Multi-criteria Recommendation Based on Trust

Completed Research Paper

Sung-Shun Weng

Chien-Hung Yeh

Abstract

With the quick rise of E-Commerce, personalized recommender systems have been created, which not only stimulate the sales of products and services but also increase customers' loyalty. Collaborative filtering has been the most common and effective technique applied to recommendation systems. In this study, we applied multi-criteria ratings in movie preference collection for building users' profiles of different aspects profile and we used regression coefficient, which was derived from multiple regression analysis, for each criterion preference level. In order to improve each criterion prediction quality and alleviate the sparsity problem of Collaborative filtering, we firstly combine users' preference levels and trust-values as trust-weight in different criterions and then set trust-weight threshold to filter users and to find recommendation members. In the final aggregation experiments, we observed that trust-based filtering method also carried out a lower MAE in overall rating prediction and the F1 value has a better recommendation performance. Finally, the approach we proposed can improve the recommendation quality in the multi-criteria recommendation environment.

Keywords: Multi-criteria, Trust-based Recommendation, Multiple Regression Analysis

Introduction

In traditional recommendation systems, single rating is mostly used to judge the degree users like the items. However, single rating cannot tell dimensions influencing users' preferences or dislikes. Adomavicius and Kwon (2007) pointed out in the research on multi-criteria recommendations that even if different users give the same rating on some item, they might have different criteria on preferences. Thus preferences of different dimensions are neglected, which might make a group of people with similar preference criteria not found. Recommendation systems with collaborative filtering approaches have their inherent limitations. For example, owing to sparse data, similar rating items of users are not sufficient, lowering similarity and reliability. Jeong et al. (2009) replaced similarity among users with trust on ratings of users when predicting ratings, further lowering the influence of inherent limitations for collaborative filtering. In the multi-criteria environment that needs feedback of more ratings, the aim of this research is to explore how to reduce the interference of malicious ratings and sparsity of rating, then to enhance accuracy of multi-criteria recommendation. Therefore, this research will get ratings of users' multi-criteria preferences on movies, evaluate factors influencing users' overall ratings with multi-criteria on movies, and take advantage of trusted

members with high-reference-value ratings of multi-criteria to enhance accuracy of prediction, expecting to minimize the error of predicting ratings with multi-criteria and improve overall recommendation efficiency through predicting ratings of trusted members with multi-criteria.
This study is expected to use multi-criteria ratings of users to obtain multi-criteria preference weights by linear regression and predict the overall ratings. By using users' preference criteria on different dimensions of the items as the preference factors of users, we can make up the single ratings that neglect diversity of customers' multi-criteria preferences. This research is to predict ratings through trusted members with multi-criteria and explore the accuracy of prediction on overall ratings of movies. The following are the purposes for this research:

(1) Use trust on users' ratings as the base to judge and predict a group under multi-criteria.
(2) Predict through trusted groups to lower the influences on prediction made by data sparsity.
(3) Use linear regression coefficients as the aggregate function for the overall rating prediction and explore the performance of multi-criteria recommendation.

Literature Review

Trust-based Recommendation

O'Donovan and Smyth (2005) pointed that if nearest neighbors are only recommended to selected products through rating similarity, it is not reliable. For example, you will consult a group of friends with similar preferences when you go to see a movie. If someone among them always dislikes some particular movie and you take his opinion into consideration, it will not be reliable. Besides similarity, trustworthy users recommended in the past should also be considered, in order to enhance the recommendation quality.

As to the disadvantage of similarity-based collaborative filtering, many researches have started to find users with prediction reference value through trust-based recommendation and make them as the criteria of group selection (Massa et al., 2004; O'Donovan et al., 2005; Victor et al., 2009). Trust means trustworthiness, reputation, reliability and credibility among members (Kwon et al., 2009). Our purpose is to find out not only members with similarity, but those with high credibility on target customers, as recommended members.

Jeong et al. (2009) brought up user credit-based collaborative filtering (UCCF) to lower the influence of data sparsity on accuracy of similarity calculation, replacing similarity among users with their rating credibility on the rated items and making rating prediction on target users' unknown items through rating credibility. To evaluate users' rating credibility and predict ratings, there are three steps as follows:

(1) Producing Nominal Rating
Form representative rating vectors with the average and median of all users' ratings on the items, or ratings with high probability density.

(2) Evaluating Users' Credibility
Form vectors of personal ratings with the set of items rated by users. Use the variability between vectors of personal ratings and representative rating vectors to show personal credibility. The higher the credibility is, the higher it correlates with overall ratings of all members. Replace traditional similarity among users with rating credibility of users.

(3) Rating Prediction and Recommendation Making
For items not rated by aimed users, get evaluation rates through aggregation function. Then recommend items ranking Top-N to aimed users.

