# A Linear Regression Approach to Multi-criteria Recommender System

Tanisha Jhalani[1], Vibhor Kant[1(✉)], and Pragya Dwivedi[2]

[1] The LNMIIT, Jaipur 302031, India
tanishajhalani75@gmail.com, vibhor.kant@gmail.com
[2] MNNIT Allahbad, Allahbad 211004, India
pragya.dwijnu@gmail.com

**Abstract.** Recommender system (RS) is a web personalization tool for recommending appropriate items to users based on their preferences from a large set of available items. Collaborative filtering (CF) is the most popular technique for recommending items based on the preferences of similar users. Most of the CF based RSs work only on the overall rating of the items, however, the overall rating is not a good representative of user preferences for an item. Our work in this paper, is an attempt towards incorporating of various criteria ratings into CF i.e., multi-criteria CF, for enhancing its accuracy through multi-linear regression. We suggest the use of multi-linear regression for determining the weights of individual criterion and computing the overall ratings of each item. Experimental results reveal that the proposed approach outperforms the classical approaches.

**Keywords:** Recommender systems · Collaborative filtering · Multi-criteria decision making · Linear regression

## 1 Introduction

During last decade, information is expanding tremendously. Instead of helping the users, this great amount of information caused the problem of information overload. To handle this explosive growth of information, a personalization tool is needed that can assist a user to get the valid and appropriate information. Recommender system (RS) is one of the most successful personalization tools that guides a user to select an appropriate item from a large set of alternatives [1,2].

Generally, recommender system employs three major filtering techniques, namely, collaborative filtering (CF), content-based filtering (CBF) and hybrid filtering (HF). Among these techniques, collaborative filtering is widely used in the recommender system. Most of the existing RSs are based on the single criterion collaborative filtering [3,4], In a single criterion CF, only overall rating of item is considered, but the overall rating of an item depends on the different criteria. So instead of considering only single criterion, multiple criteria are should be used in multi-criteria CF [3,5]. In heuristic approaches of MCCF, all criteria have same priorities, but this is not an optimal scenario because different users have different priorities on various criteria, so in [3,9], it was suggested


Y. Tan and Y. Shi (Eds.): DMBD 2016, LNCS 9714, pp. 235–243, 2016.
DOI: 10.1007/978-3-319-40973-3_23

that weights on these criteria can be computed using either some machine learning techniques or any appropriate statistical techniques. Based on the above discussion, the contributions of our paper can be summarized as follows:

- First of all, we propose the use of multi linear regression approach for deriving the individual weight for each criterion.
- Second, we aggregate similarities and ratings for different criteria using these weights.
- Third, we perform rigorous experiments, on very popular and large Yahoo movie dataset by varying the number of users and compare our approach with various benchmark algorithms for single criterion and multi-criteria CF.

The rest of this paper is organized as follows: Sect. 2 describes background related to MCCF and multi linear regression. In Sect. 3, we have discussed proposed approach. Section 4 shows experimental evaluation of our proposed approach. Finally, last Section provides some concluding remarks.

## 2 Background and Related Work

This section briefly describes collaborative filtering for multi-criteria and multi linear regression.

### 2.1 Multi-criteria Recommender System

In multi-criteria RS, user rates various criteria of an item. The complete process of multi-criteria CF can be summarized into the following three phases:

- **Phase 1 (Similarity computation):** In this phase, first multi-criteria data set is divided into k single criterion datasets (where k is the number of criteria) and then similarities are computed for each criterion separately using some similarity measures like Pearson correlation and cosine similarity [3]. Now, overall similarity is computed using any aggregation function [10,11] which is expressed as follows :

$$Sim_{aggregate}(u, u') = \sum_{c=0}^{k} w_c sim_c(u, u') \tag{1}$$

  where, $w_c$ is the weight of each criterion. In the above equation, if weights are same for all criteria, like $w_1 = w_2 = w_3 = .... = w_k$ then aggregation function is similar to the average of similarities [3]. But this technique is not appropriate for aggregation because weights may be different for each criterion and it is a challenge to find these weights. Therefore, we use multi linear regression for computing these weights for various criterion.
- **Phase 2 (Neighborhood generation):** After computing similarities between active user and remaining users, neighborhood set is formed as a collection of similar users either using nearest neighbor approach (Top N users) or threshold based approach.

– **Phase 3 (Prediction of unknown rating):** In this phase, unknown rating is predicted for each criterion separately using following prediction function [7,12]:

$$r^{p}_{u,i} = \frac{1}{\sum_{u' \in U_t} |Sim(u,u')|} \sum_{u' \in U_t} Sim(u,u') \times r_{u',i} \tag{2}$$

Now, these ratings are aggregated and overall rating is predicted for the users [12,13].

