Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands

Ya-Yueh Shih
https://dl.acm.org/profile/81100194056
Department of Information Management,
Chung Hua University, Taiwan

Duen-Ren Liu
https://scholar.google.com/citations?user=pQ7QfRsAAAAJ&hl=zh-TW
Institute of Information Management,
National Chiao Tung University, Taiwan

1 Introduction

Recommended systems are used in a variety of fields and are commonly identified as playlist generators for video and music services, product recommenders for online stores, or content recommendations for social media platforms, and web content recommendations. They assist businesses in implementing one-to-one marketing strategies, relying on custom-er purchase history to reveal customer preferences and identify products that customers may purchase. One-to-one marketing introduces a fundamentally new basis for competition in the marketplace by enabling organizations to differentiate based on customers rather than products. Recommendation systems have also been developed to research and locate experts, collaborators, and financial services.

The collaborative filtering (CF) method has been used successfully in various applications. It predicts the user's preferences for the item in a word-of-mouth manner. User preferences are estimated by looking at other users' views of "like-minded" (in the form of preference ratings). Amazon.com uses collaborative filtering to create book recommendations for customers. Content-based filtering (CBF) provides a different approach to collaborative filtering and provides recommendations by matching customer profiles (e.g., interests) with customer features (e.g., product attributes). We are exploring hybrid methods combining CF and CBF methods to overcome the disadvantages of each. This work uses customer demands derived from frequently purchased products in each industry as valuable information for integrating the CF method to make recommendations. Extended preferences derived by combining customer demands and previous purchasing preferences are used to reduce the sparsity problem of recommendation.

2 Recommendation Method

2.1. Typical KNN-based collaborative filtering

A typical KNN-based collaborative filtering (CF) method employs nearest-neighbor algorithm to recommend products to a target customer u based on the preferences of neighbors, that is, those customers having similar preferences as customer u. A customer-item matrix R represents customer preferences or customer purchase history such that, r_{ij} is one if the i^{th} customer had purchased the j^{th} product, and is zero otherwise.

$$Corr_P(c_i,c_j) = \frac{\sum_{s\in I}\left(r_{c_i,s}-\bar{r}_{c_i}\right)\left(r_{c_j,s}-\bar{r}_{c_j}\right)}{\sqrt{\sum_{s\in I}\left(r_{c_i,s}-\bar{r}_{c_i}\right)^2\sum_{s\in I}\left(r_{c_j,s}-\bar{r}_{c_j}\right)^2}}$$

The notations $\bar{r}_{c_i}$ and $\bar{r}_{c_j}$ denote the average number of products purchased by customers c_i and c_j, respectively. Moreover, the variable I denotes the set of products. Additionally, the $r_{c_i,s}$ and $r_{c_j,s}$ indicate whether customers c_i and c_j purchased product items. The k most similar customers are selected as the k-nearest neighbours of customer u. The products then are sorted based on frequency count. The N most frequent products that have not yet been purchased by target customer u are recommended to target customer u.

2.2. Weighted RFM-based CF method

This method mainly integrates the analytical hierarchy process (AHP), clustering, and association rule mining techniques for product recommendation. It employs AHP to evaluate the load (relative importance) of each RFM variable. The RFM value of each customer is normalized. The normalized RFM value is then multiplied by the relative importance of the RFM variable, $W_R, W_F, and W_M$, which are determined by the AHP.

$$Corr_{WRFM}(c_i,c_j) = \frac{\sum_{s\in V}\left(WRFM_{c_i,s}-\overline{WRFM}_{c_i}\right)\left(WRFM_{c_j,s}-\overline{WRFM}_{c_j}\right)}{\sqrt{\sum_{s\in V}\left(WRFM_{c_i,s}-\overline{WRFM}_{c_i}\right)^2\sum_{s\in V}\left(WRFM_{c_j,s}-\overline{WRFM}_{c_j}\right)^2}}$$

$\overline{WRFM}_{c_i}$ and $\overline{WRFM}_{c_j}$ are the average weighted-RFM (WRFM) value of customer c_i and c_j, respectively. The variable V denotes the set of RFM variables. The variables $WRFM_{c_i,s}$ and $WRFM_{c_j,s}$ indicate the weighted value R (F or M) of customer c_i and c_j, respectively. K-means clustering is employed to group customers with similar lifetime value or loyalty based on weighted RFM. Association rule based recommendation is then employed to recommend product items for each customer group.

2.3. Preference-based CF method

The preference-based CF method is similar to the WRFM-based CF method, except that clustering is performed based on purchase preferences.

2.4. Combining WRFM-based method and preference-based CF method

A hybrid method that groups customers by considering both the WRFM values and purchase preferences can improve the quality of recommendations. The WRFMCP method conducts association rule-based recommendations by extracting recommendation rules from customer groups clustered according to the weighted combination of the WRFM values and purchase preferences.