O'Donovan and Smyth (2005) proposed a trust-based filtering recommendation system. Their works made members with rating credibility over the threshold as the selecting criteria of prediction group, to make up the reliability of similarity-based filtering recommendation system. The group used to offer prediction is represented as Producer (p), target users are shown as Consumers (c). The error between p(i), producer' prediction value of i (consumers' common rating items), and c(i), consumer's actual rating, is used to judge whether the prediction is reliable (Formula 1).

$$Correct(i, p, c) \Leftrightarrow |p(i) - c(i)| < \varepsilon \qquad \text{(Formula 1)}$$

The group participating in recommendation in the process of recommendation is Producer (p) offering trust value. The degree of prediction accuracy in the process of recommendation made by each producer is the base of trust value of the producer. For example, Producer b makes rating prediction on i1 rated by Consumer c. When the prediction rating by Producer b is smaller than the error ε, it means that Producer b's prediction on Consumer c's preference on the items i1 is correct. Otherwise, if the prediction rating is bigger than the error ε, it means that the prediction fails. However, personal prediction does not take group prediction into consideration, but just considers that single Producer b predicts particular Consumer c's preference rating on the item i. The higher the ratio of accurate prediction is, the higher the trust value predicted by Producer on the item is.

In the research of O'Donovan and Smyth (2005), they brought up two different types of credibility measurement, profile level and item level. Profile level means that it makes the ratio of reliable sets among recommended sets by Producer p as trust value of profile level. Item level means that make the ratio of reliable sets among recommended sets on some particular item i by Producer p as trust value of item level. In measurement of trust of item level, groups similar to the items are also added. Therefore, when calculating Producer's trust value on the particular item, trust prediction on similar items is also taken into consideration (Formula 2) with a better performance on prediction.

$$Trust^{I}(p,i) = \frac{\left|\{(u_k, i_k) \in CorrectSet(u') : i_k = i \text{ and its neighbors}\}\right|}{\left|\{(u_k, i_k) \in \mathrm{Re}\,cSet(u') : i_k = i \text{ and its neighbors}\}\right|} \text{(Formula 2)}$$

Kwon et al. (2009) measured three different attributes of users' rating, expertise, trustworthiness and similarity, as the base to select reliable members. $EXPERTISE_u$ means u's trust value on item rating (O'Donovan and Smyth, 2005). $TRUST_u$ measures the similarity between u's personal rating on items and the average of all groups' ratings on all items, meaning the degree of similarity with group rating trend. $SIMILARITY_{a,u}$ means similarity on common rated items between users and target users. Find weights of all trust attributes ($K_{E,a}$、 $K_{T,a}$、 $K_{S,a}$) through linear combining, thus to evaluate users' overall trust value $V_{a,u}$ (Formula 3) on target users and find TOP-N trusted members as the prediction group.

$$V_{a,u} = K_{E,a} \quad EXPERTISE_u + K_{T,a} \quad TRUST_u + K_{S,a} \quad SIMILARITY_{a,u} + \varepsilon_a \text{ (Formula 3)}$$

Multi-criteria Recommendation

In the traditional collaborative filtering recommendation system, it can be known about customers' preferences on items when customers give ratings or hidden inferences. The system keeps customers' preferences on items via two-dimensional matrix and shows preference rating prediction on products via R: Users×Items→ R_0. R is the prediction function and R_0 is the set of items customers have not rated. However, on the application of traditional personal recommendation, multi-criteria are seldom taken into consideration in the process of recommendation. Customers just rate their preferences on different items with their single degree of preferences. Therefore, it is to be improved by using multi-criteria for recommendation. Adomavicius and Kwon (2007) pointed that multi-criteria rated data give users a chance to interpret their preferences on items through different dimensions. Through extra multi-criteria rating data, the accuracy of recommendation will be improved. Multi-criteria preference rating matrix can be shown with R: Users × Items $\rightarrow R_O \times R_1 \times \ldots \times R_k$. R_0 is the traditional single preference rating. R_i is the probable preference rating under different criteria, i (i = 1,…,k). Single preference rating only shows the degree of preference of customers. In contrast, multi-criteria rating can show why customers prefer one item. The approach learns more about preference reasons and then finds members with similar preferences. For example, the "Online Dining Guide Website", Zagat, integrates ratings under three different criteria, food, decoration and service. Yahoo!Movies not only gives ordinary overall ratings on movies, but also encourages users to rate under another four criteria, i.e., stories, actors' performances, directors and visual effect. With different rating criteria, it can be known about major factors why customers like the restaurant or the movie and interpretations with different dimensions. Adomavicius and Kwon (2007) categorized multi-criteria recommendation into two approaches, similarity-based approach and aggregation-function-based approach.

(1) Similarity-based Approach

If multi-criteria are the same as single criterion, similarity among users still needs to be calculated. Calculating approaches can be further divided into average similarity and worst case similarity. As to average similarity, it separately calculates similarity among users under k criteria. Then it calculates the average of similarity under k criteria and the average shall be the similarity between two users (Formula 4). As to worst case similarity, it makes the worse part among similarity under k criteria as the similarity between two users (Formula 5).

$$sim_{avg}(u,u') = \frac{1}{k+1}\sum_{i=0}^{k} sim_i(u,u') \qquad \text{(Formula 4)}$$

$$sim_{\min}(u,u') = \min_{i=0,\ldots,k} sim_i(u,u') \qquad \text{(Formula 5)}$$

$$sim(u,u') = \frac{1}{1+d_{user}(u,u')} \qquad \text{(Formula 6)}$$

Besides, it makes multidimensional distance as the distance of vectors under k criteria among users. After conducting distance calculation, the distance is transferred into similarity from 0 to 1 (Formula 6). Chen (2007) adopted the similarity-based approach, applied it to the field of games and achieved the purpose of multi-criteria recommendation. First, the similarity modules make the average of similarity under certain kinds of criteria among groups as the base of similarity and predict on similar groups' ratings on target users under certain kinds of criteria. The experience-direction learning modules make the similarity among users' ratings under certain kinds of criteria and the overall rating as preference weight. The recommendation modules conduct aggregating prediction on the overall rating via target users' weights under certain kinds of criteria. Finally, it recommends the Top-N games with prediction ratings higher than the threshold.

