INTEGER LINEAR PROGRAMMING FOR TRANSFORMING PAIRWISE BASED RESULTS TO THE ORIGINAL RATINGS

***Ping Ji*[1,2], *Jian Jin*[1,2]**

[1]*Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong*

[2]*The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen 518057, China*

p.ji@polyu.edu.hk, jinjian.jay@connect.polyu.hk

Keywords: integer linear programming, pairwise, optimization.

Abstract

Many pairwise models are proposed for ranking problems in the field of information retrieval. Classification problems in the field of data mining also use pairwise comparison. However, conventionally, these pairwise approaches are evaluated based evaluation metrics. The original rating for a single document or instance is not explained faithfully, which makes these algorithms cannot be evaluated by standard evaluation metrics, such as Mean Average Precision and Normalized Discounted Cumulative Gain for ranking models. In this research, the focus is on how to transform pairwise based results to the original ratings. Particularly, an integer linear programming model is formulated for this problem. In this algorithm, the objective is to minimize the number of conflicts for the predicted pairwise based relationship between instances by the assignment of rating values. An example is presented in order to clarify the proposed integer linear programming method. It validates the possibility to transform pairwise based results to the original ratings, which make them to be evaluated by standard evaluation metrics.

1 Introduction

With fast development of internet, an unimaginably vast number of information can be seen. Online information facilitates many previously challengeable things. For instance, rich information is provided in search engines. Finding relevant and desired information is getting more convenient. Also, customer opinions can be identified from a large number of online reviews efficiently for both consumers and designers. However, the extremely large size of the Web makes it a challenge to digest and utilize all these online big data. How to manage big data well and provide direct insight becomes ever more important.

This problem is also a hot topic in some relevant research areas. Many innovative models are proposed to capture and process various types of information. For example, in the area of information retrieval, different ranking algorithms are developed and they are widely used in search engines [7]. One famous category of ranking models is pairwise algorithms for learning to rank. Pairwise algorithms do not directly model the relevance degree of each document. The focus is on the relative order between document pairs. Another good example is that, due to the complexity of the multiclass classification problem, different methods are established by making pairwise comparison between instance pair. These multiclass classification methods are then utilized to classify online customer reviews, etc.

However, one major problem of pairwise algorithms is that they do not interpret the original relevance degrees or classification labels faithfully. Specially, pairwise algorithms for learning to rank do not explain the relevance degree for each document explicitly. It makes that these pairwise approaches is not convenient to be evaluated by some standard evaluation metrics, such as Mean Average Precision (MAP) and Normalized Discounted Cumulative Gain (NDCG) for ranking models. Moreover, due to various reasons, it is generally difficult to train a ranking model or a classifier without errors. Comparing with ground truth, there might be some errors in the predicted ranking list or misclassified instances for multiclass classification problems. It induces that, generally, the pairwise-based results are not able to be transformed into the original ratings for all instances faithfully. For instance, the relevance preference between three document pairs are predicted as $p(d_A)> p(d_B)$, $p(d_B)> p(d_C)$, $p(d_C)> p(d_A)$. Accordingly, the relevance degree of three documents cannot be assigned to satisfy all the relationship.

In this research, an integer linear programming problem is formulated for transforming pairwise based results to the original ratings. In this integer linear programming problem, the number of violations is modeled as integer variables. The objective of the integer linear programming model is to find the most suitable assignments of the ratings for instances and, at the same time, to minimize the total number of violations about the predicted relationship.

The rest of this paper is organized as follows: some related work about pairwise approaches is reviewed in Section 2, which is mainly on pairwise algorithms for learning to rank. The problem to be studied in this research is defined in Section 3. In Section 4, the proposed model for transforming pairwise based results to the original ratings is introduced. In Section 5, an example is presented to clarify the details of the proposed method. Finally, this research is concluded in Section 6.

2 Literature Reviews on Pairwise Approaches

In pairwise algorithms for learning to rank [6], the input space contains feature vectors of document pairs. The output space is the pairwise preference between each document pair. The hypothesis space of the pairwise approaches contains functions that take a pair of documents as input and output the relative order between them. The loss function measures the inconsistency between predicted pairwise preference and ground truth preference of document pairs. The pairwise approaches do not target at accurately predicting the relevance degree of each single document. The relative order between two documents is cared about, where the output takes values from {-1, 1}. In the pairwise approaches, ranking is usually reduced to a classification problem on document pairs. The preference between document pairs is evaluated in these pairwise approaches.