### 2.2 Multi Linear Regression

Linear regression is a statistical technique for finding the relationship between a dependent variable Y and independent variable X [14,15]. If independent variable is one then it is called simple linear regression and in case of more than one independent variables it is known as multi linear regression. Multi linear regression can be represented as follows:

$$Y = w_0 + w_1 x_1 + w_2 x_2 + ... + w_k x_k \tag{3}$$

where, $Y$ is called as dependent variables and $x_1, x_2, ..., x_k$ are independent variables. $w_0, w_1, w_2, ..., w_k$ are the weight parameters corresponding to independent variables which are computed on the basis of some observations. In proposed approach, multi linear regression is used to find the weights for different criteria.

## 3 Proposed Recommendation Approach

This section describes the proposed multi-criteria recommender system utilizing the concept of multi linear regression. Multi linear regression is used to aggregate the similarities and to find the overall ratings by using weights for each criterion. Before presenting our proposed approach, we discuss about the inputs required for our system. For multi-criteria RS, Let $U = \{u_1, u_2, u_3, ..., u_n\}$ be the set of n users , $I = \{i_1, i_2, i_3, ..., i_m\}$ is the set of m items. and $C = \{c_1, c_2, c_3, ..., c_k\}$ is the set of k criteria. The rating vectors for user u to item i is represented as $R(u,i) = (r^0_{u,i}, r^1_{u,i}, r^2_{u,i}, ...., r^k_{u,i})$, which consists of an overall rating $r^0_{u,i}$, and k multi-criteria ratings $r^1_{u,i}, r^2_{u,i}, ...., r^k_{u,i}$. Our proposed system has following three phases:

Phase 1: Multi-linear regression based similarity computation
Phase 2: Neighborhood generation
Phase 3: Multi-linear regression approach to prediction

– **Phase 1. Multi-linear regression based similarity computation:** In proposed multi-criteria RS, following two steps are required for similarity computation.

- **Step 1 (Similarity computation for each criterion):** In this step, multi-criteria ratings are divided into k single criteria ratings and then similarities are estimated between user u and u' is computed as follows:

$$sim^c(u,u') = \frac{\sum_{i\in I}(r^c_{u,i} - \bar{r}^c_u)(r^c_{u',i} - \bar{r}^c_{u'})}{\sqrt{\sum_{i\in I}(r^c_{u,i} - \bar{r}^c_u)^2}\sqrt{\sum_{i\in I}(r^c_{u',i} - \bar{r}'^c_u)^2}} \tag{4}$$

where c represents the different criteria, i.e., $c = \{1, 2, 3, ..., k\}$.

- **Step 2 (Aggregation of similarities):** In this step, overall similarity is computed using following equation:

$$sim(u,u') = w_0 + \sum_{c\in\{1,...,k\}} w_c sim^c(u,u') \tag{5}$$

where, $sim^c(u,u')$ is the similarity between user u and $u' \in U$ for criteria $c \in \{1, ..., k\}$, $w_c$ is the weight parameter for criteria $c \in \{1, ..., k\}$ and $w_0$ is the error term.

Using multi-linear regression, weight parameters are estimated on the basis of previously rated item by users which is called training data. Table 1 represents the training data.

**Table 1.** Presentation of training data

| **S.No.** | $C_1$ | $C_2$ | ... | $C_k$ | $C_0$ |
|---|---|---|---|---|---|
| 1 | $r_{1,1}$ | $r_{2,1}$ | | $r_{k,1}$ | $r_{0,1}$ |
| 2 | $r_{1,2}$ | $r_{2,2}$ | | $r_{k,2}$ | $r_{0,2}$ |
| . | . | . | | . | . |
| . | . | . | | . | . |
| n | $r_{1,n}$ | $r_{2,n}$ | | $r_{k,n}$ | $r_{0,n}$ |
| **Total** | $\sum_i r_{1,i}$ | $\sum_i r_{2,i}$ | | $\sum_i r_{k,i}$ | $\sum_i r_{0,i}$ |

where, $C_1, C_2, ..., C_k$ are different single criteria ratings and $C_0$ is the overall rating. $r_{k,i}$ is the rating of $i^{th}$; $i \in \{1, 2, ..., n\}$ training data for criteria k. Based on the training data weight values are derived using following equation in matrix form [5]:

$$\begin{bmatrix} w_1 \\ \vdots \\ w_k \end{bmatrix} = \begin{bmatrix} \sum_i u^2_{1,i} & \dots & \sum_i u_{1,i}u_{k,i} \\ \vdots & \ddots & \vdots \\ \sum_i u_{1,i}u_{k,i} & \dots & \sum_i u^2_{k,i} \end{bmatrix}^{-1} \begin{bmatrix} \sum_i u_{1,i}v_i \\ \vdots \\ \sum_i u_{k,i}v_i \end{bmatrix} \tag{6}$$

where,

$$\sum_{i\in\{1,...,n\}} u_{j,i}u_{k,i} = \sum_{i\in\{1,...,n\}} r_{j,i}r_{k,i} - \frac{\sum_{i\in\{1,...,n\}} r_{j,i} \sum_{i\in\{1,...,n\}} r_{k,i}}{n} \tag{7}$$

$$\sum_{i\in\{1,\ldots,n\}} u_{j,i}v_i = \sum_{i\in\{1,\ldots,n\}} r_{j,i}r_{0,i} - \frac{\sum_{i\in\{1,\ldots,n\}} r_{j,i} \sum_{i\in\{1,\ldots,n\}} r_{0,i}}{n} \quad (8)$$

here, n is total number of samples in training data and $j \in \{1, 2, ..., k\}$. $w_0$ is called the error term which is computed as follows.

$$w_0 = \bar{r_0} - w_1\bar{r_1} - w_2\bar{r_2} - ... - w_k\bar{r_k} \quad (9)$$

By applying these weight values in Eq. (5) overall similarity is calculated.

- **Phase 2. Neighborhood generation:**
  This phase is similar to the phase 2 of MCCF.
- **Phase 3. Multi linear regression approach to prediction:**
  In this phase, we predict the overall rating using Eq. (2). In this equation, the overall rating of an item given by these nearest neighbors is utilized. The important task in this phase is to compute the overall rating of an item through its criteria ratings. We have employed again a linear regression approach to aggregate the criteria ratings. The aggregation function for this task is expressed as follows:

$$r(u, u') = w_0 + \sum_{c\in\{1,\ldots,k\}} w_c r_c \quad (10)$$

where, $r_c$ represents the rating for criteria $c \in \{1, ..., k\}$, $w_c$ is the weight parameter for criteria $c \in \{1, ..., k\}$ and $w_0$ is the error term. these weights are calculates using Eq. (6) and then we compute overall rating. After finding overall rating, we have used Eq. (2) for predicting unknown rating to an active user. Finally, we have recommended highly some predicted items to users.

## 4 Experiments and Results

We performed various experiments to analyze the effectiveness of the proposed multi-criteria recommender system using Yahoo movie dataset, which consists of 6078 users and 976 items. Each item has five different criteria from which four are individual features and fifth is the overall rating. For experiments, 10 fold cross-validation mechanism is used. In each fold, 60 % data of each user is considered as training data and 40 % data is used as test data. Training data is used to learn the system and test data is used to analyze the performance of the system. In order to evaluate the performance of our proposed system, we have used mean absolute error (MAE), coverage, recall and f-measure as evaluation metrices:

To demonstrate the feasibility and effectiveness of proposed system we have compared our results with the following approaches:

- Single criterion CF (SCCF)
- Multi-criteria collaborative filtering using average similarity and ratings (MCCF-A)

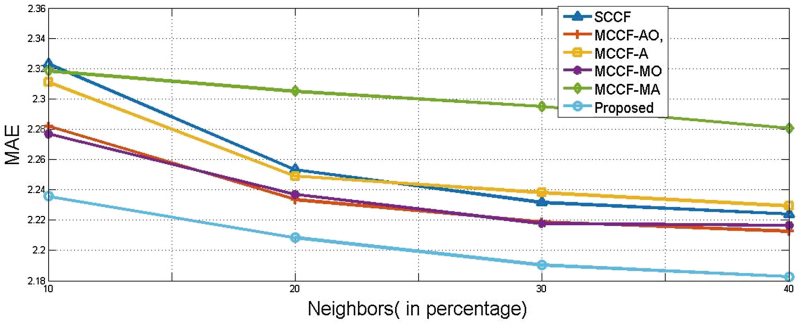


**Fig. 1.** MAE comparison on different number of neighbors

**Table 2.** Performance comparison via MAE, coverage, recall, f-measure

| | MAE | Coverage | Recall | F-measure |
|---|---|---|---|---|
| **SCCF** | 2.2406 | 0.9989 | 0.8477 | 0.8377 |
| **MCCF-AO** | 2.2367 | 0.9989 | 0.8483 | 0.8386 |
| **MCCF-A** | 2.2191 | 0.9989 | 0.8811 | 0.8452 |
| **MCCF-MO** | 2.2434 | 0.9990 | 0.8458 | 0.8367 |
| **MCCF-A** | 2.2227 | 0.9990 | 0.8786 | 0.8440 |
| **Proposed** | 2.1995 | 0.9990 | 0.9108 | 0.8500 |