(2) Aggregation-function-based Approach

The approach makes rating prediction via the aggregation-function-based approach. It is mainly divided into three parts separately, i.e., multi-criteria rating prediction, aggregation rating function generation and overall rating prediction.

(a) Multi-criteria Rating Prediction

Target users show their probable preferences on the overall rating predicted on to-be-rated movies. To consider the probable influence on overall rating brought by users' ratings on preferred criteria, the integrity of collecting multi-criteria rating data becomes the key to analyze the overall rating. After collecting users' ratings on different items under k criteria, the multi-criteria rating matrix is formed, R: Users × Items $\rightarrow X_1 \times \ldots \times X_k$. To predict the multi-criteria ratings, the approach divides the multi-criteria rating matrix into k independent rating matrices, R: Users × Items $\rightarrow X_i$ (where $i = 1,\ldots, k$). It then predicts all unknown ratings via aggregation calculation of the single rating recommendation system.

(b) Aggregation Rating Function Generation

To find out the degree of influence on the overall rating brought by users' multi-criteria rating preferences, it predicts the influence on rating of items' overall preferences brought by multi-criteria rating, aggregation function can be divided into three directions, users' personal base, item base and overall evaluation. As to personal base and item base, for particular users or items, it separately finds their multi-criteria influence weight on overall rating. As to overall evaluation, either for users or for items, the same weight is used to show the influence on overall rating brought by multi-criteria. Common aggregation function weight generation approaches are as follows: (i) domain expertise: based on previous experiences or professional knowledge, experts within the field produce suitable aggregation function for prediction. For example, prediction on overall rating is the average of all criteria, (ii) statistical techniques: it makes analysis via statistics, such as linear or non-linear regression analyzing techniques. Through linear combination, it predicts overall rating via multi-criteria rating. Factors of multi-criteria are $\hat{\beta}_0$ 、 $\hat{\beta}_1$ 、 $\hat{\beta}_2$ 、... 、 $\hat{\beta}_k$, meaning multi-criteria weight on overall rating. It can be used as the aggregation function to make subsequent prediction on common overall rating, and (iii) machine learning techniques: taking neural network technique as example, it repeatedly trains neural network via hidden layers. After then, it uses the trained neural network to determine the weights of users' multi-criteria preferences.

(c) Overall Rating Prediction
After getting multi-criteria prediction ratings on users' preferences and multi-criteria multiple regression coefficients influencing overall rating weight from the first two steps, it predicts users' overall ratings of probable preferences on to-be-rated movies via aggregation function.

Traditional collaborative filtering recommendation calculates similar groups based on all members' single ratings on items and does not measure the differences of users' preferences on items' multi-criteria. This leads to a problem that similar groups might not have the same preference criteria. Therefore, besides customer's single preference rating, multi-criteria rating's probable influence on similar groups should also be considered.

Research Methodology

The research framework of this study, considering from the perspective of multi-criteria rating, consists of users' rating trust mechanism and the process of prediction and recommendation. Standard regression coefficients are used to measure users' attention on multi-criteria. Un-rated movies' multi-criteria ratings are predicted based on the group trust. Group trust in the research indicates it not only is the credibility of users' rating on prediction, but also the attention on multi rating criteria. Finally we use regression prediction module to calculate the overall preferences on movies and recommend movies with prediction ratings over the threshold.

Module functions

(1) Multiple regression analysis

Multiple regression analysis indicates users' attention on multi-criteria preferences. It asks users to actually rate movies under different dimensions and gets multi-criteria preference features after recording users' ratings on multi-criteria. Adomavicius and Kwon (2007) brought up the linear regression approach, regarding users' multi-criteria ratings, X_1 、... 、 X_n , as independent variables and regarding to-be-measured multi-criteria common overall rating, *Y*, as the dependent variable. The ratings, based on Likert Five-point Scale, are from 1 to 5, and show users' judgment on multi-criteria preferences. We get multi-criteria regression function as aggregation function after using the SPSS software to analyze multiple regressions. Regression factors show multi-criteria influences on the overall preferences. This module can be divided into two parts, i.e., rating aggregation function and multi-criteria preference vector.

(a) Rating aggregation function
To find multi-criteria rating preference's influence on the overall rating, this study adopts the linear regression technique in statistics as aggregation function to predict the overall rating. First, we let users' common overall rating, Y, as the dependent variable and let the k multi-criteria ratings, X_1,

$X_2, \ldots, X_k$, as independent variables. We use ordinary least squares (OLS) to train linear multiple regression modules and get regression factors. Multi-criteria factors are separately as $\hat{\beta}_1$, $\hat{\beta}_2$, ..., $\hat{\beta}_k$ and C, showing multi-criteria weights on overall rating and the constant C. They are used for the aggregation function f_u, which is the following prediction of the overall rating on movies by the target user u.

