Chapter 2. Related Work

2.1 Customer lifetime value analysis and RFM evaluation

Customer lifetime value (CLV or LTV) is typically used to identify profitable customers and to develop strategies to target customers (Irvin, 1994). CLV is defined as the present value of future earning or profit of an individual customer (Berger and Nasr, 1998). There are a lot of researches on calculating customer value. Such as Pfeifer and Carraway (2000) proposed Markov Chain Models for modeling customer relationships. Berger and Nasr (1998) have proposed lifetime value calculation formula, shown as follows:

$\mathrm{CLV}_{\mathrm{i,t}} = \sum_{\mathrm{t}=0}^{\infty} \frac{\mathrm{Profit}_{\mathrm{i,t}}}{(1+d)^{\mathrm{t}}}$, Where CLV is the net present value of the future profit of customer i; d is a discount rate; $\mathrm{Profit}_{\mathrm{i,t}}$ is the function of customer profits according to time t.

Furthermore, measuring RFM is an important method for assessing customer lifetime value. Bult and Wansbeek (1995) defined the terms as: (1) R (Recency): period since the last purchase; a lower value corresponds to a higher probability of the customer's making a repeat purchase; (2) F (Frequency): number of purchases made within a certain period; higher frequency indicates greater loyalty; (3) M (Monetary): the money spent during a certain period; a higher value indicates that the company should focus more on that customer.

Hughes (1994) proposed a method for RFM scoring that involved using RFM data concerning to sort individuals into five customer groups. The latest purchase time of 20% is set to 5, and is inferred to other customers. Meanwhile, a score of 1 indicates that the most recent transaction was long time ago. Purchase frequency and monetary value are ranked using the same system. Finally, the RFM scores are obtained for each customer, with that of the best customer equaling 555 while that of the worst equals 111, and different marketing strategies are developed for different customers accordingly. Different marketing strategies could then be adopted for different customers. Stone (1995) suggested that different weights should be assigned to RFM variables depending on the characteristics of the industry. In analyzing the value of customers who used credit cards, he suggested placing the highest weighting on the Frequency, followed by the Recency, with the lowest weighting on the Monetary measure. However, he determined the RFM weightings subjectively, without employing a systematic approach.

2.2 Market segmentation

Clustering (Punj and Stewart, 1983) seeks to maximize variance among groups while minimizing variance within groups. Many clustering algorithms have been developed, including K-means, hierarchical, fuzzy c-means. The clustering method divided into hierarchical and nonhierarchical (Johnson and Wichern, 1992). Judging which of these approaches is best and how to determine when to apply it appropriately is difficult. Punj and Stwart (1983) compared the advantages, disadvantages, and outliers effecting on clustering to several other approaches, and the comparison result revealed that average Linkage and Wards were better. Punj and Stewart also noted that researchers used the average Linkage or Wards methods to identify the number of clusters, removed the outliers, and then clustered by nonhierarchical clustering approaches. Generally, K-means is one of the most widespread approaches.

K-means clustering (MacQueen, 1967) is a method commonly used to partition a set of data into groups. This scheme proceeds by selecting m initial cluster centers and then iteratively refining them. (1) Each instance d_i is assigned to its closest cluster center; (2) each cluster center C_j is updated to the mean of its constituent instances. The algorithm has converged when the assignment of instances to clusters no longer changes.

2.3 Recommender systems

A recommender Systems is a system that recommends items to users among a huge stream of available items, according to user's interests. A number of related prototypes have been developed for recommending items such as books (Mooney and Roy, 2000), web pages (Joachims et al., 1997; Lieberman, 1995), Usenet articles (Konstan et al., 1997), movies (Resnick et al., 1994), musics (Shardanand and Maes, 1995), and many more. Generally, there are four prevalent approaches to building recommender systems—content-based filtering (CBF), collaborative filtering (CF), hybrid works and association rules for product recommendation. These approaches are described as following sections.

2.3.1 Content-based filtering recommendation

2.3.1.1 Definition

The content-based filtering is based on the idea that if the user liked an item in the past then they are probably to like other similar items in the future. CBF recommender systems obtain item characteristics and compare them with user interest profiles to predict user preferences. Generally, item characterization may need various domain-specific features, each associated with their own feature extraction techniques. For instance, content-based photos recommender systems need to extract some image-based features while content-based music recommender systems extract audio-related features.

2.3.1.2 Applications

Content-based filtering (CBF) method provides recommendations by matching customer profiles (e.g. interests) with features of the content (e.g. product's attributes). Applied mostly is in textual domains, such as Krakatoa Chronicle system (Kamba et al., 1995). The Krakatoa Chronicle is a first personalized newspaper that creates a realistic rendering of a newspaper. In this system, user profile is a mapping from a set of keywords to weights (through TF-IDF similarity). Every word in the news read by the user is added to his/her profile and the weights set accordingly. Documents are recommended to users based on three parameters: the score that each article receives based on the user's profile, the average score received by each article over the community of users (the community scores), and size and composition of each article (e.g. the number of pictures). Based on these parameters, each user has accessed to a personalized newspaper, according to their interests.