A neural network was built to learn a preference function for all possible document pairs in training data [1, 8]. A boosting approach on document pairs was also utilized to combine ranking functions in RankBoost [2]. Based on SVM (Support Vector Machine), RankSVM was proposed to perform the pairwise classification [3, 5]. RankSVM differs from SVM at its constraint part and the loss function, which was built from document pairs. Based on RankSVM, a pairwise classification method was proposed to tune the weights of product characteristic for product designer from online reviews [4]. However, a single hyperplane in the feature space is employed by RankSVM, which is arguable to capture the nature of complex ranking problems [9]. Multiple hyperplanes were proposed to train a ranking model for document pairs. Finally, the ranking results predicted by each ranking model were aggregated to produce the final ranking result.

Although pairwise approaches have their own advantages, the ignored fact is that the original rating value or the relevance degree of each instance is not explained faithfully.

3 Problem Definition

In order to clarify the problem to be studied in this research, some notation will be defined step by step in this section.

In pairwise approaches, such as pairwise algorithms for learning to rank, the output space is the pairwise preference between instance (or document) pair, rather than the original ratings of each single instance. Generally, given two instances, d_i and d_j, the ground truth ratings of them are y_i and y_j, respectively. y_i often stands for the number of ratings of one online review, the relevance degree of one document, etc. It can be denoted with one integer number within a limited scope. For example, $y_i \in \{1, 2, 3, 4, 5\}$ or $y_i \in \{-3, -2, -1, 0, 1, 2, 3\}$. The differences between two kinds of rating systems are the initial number and the rating scale. For simplicity, for a single instance, the ground truth rating is defined as,

$$y_i \in \{1, 2, \ldots, T\} \tag{1}$$

T is the maximal number that y_i can be chosen. For instance, in the first exa mple, T equals 5.

The focus of pairwise approaches is not on modeling how to derive the original ratings, y_i and y_j, directly. The objective is to predict the ground truth preference between two instances, d_i and d_j. Accordingly, the ground truth preference between two instances, d_i and d_j, and the predicted preference can be denoted as $p(d_i, d_j)$ and $\hat{p}(d_i, d_j)$, respectively.

Notice that, in pairwise algorithms for learning to rank, $\hat{p}(d_i, d_j)$ is either -1 or 1, and it is denoted as $\hat{p}(d_i, d_j) \in \{-1, 1\}$, where -1 means the predicted preference of d_i is higher than that of d_j. However, there is a possibility that the preference between d_i and d_j can be equal. Mathematically, it is defined as

$$\hat{p}(d_i, d_j) \in \{-1, 0, 1\} \tag{2}$$

0 means the preference between d_i and d_j is equal.

Many pairwise approaches have been proposed, and they are applied in different areas. However, the output of pairwise approaches does not tell the predicted rating of each single instance. Specially, it does not explain the predicted rating $\hat{y}_i$ for instance d_i. What is missing in pairwise approaches is that the performance of them is not evaluated by standard evaluation metrics. For instance, Precision, Recall and F-measure are usually desired classification evaluation metrics. MAP and NDCG are utilized to evaluate the performance of ranking models. But it is important to know the original ratings of each instance.

Hence, the problem is how to transform pairwise based results to the original rating of each instance. More specifically, the problem to be studied in this research is how to assign the value of $\hat{y}_i$ for d_i in order to make the predicted preference $\hat{p}(d_i, d_j)$ to be satisfied for all instance pairs d_i and d_j.

4 Method

In this research, the problem is to assign the ratings of $\hat{y}_i$ and $\hat{y}_j$ to satisfy the predicted preference $\hat{p}(d_i, d_j)$ for all instance pairs d_i and d_j. In other words, the assignment of ratings about $\hat{y}_i$ and $\hat{y}_j$ should make the number of violations for the predicted relationship $\hat{p}(d_i, d_j)$ to be minimized.