(b) Multi-criteria preference vector

Judging from regression factors got from multiple regression analysis, we can see the influence of users' multi-criteria ratings on the overall ratings of the movies. We transfer multi-criteria preference vectors made up of k criteria standardized regression factors into users' multi-criteria preference vector (MPV), shown in MPV(u)=($\hat{\beta}_1$, $\hat{\beta}_2$,…, $\hat{\beta}_k$), meaning users' different rating criteria on movies. Factor $\hat{\beta}_k$ means probable influence on users' overall preference and MPV(u) is the list of users' preference vectors.

(2) Multi-criteria rating evaluation module

To predict users' preference ratings on to-be-rated movies, this study uses the trust-based approach as the method of rating prediction, in order to find users or ratings with higher trust value as the base of multi-criteria preference prediction and avoid prediction errors caused by low trust value or malicious inaccurate ratings in the process of rating prediction. Before predicting multi-criteria ratings on users' to-be-rated movies, we first divide multi-criteria rating matrix into single criterion rating matrixes. Then we separately predict unknown multi-criteria ratings with the single criterion rating matrix.

(a) Calculation of the trust value of rating

Users' trust-based prediction uses the error among similar users' prediction values on items as the base to set up trust value. The more users with accurate ratings are predicted, the higher the trust value is. Under k criteria, each user has his/her own item prediction trust values on different items. This study refers to trust-based recommendation brought up by O'Donovan and Smyth (2005) to predict target users' unknown ratings. The steps to set up trust values are as follows:

(i) Calculation of similarity

To measure each user's influence weights on overall preference under different criteria, after getting MPV among users, we calculate similarity among users via Cosine Measure (Formula 7). We calculate similarity among groups via multi-criteria characteristics vectors among users in order to lower unreliability of similarity calculation caused by lack of rated items owing to sparse quantities.

$$sim_{MPV}(u,u') = \cos(MPV(u), MPV(u')) = \frac{\sum_{i=1}^{k} u(\hat{\beta}_i)u'(\hat{\beta}_i)}{\sqrt{\sum_{i=1}^{k} u(\hat{\beta}_i)^2}\sqrt{\sum_{i=1}^{k} u'(\hat{\beta}_i)^2}} \quad \text{(Formula 7)}$$

(ii) Prediction of personal rating

Before calculating users' trust values, members' ratings on common rated items should have to be predicted. We judge whether the prediction is correct by checking whether prediction error is within the error limits. As to common rated items, i, $u'(i)$ means that u' rates on i. We predict the other member, u, via $sim_{MPV}(u,u')$. Prediction rating is shown as $P_{u'}u(i)$ in Formula 8.

$$P_{u'}u(i) = \overline{u} + (u'(i) - \overline{u}')sim_{MPV}(u,u') \quad \text{(Formula 8)}$$

(iii) Calculation of item trust value

When the error between prediction rating of u', $P_{u'}u(i)$, and common member u's actual rating, $u(i)$, is smaller than ε, it is regarded as correct prediction, represented as $CorrectSet(u')$ in Formula 9. The set on i predicted by u' is shown as $\mathrm{Re}\,cSet(u')$. The set of predictions that their errors are smaller than error ε is represented as $CorrectSet(u')$. The more values that $CorrectSet(u')$ has, the more accurate ratio is, meaning that the bigger trust value of u' on i's rating shown in Formula 10.

$$Correct_k(i,u',u) \Leftrightarrow \left|P_{u'}u(i) - u(i)\right| < \varepsilon \qquad \text{(Formula 9)}$$

$$Trust_k(u',i) = \frac{\left|\{(u_k, i_k) \in CorrectSet(u') : i_k = i\}\right|}{\left|\{(u_k, i_k) \in \mathrm{Re}\,cSet(u') : i_k = i\}\right|} \qquad \text{(Formula 10)}$$

After finishing calculating users' trust values on the movies, we record users' trust values on the criteria (C1, C2,…, Ck) into the trust value matrix. The bigger the trust value is, the more credible users' ratings on the items are and the higher the reference value of prediction is, thus lowering malicious rating's influence on prediction.

(b) Calculation of trust weight

In multi-criteria rating environment, each member has different degree of influence on overall rating under different criteria. If we just take rating trust value as the judgment on trust groups, we might find false reference groups. Therefore, besides using users' rating trust values as the judgment on trust groups, users' preference factors, $\hat{\beta}_k$, on different rating criteria are also regarded as the bases to judge the trust groups. Therefore, trust weight means each rating member's trust weight on different items under multi-criteria. For trust weight calculation, this study combines similar trust values of u' on predicted item i, represented as $Trust_k(u',i)$, and the preference factors under k criteria, $\hat{\beta}_k$, and then transfers them into the trust weight $w_k(u',i)$, shown in Formula 11. The purpose of combining the two values into the weight is to lower the errors of judgment on trust members when only evaluating a single value.

$$w_k(u',i) = \frac{2(\hat{\beta}_k)(trust_k(u',i))}{\hat{\beta}_k + trust_k(u',i)} \qquad \text{(Formula 11)}$$

