-2.56 - 2.73x_1 + 0.001x_2 - 0.15x_3 + 4.474x_4 + 2.761x_5\right)} \tag{6}$$

Assume that there are two new members: Member 1(male, mean monthly salary 6200 yuan, graduated from the university, found the hairdresser's from advertising) and Member 2 (female, mean monthly salary 4500 yuan, graduated from university, saw the hairdresser's when shopping). Their basic information can be respectively described as

$$X_1 = \begin{pmatrix} 0 & 6200 & 16 & 0 & 1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 4500 & 16 & 0 & 0 \end{pmatrix}.$$

According to the formula (2), the probability estimates that they will accept scalp maintain can be calculated respectively,

$$\hat{p}_1 = \frac{\exp\left(-2.56 + 0.001 \times 6200 - 0.15 \times 16 + 2.761 \times 1\right)}{1 + \exp\left(-2.56 + 0.001 \times 6200 - 0.15 \times 16 + 2.761 \times 1\right)} = 0.9820 \tag{7}$$

$$\hat{p}_2 = \frac{\exp\left(-2.56 - 2.73 + 0.001 \times 4500 - 0.15 \times 16\right)}{1 + \exp\left(-2.56 - 2.73 + 0.001 \times 4500 - 0.15 \times 16\right)} = 0.0395 \tag{8}$$

Assume that the default to recommend the consumption to a candidate is more than 0.7. Since $\hat{p}_1 > 0.7, \hat{p}_2 < 0.7$, the scalp maintain should be recommended to Member 1.

5. Conclusion

Hairdressing is a highly personalized industry. And consumers have better personalized experience with the help of personalized recommendations. So a recommender system should be built in a hairdresser's. According to the membership information and consumer behavior, what to recommend to a particular customer may be decided by the system. This paper discusses recommendation algorithm based on customer ratings and customer classification method based on Logistic Regression Model and gives simulation examples. When similar customer set of Customer a is produced or the score about hairstyle which Customer a might like is predicted, loss scores of the rest customers must be added. There are many kinds of algorithm about how to add the loss scores. The survey shows that customers tend to use the stylist scores to add the loss data. It shows the importance of an expert system in the hairdressing industry.

References

[1] Shardanand, U. and Maes, P. (1995) Social Information Filtering: Algorithms for Automating "Word of Mouth". *Pro-*

ceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems, ACM, New York, 210-217. http://dx.doi.org/10.1145/223904.223931

[2] Konstan, J., Miller, B., Malzt, D., Herlocer, J., Gordon, L. and Riedl, J. (1997) GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, **40**, 77-87. http://dx.doi.org/10.1145/245108.245126

[3] Mobasher, B., Cooley, R. and Srivastava, J. (2000) Automatic Personalization Based on Web Usage Mining. *Communications of the ACM*, **43**, 142-151. http://dx.doi.org/10.1145/345124.345169

[4] Sarvar, B., Karypis, G., Konstan, J. and Riedl, J. (2001) Item Based Collaborative Filtering Recommendation Algorithms. *International Conference on the World Wide Web*, **4**, 285-295

[5] Adomavicius, G. and Tuzhilin, A. (2005) Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge & Data Engineering*, **17**, 734-749. http://dx.doi.org/10.1109/TKDE.2005.99

[6] Koren, Y. (2008) Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model. *Knowledge Discovery and Data Mining*, Las Vegas, 24-27 August 2008, 426-434.

[7] Shi, Y., Larson, M. and Hanjalic, A. (2011) Tags as Bridges between Domains: Improving Recommendation with Tag-Induced Cross-Domain Collaborative Filtering. *User Modeling*, *Adaption & Personalization-international Conference*, **6787**, 305-316.

[8] Deshpande, M. and Karypis, G. (2014) Item-Based Top-N Recommendation Algorithms. *ACM Transaction on Information System*, **22**, 143-177. http://dx.doi.org/10.1145/963770.963776

[9] Jannach, D., Zanker, M., Felfering, A. and Friedrich, G. (2013) Recommender System an Introduction. Posts & Telecom Press, Beijing.

[10] Wang, J.C. and Guo, Z.G. (2001) Logistic Regression Model: Method and Application. Higher Education Press, Beijing.

[11] Liang, X. (2012) Recommender System Practice. Posts & Telecom Press, Beijing.