**Table 3.** Performance comparison of the proposed approach with other approach for different number of users

| | Performance Measures | Y_1000 | Y_2000 | Y_3000 | Y_4000 | Y_5000 | Y_6078 |
|---|---|---|---|---|---|---|---|
| **SCCF** | **MAE** | 2.3819 | 2.2.2803 | 2.2939 | 2.2539 | 2.2343 | 2.2406 |
| | **F-Measure** | 0.8132 | 0.8280 | 0.88271 | 0.8370 | 0.8403 | 0.8377 |
| **MCCF-AO** | **MAE** | 2.3861 | 2.2952 | 2.2982 | 2.2466 | 2.2390 | 2.2367 |
| | **F-Measure** | 0.8120 | 0.8291 | 0.8283 | 0.8369 | 0.8405 | 0.8386 |
| **MCCF-A** | **MAE** | 2.2396 | 2.2596 | 2.2692 | 2.2217 | 2.2164 | 2.22191 |
| | **F-Measure** | 0.8165 | 0.8344 | 0.8354 | 0.8450 | 0.8455 | 0.8452 |
| **MCCF-MO** | **MAE** | 2.3842 | 2.2927 | 2.2934 | 2.2506 | 2.2387 | 2.2434 |
| | **F-Measure** | 0.8088 | 0.8267 | 0.8290 | 0.8354 | 0.8404 | 0.8367 |
| **MCCF-MA** | **MAE** | 2.3521 | 2.2622 | 2.2586 | 2.2293 | 2.2183 | 2.2227 |
| | **F-Measure** | 0.8137 | 0.8336 | 0.8364 | 0.8447 | 0.8454 | 0.8440 |
| **Proposed** | **MAE** | 2.2212 | 2.2352 | 2.1987 | 2.1731 | 2.2030 | 2.1995 |
| | **F-Measure** | 0.8457 | 0.8394 | 0.8435 | 0.8509 | 0.8484 | 0.8500 |

– Multi-criteria collaborative filtering using average similarity &overall rating (MCCF-AO)
– Multi-criteria collaborative filtering using minimum similarity &average rating (MCCF-MA)
– Multi-criteria collaborative filtering using minimum similarity &overall rating (MCCF-MO).

### 4.1 Experiment 1

In this experiment, we calculate the predictive and classification accuracy of proposed approach via MAE, coverage, recall and f-measure. Table 2. presents results for these measures by taking 30 % most similar user as neighbors and shows that our proposed approach outperformed in terms these measure. Figs. 1 and 2, show the results for different percentages of users ( 10 %, 20 %, 30 % and 40 %) on MAE and f-measure. It reveals that proposed approach has minimum MAE and maximum f-measure.

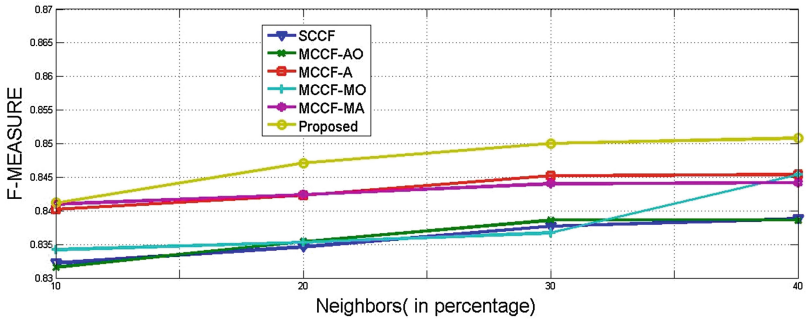


**Fig. 2.** F-measure comparison on different number of neighbors

### 4.2 Experiment 2

This experiment reflects the scalability of proposed approach. For this experiments we choose six different subsets of Yahoo movie dataset, called Y_1000, Y_2000, Y_3000, Y_4000, Y_5000, Y_6078. Table 3. depicts the effectiveness of proposed approach under varying number of participating users. Fig. 3 depicts the results of F-measure for different scheme on different subset of dataset.