(c) Prediction of multi-criteria rating

To avoid huge diversity of multi-criteria preference among users, multi-criteria preference similarity threshold is set as the original measurement on selection of the predicted groups. The similarity of multi-criteria preference vectors between u' and target user, u, should be bigger than the threshold, $T_{sim} = 0.5$, and then they can be regarded as similar groups on multi-criteria preferences. Besides, as to multi-criteria rating prediction, this study refers to trust-based filtering prediction approach brought up by O'Donovan & Smyth (2005) to select trust groups. The trust weight obtained through the above steps (Formula 11) not only means reliability of members' ratings on the item, but also the criterion preference's influence on overall rating. The threshold of trust weight is set and represented as T_{Trust_Weight}. Members with weights over the threshold are trust groups to be predicted on ratings (Formula 12). Through trust group weight, we predict the k criteria rating of the target user, u, on the un-rated movie, i, via Formula 13.

$$U'^{T}(i) = \{u'_k \in U'_k(i) : w_k(u'i) > T_{Trust\_Weight} \text{ and } sim_{MPV}(u,u') > 0.5\} \qquad \text{(Formula 12)}$$

$$u_k(i) = \overline{u}_k + \frac{\sum_{u' \in U'(i)} (u'_k(i) - \overline{u}'_k) \times w_k(u', i)}{\sum_{u' \in U'(i)} w_k(u', i)} \qquad \text{(Formula 13)}$$

(3) Recommendation module

In the phase of multi-criteria rating evaluation module, we predict users' k criteria ratings ($X_1^{'}, X_2^{'}, \ldots X_k^{'}$) on to-be-rated movies via group trust-based approach. Through multiple regression analysis, we get the aggregation function, f, meaning users' prediction on overall rating. Finally we use the aggregation function to predict users' overall ratings on probable preferring to-be-rated movies. This study recommends movies with ratings equal or greater than 4.

Experiments and Result Analysis

Method of evaluation and analysis

In order to evaluate the predicted quality of movie recommendation, the mean absolute error (MAE) and F1 value that are commonly used by recommendation systems are adopted in this study.

(1) Mean absolute error

In order to verify the difference of the predicted quality of movie recommendation, and use it as a reference for selecting parameters with better predictive ability, this study uses the mean absolute error frequently used by recommendation systems as the predictive indicator of measurement, see formula 14, where r_i is the real rating of item i, p_i is the predicted rating of item i, and MAE is the average value of errors between the actual rating of items and the predicted rating of items, which is used as the measurement value. When the MAE is lower, it means the quality predicted by the system is higher.

$$MAE = \frac{\sum_{i=1}^{N} |r_i - p_i|}{N} \qquad \text{(Formula 14)}$$

(2) Experiment assessment of recommendation performance

In the study of recommendation system, indicators such as precision, recall and F1 are often used for assessing the recommendation result (Herlocker et al., 2004; Huang et al., 2004; Sarwar et al., 2000). Precision refers to the proportion highly preferred by the user in the recommended items, and recall refers to the recommended proportion actually obtained from the items highly preferred by all users. Precision is often inversely proportional to recall, so the indicator, F1, combining precision and recall is used for avoiding deviation produced by a single assessment indicator. This experiment will carry out recommendation assessment for all test data sets, so movies relevant and irrelevant to high interest, and selected and not selected movies must be defined in the test data. The classification is listed in table 1.

Table 1. Classification of recommended items

	Selected	Not Selected	Total
Relevant	N_{rs}	N_{rn}	N_r
Irrelevant	N_{is}	N_{in}	N_i
Total	N_s	N_n	N

$$\Pr ecision = \frac{N_{rs}}{N_S} \qquad \text{(Formula 15)}$$

$$\operatorname{Re} call = \frac{N_{rs}}{N_r} \qquad \text{(Formula 16)}$$

$$F1 = \frac{2 \times \operatorname{Re} call \times \Pr ecision}{\operatorname{Re} call + \Pr ecision} \qquad \text{(Formula 17)}$$

In this study, movies with actual rating equal to and higher than 4 in the test data set are viewed as movies (Nr) related to user high interest. The aggregate function will predict movies with rating equal to and higher than 4, and will view them as the recommended movies (Ns). The recommendation precision is the proportion of the actually and accurately predicted movies (Nrs) to all recommended movies (Ns), expressed in the formula 15, recall is the proportion of the actually and accurately predicted movies (Nrs) to movies (Nr) related to the high interest of all users, expressed in formula 16. And F1 (see formula 17) is used for assessing the recommendation effectiveness.

Experiment environment and data source

Our movie data and introductions are from the movie information portal website "True Movie" (http://www.truemovie.com/) and KingNet (http://www.kingnet.com.tw/), as well as movie preview links of YouTube. They are used for providing users with review of movies. These websites supply complete movie reference information to help users to assess movies. We collected totally 130 movies from these movie platforms as the experiment basis from January, 2007 to March, 2013. Users can rate and collect movies according to their impression after watching them. In addition to the general overall rating, in order to collect user's preferences to different aspects of movies, the movie rating in this experiment is based on the Yahoo!Movies website and movie comment blog, "Donald Kwan Movies" , as references. The criteria such as "screenplay/story", "acting", "cast" and "sound & visual effect" are the four criteria used for rating movies (See table 2). Users rate the impressed movies or favored movies according to the 5 rating criteria after reading the guidelines of rating criteria. The rating ranges from 1 to 5, which respectively represents "dislike", "not bad", "ordinary", "like" and "like very much".