2.5. Hybrid Works

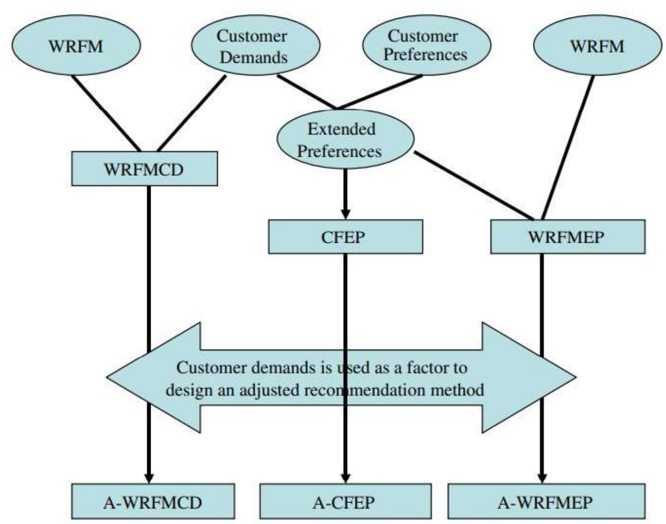

Figure 1: Flow chart for product recommendation.

Providing accurate recommendations under sparse data conditions (some preference ratings) is a primary challenge for collaborative filtering. It is difficult to find users with similar preferences if the user-rating matrix is too sparse, leading to sparsity problems for the CF method. A hybrid approach has been proposed to overcome the drawbacks of CF and

CBF methods. We combine content-based filtering and collaborative filtering to improve the accuracy of recommendation. Two such approaches, weighted model and meta-level model, were proposed that used different strategies to combine content-based and collaborative filtering. The weighted model used linear combinations of predictive results for collaborative and content-based filtering. The meta-level model employed a sequential combination of collaborative and content-based filtering, where the output generated by content-based filtering is used as input for collaborative filtering.

3 Proposed Recommendation Method

3.1. Customer Demands

Customer demands for each industry are used as valuable information in this work. These demands are determined by simple statistics that calculate the frequency count of each product item purchased by customers in each industry. If the frequency count of a product item purchased in an industry is greater than a given threshold h, then customers in such industries tend to have a demand for the item. The element r_{ij} of a customer-demand matrix CD represents whether the i^{th} customer tends to have a demand for the j^{th} product. The similarity of customer demands $Corr_{cd}$ among customers can also be measured by computing the Pearson correlation coefficient.

3.2. Customers' extended preferences

Limited to available content information, in which product features are not provided in the data set, customer demands are integrated with purchase preferences to reduce the sparsity of customer item matrix R. The element r_{ij} of the extended-preferences matrix EP represents whether the i^{th} customer had purchased or tend to have the demand for the j^{th} product.

3.3. Combining WRFM and customer demands

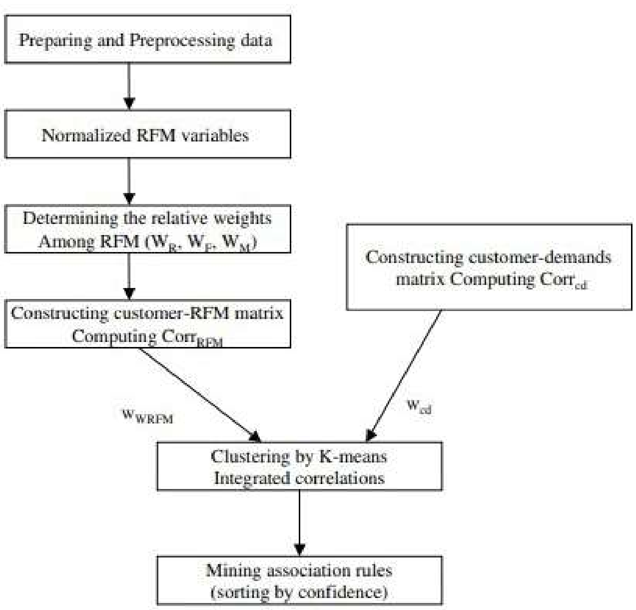

Figure 2: WRFMCD method for product recommendation.

The data set is pre-processed to extract customer transactions. Unreasonable records such as those of customers who have a non-zero amount of purchases but have never made any transactions are also removed. The WRFMCD method first establishes a customer-WRFM and a customer-demand matrix. RFM values of customer transactions are extracted to measure the customers' CLV. Then, the WRFM-based and the customer-demand correlation coefficients are computed using the Pearson correlation coefficient, respectively.

$$Corr_{WRFMCD}(c_i,c_j) = w_{WRFM} \times corr_{WRFM}(c_i,c_j) + w_{cd} \times corr_{cd}(c_i,c_j)$$

w_{WRFM} and w_{cd} represent the relative importance (weights) of the CLV elements and customer demands, respectively. If $w_{WRFM} = 0$, customer demands are used for recommendations. If $w_{WRFM} = 1$, the method becomes a WRFM-based CF method. Subsequently, K-means clustering is used to group customers with similar CLV and customer demands based on weighted correlation coefficients. Finally, the association rule mining approach is applied to extract recommendation rules from each group derived from K-means clustering. Candidate products are sorted by their associated confidence values, and the top-N highest-ranked candidate products are recommended to users. Customer demands can also be used as a factor to design an adjusted recommendation method, termed the adjusted WRFMCD (A-WRFMCD) method.