2.3.1.3 Advantages and disadvantages

CBF presents two key advantages: (i) no first-rater problem, and (ii) no sparsity problem. The first advantage is because CBF recommends an item to a user if the user's profile and the text of the item share words in common. The second advantage is due to the fact that, in textual domains, for most items can be computed a similarity between its text and the user's profile.

However, CBF method suffers the limitations of not being able to provide serendipitous recommendations, and it cannot successfully analyze the content in some domains. First, CBF systems not able to provide serendipitous recommendation, this is due to the fact that techniques analyze the content of the texts, then recommending items with similar content, without between other subjects. For instance, if a text uses the word "car" and other text users the work "automobile", a technique might not consider these two texts similar. Second limitation is that current technology difficult for a computer to analyze such content likes video and audio streams, thus CBF systems is hard to recommend items to users.

2.3.2 Collaborative recommendation

2.3.2.1 Definition

Collaborative recommendation (or collaborative filtering) predicts user preferences for items in a word-of-month manner. That is, user preferences are predicated by considering the opinions (in the form of preference ratings) of other "like-minded" users. In particular, one can define a similarity measure between a pair of user preference ratings to define the like-mindedness between users (called memory-based methods in Breese et al., 1998). As preference ratings are used instead of domain-specific features, the applicability of collaborative recommender systems is more universal. For instance, if the system finds that you like computer books and at the same time are similar in taste to a group of users who like both computer books and science fictions, it will then recommend science fictions to you.

2.3.2.2 Typical KNN-based CF method

Collaborative filtering is a successful recommendation method, which has been widely used in various applications. A typical KNN-based collaborative filtering (CF) method (Resnick et al., 1994; Shardanand and Maes, 1995, Sarwar et al., 2000) 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. Notably, preferences generally are defined in terms of customer purchasing behavior, namely, purchased/non-purchased (binary choice) of shopping basket data, or taste, namely, preference rating on product items.

The typical KNN-based CF method is detailed as follows. Customer preferences, namely, customer purchase history, are represented as a customer-item matrix ***R*** such that, r_{ij} is one if the ith customer purchased the jth product; and is zero otherwise. The similarity of preferences among customers can be measured in various ways. A common method is to compute the Pearson correlation coefficient defined as Eq. 1:

$$corr_P(c_i,c_j)=\frac{\sum_{s\in I}(r_{c_i,s}-\bar{r}_{c_i})(r_{c_j,s}-\bar{r}_{c_j})}{\sqrt{\sum_{s\in I}(r_{c_i,s}-\bar{r}_{c_i})^2\sum_{s\in I}(r_{c_j,s}-\bar{r}_{c_j})^2}} \tag{1}$$

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 $r_{ci,s}$ and $r_{cj,s}$ indicate whether customers c_i and c_j purchased product item s. Customers are ranked by their similarity measures in relation to the target customer u, as determined using the Pearson correlation coefficient. The k most similar (highest ranked) customers are selected as the k-nearest neighbors of customer u. Finally, the top-N recommended products are determined from the k-nearest neighbors of u, as follows. The frequency count of products is calculated by scanning the purchase data of the k-nearest neighbors. 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 selected as the top-N recommendations.

2.3.2.3 Applications

Collaborative systems have been widely used in many areas, such as GroupLens system (Resnick et al., 1994) applied CF to recommend Usenet News. Ringo (Shardanand and Maes, 1995) recommends music album. The GroupLens system is a collaborative system to recommended Usenet Net News to users. Users explicitly rated the news in a 1-5 ratings scale and the system aggregates their votes and generates neighborhoods using the Pearson correlation coefficient. Recommendations are given as a weighted average among the neighbor's ratings. GroupLens is considered the first successful system that employs CF and the most cited works in this field. The Ringo system, developed at the Massachusetts Institute of Technology, used the same approach of GroupLens but through a 1-7 rating scale. Ringo also proposed a different coefficient, the constrained Pearson, to compute similarity. This coefficient has the same formula of the original Pearson but instead of using the average of the rating, it uses 4, which is the midpoint of its seven-point rating scale. Constrained Pearson performed better than its standard approach, but it reduced its coverage. In Ringo, Users explicitly enter their ratings to get recommendations of audio CDs.