Table 2. Rating criteria description

Grade criteria	Description
Like as a whole	Rating based on like as a whole and impression.
Screenplay & Story	Total epitasis, smooth of story flow, dialogue design, and message and meaning produced by story.
Acting	Degree of impression left by acting of actors and actresses, or acting quality of actors or actresses.
Cast	Cast, scale or fame of director and actors and actresses.

Sound & Visual Effect	Action design, animation technique, set design and story supplementary, poster, background music, and sound effect.

Current movie communities mostly use movie introduction and comments as the main information, and largely use a single rating as the feedback and collection of customers as for user rating. In this study, in order to verify if users will prefer movies based on different criteria, and if users will have an influence on the preference of movie as a whole, a multi-criteria movie rating platform is established for collecting ratings based on multi-criteria preference of users. After collecting movie ratings based on multi-criteria, the method proposed in this study is used for analyzing the trust-based community of the rating community. Users with the potential interest in rating prediction will be recommended.

Table 3. Multiple regression coefficient

	Unstandardized Coefficients		Standardized Coefficients		
	B	Std. Error	Beta	t	Sig.
(Constant)	.257	.088		2.939	.003
Screenplay/story, acting, cast, and sound & visual effect	.538	.021	.562	25.110	.000
	.254	.026	.235	9.625	.000
	.016	.019	.018	.846	.398
	.135	.017	.146	7.720	.000

(1) Statistics assessment

In order to assess if the four rating factors (independent variables) such as the collected “screenplay/story”, “acting”, “cast” and “sound & visual effect” have an significant influence on the total rating (dependent variable), the SPSS statistics software is used for conducting multiple regression analysis so as to find out the criteria with remarkable effect to carry out the subsequent experiments. The result is listed in table 3. The significant level “α=0.05” is used for verifying if each criterion has an influence on the total rating. We can know from table 3 that in the “cast” criterion, its t-value=0.846 and its p-value=0.398 > 0.05, so this criterion has no significant influence on the total rating. Therefore, the “cast” criterion will be removed, and the criteria such as “screenplay/story”, “acting” and “sound & visual effect” which have significant influences on the total rating will be used as the assessment criteria of this experiment.

(2) Division of data set

In order to discuss if the trust base will enhance the predictive effect, in this study all rating data will be divided into 3 different levels of data sets. The Sparsity Level refers to the percentage of the rating number under the rating matrix. The higher the sparsity level is, the lower the percentage of the rating number in the rating matrix will be, and vice versa. Formula 18 is used for calculating the sparsity levels of different sub-data sets, where $|Users|$ is the number of users in the data set, $|Movies|$ is the number of movies, and $|Ratings|$ is the rating sum in the data set. In this study, three data sets with different sparsity levels will be compared. They are respectively Data Set 1, Data Set 2 and Data Set 3, of which the sparsity level of Data Set 1 is: $1-\frac{405}{41\times 100}=0.9012$, the sparsity level of Data Set 2 is: $1-\frac{735}{41\times 117}=0.8468$, and the sparisty level of Data Set 3 is: $1-\frac{1049}{41\times 128}=0.8001$. This study uses the three data sets (See table 4) as a basis and compares them, and changes in effect of three data sets with different sparsity levels predicted by the trust base will be used for predicting the subsequent movie ratings and comparing the assessments of recommendation efficiency.

$$Sparsity\ Level = 1 - \frac{|Ratings|}{|Users| \times |Movies|} \qquad \text{(Formula 18)}$$

Table 4. Division of data set

	Number of users	Number of movies	Rating number	Sparsity Level
Data Set 1	41	100	405	0.9012
Data Set 2	41	117	735	0.8468
Data Set 3	41	128	1049	0.8001

Experiment of recommendation analysis

The experiment of recommendation analysis consists of two phases. The phase 1 is used as the parameter of trust-based community for dividing the experiment into 2 parts. Experiment 1 is the experiment set by the error range, ε, predicted by the movie co-rating members. In the predictive set, the percentage lower than the set, $CorrectSet(u')$, of the error range, ε, in the user recommendation set, $\mathrm{Re}\,cSet(u')$, is used for judging the rating trust value. The better error range will be found from the collecting rating data to judge if the user prediction is correct or not. Experiment 2 is the experiment set by the threshold value of the trust weight. It is necessary to find out the trust-based community of a movie based on each criterion when rating prediction of a movie is conducted, and members of the trust-based community shall be regarded as the basis of prediction of the trust base. After setting the parameter of experiment in the phase 1, the experiment in phase 2 is set up based on the parameter so as to assess the rating prediction and the recommendation effectiveness.