In particular, there are two instances, d_i and d_j, with $\hat{p}(d_i, d_j) = 1$. It implies that the predicted rating $\hat{y}_i$ for instance d_i should be bigger than $\hat{y}_j$ if the predicted preference $\hat{p}(d_i, d_j) = 1$ is satisfied. Let $\alpha \in \{0, 1\}$ be the flag that denotes whether the relationship is satisfied or not by the assignment of values for $\hat{y}_i$ and $\hat{y}_j$. It can be mathematically formalized as

$$\begin{aligned}&\hat{y}_i - \hat{y}_j \geq 1 - M \cdot \alpha \\ &\hat{y}_j - \hat{y}_i \geq 1 - M \cdot (1-\alpha) \\ &\alpha \in \{0,1\} \\ &\hat{y}_i, \hat{y}_j \in \{1,2,\ldots,T\}\end{aligned} \tag{3}$$

In Model (3), if the assignment of values for $\hat{y}_i$ and $\hat{y}_j$ is not satisfied, α will equal to one, otherwise α will be zero. M is a sizable number. For instance, M equals to 10^3.

Likewise, if two instances, d_i and d_j are confined as $\hat{p}(d_i, d_j) = -1$, the equivalent model can be formulated as follows:

$$\begin{aligned}&\hat{y}_j - \hat{y}_i \geq 1 - M \cdot \beta \\ &\hat{y}_i - \hat{y}_j \geq 1 - M \cdot (1-\beta) \\ &\beta \in \{0,1\} \\ &\hat{y}_i, \hat{y}_j \in \{1,2,\ldots,T\}\end{aligned} \tag{4}$$

β is either zero or one, denoting whether the relationship $\hat{p}(d_i, d_j) = -1$ is satisfied or not by the assignment of values for $\hat{y}_i$ and $\hat{y}_j$.

The third case is that the preference between d_i and d_j is equal, which means $\hat{p}(d_i, d_j) = 0$. If γ is symbolized whether the equation relationship is satisfied or not, a slightly different model can be derived as,

$$\begin{aligned}&\hat{y}_i - \hat{y}_j \leq M \cdot \gamma \\ &\hat{y}_j - \hat{y}_i \leq M \cdot \gamma \\ &\gamma \in \{0,1\} \\ &\hat{y}_i, \hat{y}_j \in \{1,2,\ldots,T\}\end{aligned} \tag{5}$$

According to Model (3), Model (4), and Model (5), α, β and γ are utilized to denote whether the corresponding relationship is satisfied or not. Hence, the sum of α, β and γ represents the total number of the relation that are not dissatisfied by the assignment of values for $\hat{y}_i$ and $\hat{y}_j$ for all instance pairs d_i and d_j. Hence, the question of transforming pairwise based results to the original ratings faithfully turns to minimize the sum of α, β and γ.

Combing Model (3), Model (4), and Model (5), the final model is shown in Model (6).

$$\begin{array}{llcc} & \min\{\sum \alpha_i + \sum \beta_j + \sum \gamma_k\} & & \\ s.t. & \hat{y}_{ai} - \hat{y}_{bi} & \geq & 1 - M \cdot \alpha_i \\ & \hat{y}_{bi} - \hat{y}_{ai} & \geq & 1 - M \cdot (1-\alpha_i) \\ & \hat{y}_{bj} - \hat{y}_{aj} & \geq & 1 - M \cdot \beta_j \\ & \hat{y}_{aj} - \hat{y}_{bj} & \geq & 1 - M \cdot (1-\beta_j) \\ & \hat{y}_{ak} - \hat{y}_{bk} & \leq & M \cdot \gamma \\ & \hat{y}_{bk} - \hat{y}_{ak} & \leq & M \cdot \gamma \\ & \alpha_i, \beta_j, \gamma_k & \in & \{0,1\} \\ & \hat{y}_{ai}, \hat{y}_{bi}, \hat{y}_{aj}, \hat{y}_{bj}, \hat{y}_{ak}, \hat{y}_{bk} & \in & \{1,2,\ldots,T\} \end{array} \tag{6}$$

5 An Example

An example will be shown in this section to illustrate how to make Model (6) to be applied for transforming pairwise based results to the original ratings.

Suppose there are 15 instances, d_1, d_2, …, d_{15}. The ground truth value y_i is confined as $y_i \in \{1, 2, 3, 4, 5\}$ and they are

$y_1=1$, $y_2=4$, $y_3=5$, $y_4=3$, $y_5=1$, $y_6=4$, $y_7=2$, $y_8=2$, $y_9=5$, $y_{10}=5$, $y_{11}=3$, $y_{12}=3$, $y_{13}=4$, $y_{14}=3$, $y_{15}=1$

Accordingly, the ground truth relationship between each instance pairs can be derived as Table 1 shows.