2.3.2.4 Advantages and disadvantages

Collaborative filtering algorithms provide two key advantages to information filtering that are not provided by Content-based filtering (Balabanovic and Shoham, 1997; Herlocker et al., 1999): (i) independence of content; and (ii) the ability to provide serendipitous recommendations. First of all, CF systems can perform in domains where there is not much content associated with items. This system can suggest items to users based on the rating of items instead of the contents of the items. Second, CF can recommend items to users that they don't expect to receive, but are good recommendations (serendipity). This is due to the recommended items are relevant to the user, but do not contain content from the user's profile. For instance, users John and Mary have the same tastes about romance movies. If John rates very highly to a drama movie, then the drama movie might be a good recommendation to Mary.

Although CF has been successfully used in various domains, there still suffers from two problems (Balahanovic and Shoham, 1997; Claypool et al., 1999; Herlocker and Konstan et al., 1999): (i) the first-rater problem; (ii) the sparsity problem. For the first problem, that an item cannot be recommended until a user has rated it. The second problem is the sparsity problem occurs because in a real domain, a user is very likely to rate only a small percentage of the existing items, making it difficult to create neighborhoods due to the lack of overlap of tastes. In online retailers such as Amazon.com there are millions of books that a user could never possibly rate.

2.3.3 Hybrid Methods

2.3.3.1 Definition

Here we summarized the advantages and disadvantages of CBF and CF to Table 1. The goal of hybrid methods is to combine different techniques to mutually eliminate their drawbacks.

Table 1. Advantages and disadvantages of CBF and CF methods

	Content-based approach (CBF)	Collaborative filtering approach (CF)
Feature	Extract items characteristics and compare them with user interest profiles	Predicts user preferences for items in a word-of-mouth manner
Advantages	No first-rater problem and no sparsity problem	Independence of content and it can provide serendipitous recommendations
Disadvantages	Cannot successfully analyze the content and have the over-specialization problem	The first-rater problem and the sparsity problem

As mentioned previous, in practice, it's hard to require individual users to provide too much preference ratings. In the other hand, most users do not rate most items and hence the user ratings matrix is typically very sparse. Therefore, the probability of finding a set of users with similar ratings is usually low. That's the sparsity problem of CF method. Hybrid works are usually proposed to overcome drawbacks of CF and CBF methods to improve the recommendation accuracy.

2.3.3.2 Methods

Different filtering systems can be combined in many ways. Li and Kim (2003) summarized two groups of hybrid systems, which combine content-based and collaborative filters together.

One group is the linear combination of results of collaborative and content-based filtering. The method used by this group is described by Claypool (1999) and is applied to recommend news in an online newspaper. It uses an adaptive weighted average to combine the predictions of content-based and the collaborative filtering. The users rate items explicitly. Every time as the number of users accessing an item increases, the weight of the collaborative component tends to increase. However, weights of collaborative and content-based components are difficult to decide. Following Burke's taxonomy (2002), such kinds of hybrid systems are named a weighted model.

The other group is the sequential combination of collaborative and content-based filtering. The sequential combination is based on measuring the similarity between the user and product profiles (features of product items) for products not yet rated by the users. Herein the user profile is composed of the user preferences of each product features (describes the user's interests). The similarity is then used to predict ratings of unrated products. This process aims to convert a sparse user ratings matrix into a full ratings matrix. Collaborative filtering then can use the denser matrix to provide recommendations. For example, Melville et al. (2002) presented a hybrid approach Content-Boosted Collaborative Filtering (CBCF) to make movie recommendations. They create a pseudo user-ratings vector for every user *u* in the database. A pseudo user-ratings vector contains the user's actual ratings and content-based predictions for the unrated items. Then performing collaborative filtering based on this dense matrix. Following Burke's taxonomy (2002), the CBCF is the meta-level hybrid model, where the model generated by one technique is used as the input of another technique.

2.4 Association rules for product recommendation

Association rule mining (Agrawal et al. 1993; Srikant and Agrawal, 1995; Yun et al., 2003) is a widely used data mining technique to generate recommendations in recommender systems. Accordingly, this work employs association rule mining to discover the relationships among product items based on patterns of co-occurrence across customer transactions.

2.4.1 Association rule mining

Association rule mining aims to find an association between two sets of products in the transaction database. Agrawal et al. (1993) formalized the problem of finding association rules that satisfy minimum support and minimum confidence requirements. Let I be a set of product items and D be a set of transactions, each of which includes a set of products that are purchased together. An association rule is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \Phi$. X is the antecedent and Y is the consequent of the rule herein. Two measures, support and confidence, are used to indicate the quality of an association rule. The support of a rule is the percentage of transactions that contain both X and Y, whereas the confidence of a rule is the fraction of transactions that contain X, that also contain Y. An example of an association rule in the basket market analysis domain is: "90% of transactions that contain bread and butter also contain milk; 30% of all transactions contain the three of them". Herein, X = {bread, butter}, Y = {milk}, 90% is called the confidence of the rule, and 30% the support of the rule.