(1) Experiments of parameter setup

(a) Experiment 1: Predicted error range (ε) set
In the process of establishing the user rating trust value, the correct proportion predicted between the error range between the item predictive rating and the actual rating will be used as the trust value of this item by the user. When the tolerable error range is lower, and the correct predicted definition is stricter, the trust value of the item ratingd by the user is lower, and vice versa. Therefore, the predicted error range has an influence on the distribution number of the community rating trust value.

(b) Experiment 2: Threshold of trust weight set
The trust weight consists of the trust value of a movie rating by the community and the preferred coefficient of each criterion, the trust-based community must be found before a movie with unknown preference is rated by the target user based on each criterion. After setting the threshold value of the trust weight, the community whose value is higher than the threshold value will be used as a standard for selecting predictive community. The data sets with different sparsity levels have different rating numbers and densities. If a higher threshold value is used as the conditions for selecting the trust-based community of intensive data, it may ignore the members with higher trust value. Therefore, as for data sets with different sparsity levels, a better predicted threshold value must be found as the conditions for selecting the trust-based community based on each criterion.

In this phase, the design weight will accumulate different threshold values ($T_{Trust\_Weight} = 0.1$ 、 $T_{Trust\_Weight} = 0.2$ 、 $T_{Trust\_Weight} = 0.3$ 、 $T_{Trust\_Weight} = 0.4$ 、 $T_{Trust\_Weight} = 0.5$) at the rate of 0.1 to assess the threshold of the trust-based community in the data sets with different sparsity levels, and 20% of rating data will be randomly selected as the test date of each data set, and the test data will be predicted. After completion of experiment 1 and experiment 2 in the phase 1, the error range for judging if the user rating is right or not and the threshold value of the trust weight of each criterion under the data sets with different sparsity levels are respectively determined as the conditions for selecting the best predictive community. Parameters (See table 5) determined in this part is used for assessing the next prediction.

Table 5. Set value of parameter

Rating criteria	Threshold value of trust weight			Error range of user prediction
	Data Set 1 (sparsity level: 0.9012)	Data Set 2 (sparsity level: 0.8468)	Data Set 3 (sparsity level: 0.8001)	
Screenplay /Story	$T_{Trust\_Weight} = 0.3$	$T_{Trust\_Weight} = 0.2$	$T_{Trust\_Weight} = 0.1$	$\|\varepsilon\|<1$
Acting	$T_{Trust\_Weight} = 0.2$	$T_{Trust\_Weight} = 0.2$	$T_{Trust\_Weight} = 0.1$	
Sound &Visual effect	$T_{Trust\_Weight} = 0.3$	$T_{Trust\_Weight} = 0.2$	$T_{Trust\_Weight} = 0.1$	

Conclusion

The multi-criteria decision-making issue has been widely used in studying on both the decision supporting area and the marketing area. However, in the research area of the current recommendation system, the overall preference of the single rating is used as the recommendation basis. O'Donovan and Smyth (2005) considered the preferences of users for different movie criteria as the recommendation basis so as to increase the overall recommendation accuracy. In the multi-criteria recommendation environment requiring collecting many ratings, how to increase the reliability of rating prediction under each criterion, facilitate increasing the accuracy of the overall rating prediction, and increase the recommendation efficiency are the issues discussed in this study.

In order to ensure that an appropriate trust community can be found as the rating prediction under each criterion, the trust values and preferences of all members for each item are considered as the trust weight in this study. From the experiment conducted in phase 1, the threshold value of the trust weight is found out as the standard for selecting the trust community. It is found that the MAE effect of the predictive rating is better than the similarity-based prediction when the filtering community of the trust-based weight is used as the predictive basis under rating data with different sparsity levels in the rating predictive experiment of each criterion during the experiment 3 of phase 2. The use of the trust-based filtering can not only consider the trust community with more reference values as the prediction under each criterion, but the way only using the similarity-based prediction which may affect the predictive quality when the rating is relatively sparse can be avoided.

In predicting the overall rating in the experiment of phase 2, the trust-based filtering is used as the rating predictive community under each criterion, and it therefore has a better performance when the MAE value is predicted in the overall rating. The error dispersion judged by the standard deviation of error also has a small difference. Finally form the recommendation comparison conducted in the experiment 5, we found that the F1 value of the recommendation effectiveness is better than the traditional similarity-based prediction no matter what kind of overall rating aggregate function is used when the rating of each criterion uses the trust-based filtering approach to make prediction. By comparing with the overall rating aggregate functions, we found that the User-reg overall aggregate function has better recommendation effectiveness when the rating data are sufficient except that the insufficient rating data will make the recommendation efficiency of the User-reg overall aggregate function to be worse in the case that the data is sparser.

In the recommendation system, the way to increase the reliability of the predictive rating so as to enhance the recommendation effectiveness, and to reduce the cost for searching user information and increase customer loyalty is the main issue when a large amount of ratings are collected. In the multi-criteria recommendation environment, the preferred interest information of users based on different criteria is collected. In order to enable more rating information to increase the recommendation accuracy, the trust community will be used for offsetting the shortage of the collaborative filtering prediction so as to enhance the multi-criteria recommendation efficiency in this study. Therefore, this study has the following contributions:

(1) Members with higher rating reliability and preference are used as the trust members, and the threshold value of the trust weight is used as the standard for selecting the trust community under each criterion so as to enhance the predictive effect of the trust community.
(2) The trust community used as the rating prediction can obviously increase the efficiency, and can reduce the inherent loss of the similarity-based prediction when the data with different sparsity levels is cut and the rating data is sparser.
(3) In this study the perdition of the trust community is used in the multi-criteria recommendation environment, so the rating error of each criterion for movies with unknown user preference can be reduced under much rating information of the multi-criteria, and hence the overall rating predictive ability and the recommendation efficiency can be enhanced. Therefore, the feasibility of the trust-based recommendation framework has been verified in this study under the multi-criteria environment.