Take d_1 and d_2 for example. The ground truth ratings for them are y_1=1 and y_2=4. Hence, the ground truth relationship between d_1 and d_2 is $p(d_1, d_2) = -1$, as illustrated in Table 1.

However, due to various reasons, one classifier predicts the above relationship for instance pairs with some errors. The incorrect instance pairs are:

$$\begin{array}{llllll} \hat{p}(10,6) & \hat{p}(13,1) & \hat{p}(1,8) & \hat{p}(13,4) & \hat{p}(6,8) & \hat{p}(11,3) \\ \hat{p}(14,2) & \hat{p}(1,6) & \hat{p}(15,2) & \hat{p}(1,2) & \hat{p}(2,13) & \hat{p}(3,7) \\ \hat{p}(14,8) & \hat{p}(12,1) & \hat{p}(1,15) & \hat{p}(2,5) & \hat{p}(8,15) & \hat{p}(12,15) \\ \hat{p}(11,2) & \hat{p}(8,4) & \hat{p}(8,9) & \hat{p}(9,6) & \hat{p}(5,11) & \hat{p}(11,7) \\ \hat{p}(8,13) & \hat{p}(3,5) & \hat{p}(13,3) & \hat{p}(9,4) & \hat{p}(10,2) & \hat{p}(3,2) \end{array} \tag{7}$$

However, this information is unknown, given the predicted relationship. Accordingly, the predicted relationship between each instance pair is suggested in Table 2.

Now, the objective is to assign the rating values for all instances, which is able to satisfy the predicted relationship $\hat{p}$. In other words, the assignment of the rating values for all instances should minimize the total number of violations of the predicted relationship.

Take $\hat{p}(d_3, d_1) = 1$ for example. According to Model (3), $\hat{y}_3$ and $\hat{y}_1$ should be confined like,

$$\begin{aligned}&\hat{y}_3 - \hat{y}_1 \geq 1 - M \cdot \alpha \\ &\hat{y}_1 - \hat{y}_3 \geq 1 - M \cdot (1-\alpha) \\ &\alpha \in \{0,1\} \\ &\hat{y}_1, \hat{y}_3 \in \{1,2,3,4,5\}\end{aligned} \tag{8}$$

Similarly, according to Model (4), $\hat{p}(d_1, d_4) = -1$ makes $\hat{y}_1$ and $\hat{y}_4$ should be confined like,

$$\begin{aligned}&\hat{y}_1 - \hat{y}_4 \geq 1 - M \cdot \beta \\ &\hat{y}_4 - \hat{y}_1 \geq 1 - M \cdot (1-\beta) \\ &\beta \in \{0,1\} \\ &\hat{y}_1, \hat{y}_4 \in \{1,2,3,4,5\}\end{aligned} \tag{9}$$

Also, according to Model (5), $\hat{p}(d_1, d_5) = 0$ makes $\hat{y}_1$ and $\hat{y}_5$ should be confined like,

$$\begin{aligned}&\hat{y}_1 - \hat{y}_5 \leq M \cdot \gamma \\ &\hat{y}_5 - \hat{y}_1 \leq M \cdot \gamma \\ &\gamma \in \{0,1\} \\ &\hat{y}_1, \hat{y}_5 \in \{1,2,3,4,5\}\end{aligned} \tag{10}$$

Table 1. The ground truth relationship between each instance pair.

P	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
d_1	0	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	0
d_2	1	0	-1	1	1	0	1	1	-1	-1	1	1	0	1	1
d_3	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1
d_4	1	-1	-1	0	1	-1	1	1	-1	-1	0	0	-1	0	1
d_5	0	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	0
d_6	1	0	-1	1	1	0	1	1	-1	-1	1	1	0	1	1
d_7	1	-1	-1	-1	1	-1	0	0	-1	-1	-1	-1	-1	-1	1
d_8	1	-1	-1	-1	1	-1	0	0	-1	-1	-1	-1	-1	-1	1
d_9	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1
d_{10}	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1
d_{11}	1	-1	-1	0	1	-1	1	1	-1	-1	0	0	-1	0	1
d_{12}	1	-1	-1	0	1	-1	1	1	-1	-1	0	0	-1	0	1
d_{13}	1	0	-1	1	1	0	1	1	-1	-1	1	1	0	1	1
d_{14}	1	-1	-1	0	1	-1	1	1	-1	-1	0	0	-1	0	1
d_{15}	0	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	0

Table 2. The predicted relationship between each instance pair.