The support of an association rule indicates how frequently that rule applies to the data. Higher support of a rule corresponds to a stronger correlation between the product items. The confidence is a measure of the reliability of an association rule. The higher the confidence of a rule corresponds to a more significant correlation between product items. The *apriori* algorithm (Agrawal et al. 1993; Agrawal and Srikant, 1994) is typically used to find association rules by discovering frequent *itemsets* (sets of product items). An *itemset* is considered to be frequent if the support of that *itemset* exceeds a user-specified minimum support. Moreover, association rules that meet a user-specified minimum confidence, can be generated from the frequent *itemsets*.

2.4.2 Association rule based recommendation

Sarwar et al. (2000) described the method of association rule-based recommendation as follows. For each customer, a customer transaction is created to record all the products previously purchased by the customer. The association rule mining algorithm is then applied to find all the recommendation rules that satisfy the given minimum support and minimum confidence constraints. The top-N products to be recommended to a customer u, is then determined as follows. Let X_u be the set of products previously purchased by customer u. The method first finds all the recommendation rules $X \Rightarrow Y$, for which $X \subseteq X_u$. Then, for each extracted recommendation rule, all the products in Y that have not yet been purchased by customer u are candidate products for recommendation. Each candidate product is associated with the confidence of the corresponding recommendation rule. The candidate products are sorted by associated confidence value, where the N highest ranked candidate products are selected as the recommendation set.

2.5 Evaluation metrics

Two metrics, precision and recall, are commonly used to measure the quality of a recommendation. These are also used measures in information retrieval (Salton and McGill 1983; van Rijsbergen, 1979). Product items can be classified into products that customers are interested in purchasing, and those that they are not interested in purchasing. A recommendation method may recommend interesting or uninteresting products. The recall-metric indicated the effectiveness of a method for locating interesting products. The precision-metric represented the extent to which the product items recommended by a method really are interesting to customers.

Recall is the fraction of interesting product items that can be located.

$$Recall = \frac{\text{number of correctly recommended items}}{\text{number of interesting items}}$$

Precision is the fraction of recommended products (predicted to be interesting) that are really found to be interesting.

$$Precision = \frac{\text{number of correctly recommended items}}{\text{number of recommended items}}$$

Items interesting to customer u were those products purchased by u in the test set. Correctly recommended items were those that match interesting items. However, increasing the number of recommended items tended to reduce the precision and increase the recall. An F1-metric (van Rijsbergen, 1979) could be used to balance the trade-off between precision and recall. F1 metric assigned equal weight to precision and recall and was given by,

$$F1 = \frac{2 \times recall \times precision}{recall + precision}$$

Each metric was computed for each customer, and the average value computed for each cluster, as well as the overall average (over all customers) as measures of the quality of the recommendation.

2.6 AHP approach

The Analytic Hierarchy Process (AHP) is a mathematical technique for multi-criteria decision making (Saaty, 1980; 1994). It is used to determine the relative critically weighting of indicators. The three main steps of the AHP are described as follows.

Step1: Perform pairwise comparisons. The aim of this step is to specify pairwise comparisons by decision makers. The comparisons of any two criteria A and B is made using questions of this type, how much A is more important than B.

Generally, the decision makers most often use 9-point scale. The scale is explained in Table 2.

Table 2. Relative degree of importance for pairwise comparisons

Comparative importance	Description	Explanation
1	Equally importance	Two activities contribute equally to the objective
2 3	Intermediate between equal and weak Weak importance of one over another	Experience and judgment slightly favor one activity over another
4 5	Intermediate between weak and strong Essential or strong importance	Experience and judgment strongly favor one activity over another
6 7	Intermediate between strong and demonstrated Demonstrated importance	An activity is strongly favored and its dominance is demonstrated in practice
8 9	Intermediate between demonstrated and absolute Absolute or extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation

Step 2: Assess the consistency of pairwise judgments. Evaluators may make inconsistent judgments when making pairwise comparisons, such as A is more important than B, B is more important than C. And thus, A should be more important than C. The consistency of the judgment matrix can be determined by a measure called the consistency ratio (CR), defined as:

$$CR=CI/RI$$

where CI is called the consistency index and RI, the random index. If CR of the matrix is higher, it means that the input judgments are not consistent, and therefore are not reliable. Generally, a consistency ratio of 0.1 or less is considered acceptable. If the value exceed this, then the pariwise judgments may be revised before the weighting of indicators are computed.

Step3: Computing the relative weights. This determines the weight of each decision element was computed based on the pairwise comparison. There are different techniques to find the best estimation methods, such as the Eigenvalue method, mean transformation or row geometric mean.