References

Adomavicius, G. and Kwon, Y., 2007. "New Recommendation Techniques for Multicriteria Rating Systems, " *IEEE Intelligent Systems*, (22: 3), pp. 48-55.

Adomavicius, G., and Tuzhilin, E. 2005. "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, " *IEEE Transactions on Knowledge and Data Engineering*, (17:6), pp. 734-749.

Ansari, A., Essegaier, S., and Kohli, R. 2000. "Internet Recommendation Systems, *Journal of Marketing Research*, " (37:3), pp. 363-375.

Bichler, M. 2001. "An Experience Analysis of Multi-Attribute Auction, " *Decision Support System*, (29:10), pp. 249-268.

Billsus, D. and M. Pazzani. 2000. "User Modeling for Adaptive News Access," *Journal of User Modeling and User-Adapted Interaction*, (10:2-3), pp. 147-180.

Breese, J. S., Heckerman, D. and Kadie, C. 1998. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 43-52.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., and Sartin, M. 1999. "Combining Content based and Collaborative Filters in an Online Newspaper," in *Proceedings of the ACM SIGIR '99 Workshop on Recommender Systems: Algorithmsand Evaluation*, University of California, Berkeley.

Green, P. E., Krieger, A. M., and Wind, Y. 2001. "Thirty Years of Conjoint Analysis: Reflections and Prospects," *Interfaces*, (31:3), pp. 56-73.

Herlocker, J. L., Konstan, J. A., and Terveen, L. G. 2004. "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, (22), pp. 5-53.

Huang, Z., Chen, H., and Zeng, D. 2004. "Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering," *ACM Transactions on Information Systems*, (22:1), pp. 116-142.

Jeong B., Lee J., and Cho H. 2009. "User credit-based collaborative filtering," *Expert Systems with Applications*, (36:3), pp. 7309-7312.

Kim, B. M., Li, Q., Kim, S. G., and Kim, J. Y. 2006. "A New Approach for Combining Content-based and Collaborative Filters," *Journal of Intelligent Information Systems archive*, (27:6), pp. 79-91.

Kwon, K., Cho J., and Park, Y. 2009. "Multidimensional credibility model for neighbor selection in collaborative recommendation," *Expert Systems with Applications* (36:3), Part 2, pp. 7114–7122.

Lam, S. K., McNee, S. M., Konstan, J. A., and Riedl, J. 2002. “Getting to Know You: Learning New User Preferences in Recommender System,” in *Proceedings of the International Conference on Intelligent User Interfaces*, pp. 127-134.

M. Bichler, 2000. “An Experimental Analysis of Multi-Attribute Auctions,” *Decision Support Systems*, (29:10), pp. 249-268.

Massa, P., and Bhattacharjee, B. 2005. “Using Trust in Recommender Systems: An Experimental Analysis,” in *Proceedings of 2nd International Conference on Trust Management*, Oxford, England.

O'Donovan, J., and Smyth, B. 2005. “Trust in recommender systems. in *Proceedings of the 10th international conference on Intelligent user interfaces*, pp. 167-174.

Papagelis, M., Plexousakis, D., and Kutsuras T. 2005. “Alleviating the Sparsity Problem of Collaborative Filtering Using Trust Inferences, in *Proceedings of the 3rd International Conference on Trust Management*, pp. 224-239.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews, in *Proceedings of CSCW 94 Conference on Computer Supported Cooperative Work*, pp. 175-186.

Sarwar, B., Karypis, G., Konstan, J. A. and Riedl, J. 2001. “Item-based Collaborative Filtering Recommendation Algorithms,” in *Proceedings of Tenth International WWW Conference, 2001*, pp. 285-295.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. 2000. “Analysis of Recommendation Algorithms for E-Commerce,” in *Proceedings of the ACM E-Commerce 2000 Conference*, pp. 158-167.

Schafer, J. B., Konstan, J. A., and Riedl, J. 2000. “E-Commerce Recommendation Applications, *Journal of Data Mining and Knowledge Discovery*,” (5:1), pp. 115-152.

Schein, A. I., Popescul, A., Ungar L. H., and Pennock, D. M. 2002. “Methods and Metrics for Cold-Start Recommendations,” in *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, Tampere*, pp.253-260.

Shardanand, U., and Maes, P. 1995. Social Information Filtering: Algorithms for Automating ‘Word of Mouth’, in *Proceedings of the Conference on Human Factors in Computing Systems (CHI95)*, 1995, pp. 210-217.

Victor, P., Corneilis, C., and Cock M. De. 2009. “Trust-and Distrust-Based Recommendations for Controversial Reviews,” in *Proceedings of the WebSci'09: Society On-Line*, Athens, Greece.