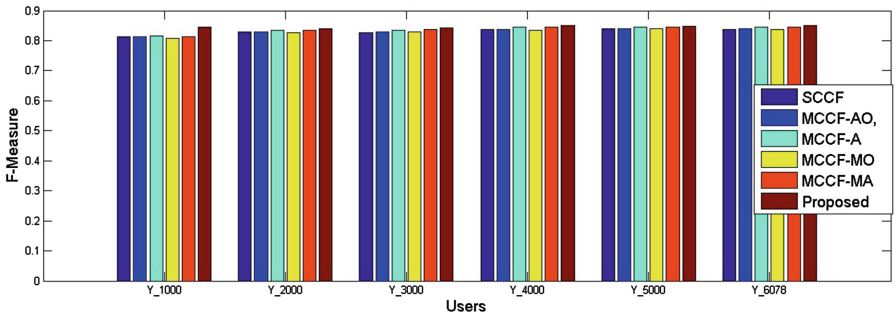


**Fig. 3.** F-measure comparison for different users (Color figure online)

## 5 Conclusion

In this work, we have presented linear regression based multi-criteria recommender system (MCRS) where linear regression is used to aggregate similarity components on various criteria and to compute overall rating. Generally different users have different priorities on various criteria where they evaluate these criteria based on their perceptions. The aggregation of similarities based on each criterion is quite challenging task in the area of MCRS because the used weights

in aggregation task are not optimal. We have used linear regression approach to compute these weights optimally. Experimental results on a popular Yahoo dataset demonstrated that the adoption of linear regression approach in MCRS has produced quality recommendation and established that our proposed approach outperformed other heuristic approaches.

In our future work, we are planning to handle uncertainty associated with user preferences using fuzzy sets [16] and we will explore some new methods for dealing with correlation based similarity problems.

## References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Engg. **17**(6), 734–749 (2005)
2. Bobadilla, J., Ortega, F., Hernando, A., Gutirrez, A.: Recommender systems survey. Knowl. Based Syst. **46**, 109–132 (2013)
3. Adomavicius, G., Kwon, Y.: New recommendation techniques for multicriteria rating systems. IEEE Int. Syst. **22**(3), 48–55 (2007)
4. Soboroff, I., Nicholas, C.: Combining Content and Collaboration in Text Filtering. In: International Joint Conference on Artificial Intelligence, pp. 86–92 (1999)
5. Balabanovi, M., Shoham, Y.: Fab: content-based collaborative recommendation. ACM Comm. **40**(3), 66–72 (1997)
6. Kant, V.: A user-oriented content based recommender system based on reclusive methods and interactive genetic algorithm. In: Bansal, J.C., Singh, P.K., Deep, K., Pant, M., Nagar, A.K. (eds.) Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). Advances in Intelligent Systems and Computing, vol. 201, pp. 543–554. Springer, India (2013)
7. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In: ACM Conference on Computer Supported Cooperative Work, pp. 175–186. ACM (1994)
8. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: 14th Conference on Uncertainty in Artificial Intelligence, pp. 43–52. Morgan Kaufmann Publishers Inc., San Francisco (1998)
9. Al-Shamri, M.Y.H., Bharadwaj, K.K.: Fuzzy-genetic approach to recommender systems based on a novel hybrid user model. Expert Syst. Appl. **35**(3), 1386–1399 (2008)
10. Delgado, J., Ishii, N.: Memory-based weighted majority prediction. In: SIGIR Workshop on Recommender System. Citeseer (1999)
11. Jannach, D., Karakaya, Z., Gedikli, F.: Accuracy improvements for multi-criteria recommender systems. In: 13th ACM Conference on Electronic Commerce, pp. 674–689. ACM (2012)
12. Winarko, E., Hartati, S., Wardoyo, R.: Improving the prediction accuracy of multicriteria collaborative filtering by combination algorithms. Int. J. Adv. Comput. Sci. App. **52**(4), 52–58 (2014)
13. Bilge, A., Kaleli, C.: A multi-criteria item-based collaborative filtering framework. In: 11th International Joint Conference on Computer Science and Software Engineering, pp. 18–22. IEEE (2014)
14. Agarwal, B., L.: Basic Statistics. New Age International (2006)

15. Kutner, M.H.: Applied Linear Statistical Models, vol. 4. Irwin, Chicago (1996)
16. Kant, V., Bharadwaj, K.: Integrating collaborative and reclusive methods for effective recommendations: a fuzzy bayesian approach. Int. J. Int. Syst. **28**(11), 1099–1123 (2013)