3.4. Combining WRFM and extended preferences

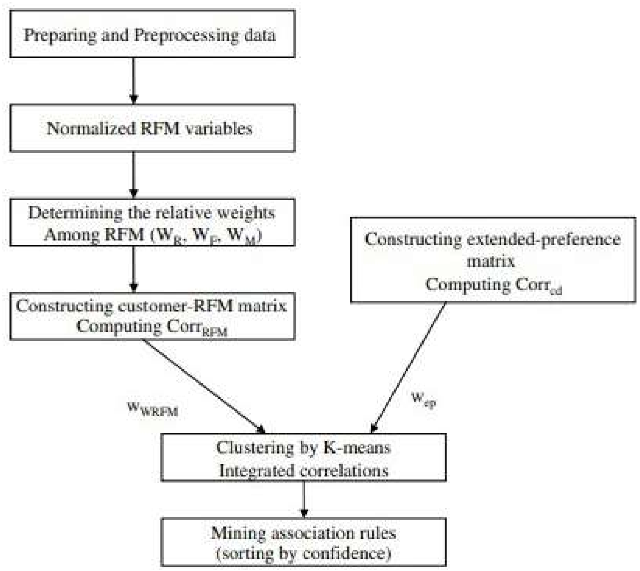

Figure 3: WRFMEP method for product recommendation.

This work proposes another hybrid method, namely the WRFMEP method, by which customers are clustered by integrating the dimensions of customer lifetime value and extended preferences. Customer RFM values are normalized and then multiplied by the relative importance of the RFM variables to adjust the importance of customer lifetime value and extended preferences in clustering.

$$Corr_{WRFMEP}(c_i,c_j) = w_{WRFM} \times corr_{WRFM}(c_i,c_j) + w_{ep} \times corr_{ep}(c_i,c_j)$$

The K-means technique uses integrated correlation coefficients to cluster customers. The association rule mining approach is then applied to extract recommendation rules from each group derived by K-means clustering. Candidate products are sorted by their associated confidence values, and the top-N highest-ranked candidate products are recommended to users. The adjusted WRFMEP (A-WRFMEP) method uses customer demands to adjust and re-rank candidate recommended products.

4 Pros:

- Handle fresh items in a much better way if we compared it with the basic collaborative method.
- We have improved the sparsity of the interaction-matrix if we compared it with the basic collaborative method.
- We can get benefits from more than one recommendation method.

- We have improved the accuracy.
- Extended preferences, derived by combining customer demands and purchase preferences, are useful for improved recommendation quality.

5 Cons:

- Computationally expensive.
- Size of the interaction-matrix. The dimension of the interaction-matrix is N $\times$ M. Where N is the number of users and M is the number of products.
- The sparsity of the interaction-matrix.

6 Result

The performance ranking of those methods with extended preferences is WRFMEP > CFEP > EP-based k-NN method; while the ranking of those methods without considering extended preferences is WRFMCP method > WRFM-based CF method > preference-based CF method > KNN-based method.

7 Future Work

Further studies are needed to evaluate the application of the proposed approach to other application domains. Second, the current work focused on product recommendation of retail transaction data that includes a binary option of shopping basket data; Customer choice is indicated as customers purchased the product; And zero, otherwise. Further investigation is needed to evaluate the effectiveness of the proposed methods for data sets with non-binary preference ratings. Finally, due to limitations of the available content information of the respective data sets, this work cannot address new and unseen objects. Further studies are required to validate the proposed methods on other essential matters that may support more material information.

8 bibliography:

[1] Recommendation systems,
`https://youtu.be/jhFFOmvPne4`

[2] The AI University, WRFM-based CF method
`https://youtu.be/fdUofaT8gUw`

[3] The AI University, WRFM-based CF method
`https://youtu.be/VE0m2FLoJCk`

[4] The AI University, WRFM-based CF method
`https://youtu.be/JCV7pPxJDdk`

[5] The AI University, WRFM-based CF method
`https://youtu.be/jBffqlNKMtI`

[6] The AI University, Customer Lifetime Value
`https://youtu.be/0q_uCTCLQuo`

[7] Maciej Kula, Basics of Hybrid method
`https://youtu.be/EgE0DUrYmo8`

[8] Rayan Gillard, Hybrid methods
`https://youtu.be/5VylUiV5zsc`

[9] Edureka!, Recommendation Rule set 1
`https://youtu.be/guVvtZ7ZClw`

[10] Augmented Startups, Recommendation Rule set 2
`https://youtu.be/WGlMlS_Yydk`