$\hat{p}$	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
d_1	0	1	-1	-1	0	1	-1	1	-1	-1	-1	1	1	-1	0
d_2	-1	0	1	1	-1	0	1	1	-1	1	-1	1	0	-1	-1
d_3	1	-1	0	1	-1	1	-1	1	0	0	-1	1	-1	1	1
d_4	1	-1	-1	0	1	-1	1	-1	1	-1	0	0	1	0	1
d_5	0	1	1	-1	0	-1	-1	-1	-1	-1	1	-1	-1	-1	0
d_6	-1	0	-1	1	1	0	1	-1	1	1	1	1	0	1	1
d_7	1	-1	1	-1	1	-1	0	0	-1	-1	1	-1	-1	-1	1
d_8	-1	-1	-1	1	1	1	0	0	1	-1	-1	-1	1	1	-1
d_9	1	1	0	-1	1	-1	1	-1	0	0	1	1	1	1	1
d_{10}	1	-1	0	1	1	-1	1	1	0	0	1	1	1	1	1
d_{11}	1	1	1	0	-1	-1	-1	1	-1	-1	0	0	-1	0	1
d_{12}	-1	-1	-1	0	1	-1	1	1	-1	-1	0	0	-1	0	-1
d_{13}	-1	0	1	-1	1	0	1	-1	-1	-1	1	1	0	1	1
d_{14}	1	1	-1	0	1	-1	1	-1	-1	-1	0	0	-1	0	1
d_{15}	0	1	-1	-1	0	-1	-1	1	-1	-1	-1	1	-1	-1	0

Finally, combining three different cases, according to Model (6), an integer linear programming can be formulated for all these 15 instances. The optimization model is:

$$\min\left(\sum_{i=1}^{40}\alpha_i + \sum_{j=1}^{49}\beta_j + \sum_{k=1}^{16}\gamma_k\right)$$

Subject to:

$$\hat{y}_{3} - \hat{y}_{1} \geq 1 - M \cdot \alpha_{1} \qquad \hat{y}_{1} - \hat{y}_{3} \geq 1 - M \cdot (1 - \alpha_{1})$$
$$\hat{y}_{4} - \hat{y}_{1} \geq 1 - M \cdot \alpha_{2} \qquad \hat{y}_{1} - \hat{y}_{4} \geq 1 - M \cdot (1 - \alpha_{2})$$
$$\hat{y}_{5} - \hat{y}_{2} \geq 1 - M \cdot \alpha_{3} \qquad \hat{y}_{2} - \hat{y}_{5} \geq 1 - M \cdot (1 - \alpha_{3})$$
$$\hat{y}_{5} - \hat{y}_{3} \geq 1 - M \cdot \alpha_{4} \qquad \hat{y}_{3} - \hat{y}_{5} \geq 1 - M \cdot (1 - \alpha_{4})$$
$$\hat{y}_{6} - \hat{y}_{4} \geq 1 - M \cdot \alpha_{5} \qquad \hat{y}_{4} - \hat{y}_{6} \geq 1 - M \cdot (1 - \alpha_{5})$$
$$\hat{y}_{6} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{6} \qquad \hat{y}_{5} - \hat{y}_{6} \geq 1 - M \cdot (1 - \alpha_{6})$$
$$\hat{y}_{7} - \hat{y}_{1} \geq 1 - M \cdot \alpha_{7} \qquad \hat{y}_{1} - \hat{y}_{7} \geq 1 - M \cdot (1 - \alpha_{7})$$
$$\hat{y}_{7} - \hat{y}_{3} \geq 1 - M \cdot \alpha_{8} \qquad \hat{y}_{3} - \hat{y}_{7} \geq 1 - M \cdot (1 - \alpha_{8})$$
$$\hat{y}_{7} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{9} \qquad \hat{y}_{5} - \hat{y}_{7} \geq 1 - M \cdot (1 - \alpha_{9})$$
$$\hat{y}_{8} - \hat{y}_{4} \geq 1 - M \cdot \alpha_{10} \qquad \hat{y}_{4} - \hat{y}_{8} \geq 1 - M \cdot (1 - \alpha_{10})$$
$$\hat{y}_{8} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{11} \qquad \hat{y}_{5} - \hat{y}_{8} \geq 1 - M \cdot (1 - \alpha_{11})$$
$$\hat{y}_{8} - \hat{y}_{6} \geq 1 - M \cdot \alpha_{12} \qquad \hat{y}_{6} - \hat{y}_{8} \geq 1 - M \cdot (1 - \alpha_{12})$$
$$\hat{y}_{9} - \hat{y}_{1} \geq 1 - M \cdot \alpha_{13} \qquad \hat{y}_{1} - \hat{y}_{9} \geq 1 - M \cdot (1 - \alpha_{13})$$
$$\hat{y}_{9} - \hat{y}_{2} \geq 1 - M \cdot \alpha_{14} \qquad \hat{y}_{2} - \hat{y}_{9} \geq 1 - M \cdot (1 - \alpha_{14})$$
$$\hat{y}_{9} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{15} \qquad \hat{y}_{5} - \hat{y}_{9} \geq 1 - M \cdot (1 - \alpha_{15})$$
$$\hat{y}_{9} - \hat{y}_{7} \geq 1 - M \cdot \alpha_{16} \qquad \hat{y}_{7} - \hat{y}_{9} \geq 1 - M \cdot (1 - \alpha_{16})$$
$$\hat{y}_{10} - \hat{y}_{1} \geq 1 - M \cdot \alpha_{17} \qquad \hat{y}_{1} - \hat{y}_{10} \geq 1 - M \cdot (1 - \alpha_{17})$$
$$\hat{y}_{10} - \hat{y}_{4} \geq 1 - M \cdot \alpha_{18} \qquad \hat{y}_{4} - \hat{y}_{10} \geq 1 - M \cdot (1 - \alpha_{18})$$
$$\hat{y}_{10} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{19} \qquad \hat{y}_{5} - \hat{y}_{10} \geq 1 - M \cdot (1 - \alpha_{19})$$
$$\hat{y}_{10} - \hat{y}_{7} \geq 1 - M \cdot \alpha_{20} \qquad \hat{y}_{7} - \hat{y}_{10} \geq 1 - M \cdot (1 - \alpha_{20})$$
$$\hat{y}_{10} - \hat{y}_{8} \geq 1 - M \cdot \alpha_{21} \qquad \hat{y}_{8} - \hat{y}_{10} \geq 1 - M \cdot (1 - \alpha_{21})$$
$$\hat{y}_{11} - \hat{y}_{1} \geq 1 - M \cdot \alpha_{22} \qquad \hat{y}_{1} - \hat{y}_{11} \geq 1 - M \cdot (1 - \alpha_{22})$$
$$\hat{y}_{11} - \hat{y}_{2} \geq 1 - M \cdot \alpha_{23} \qquad \hat{y}_{2} - \hat{y}_{11} \geq 1 - M \cdot (1 - \alpha_{23})$$
$$\hat{y}_{11} - \hat{y}_{3} \geq 1 - M \cdot \alpha_{24} \qquad \hat{y}_{3} - \hat{y}_{11} \geq 1 - M \cdot (1 - \alpha_{24})$$
$$\hat{y}_{11} - \hat{y}_{8} \geq 1 - M \cdot \alpha_{25} \qquad \hat{y}_{8} - \hat{y}_{11} \geq 1 - M \cdot (1 - \alpha_{25})$$
$$\hat{y}_{12} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{26} \qquad \hat{y}_{5} - \hat{y}_{12} \geq 1 - M \cdot (1 - \alpha_{26})$$
$$\hat{y}_{12} - \hat{y}_{7} \geq 1 - M \cdot \alpha_{27} \qquad \hat{y}_{7} - \hat{y}_{12} \geq 1 - M \cdot (1 - \alpha_{27})$$
$$\hat{y}_{12} - \hat{y}_{8} \geq 1 - M \cdot \alpha_{28} \qquad \hat{y}_{8} - \hat{y}_{12} \geq 1 - M \cdot (1 - \alpha_{28})$$
$$\hat{y}_{13} - \hat{y}_{3} \geq 1 - M \cdot \alpha_{29} \qquad \hat{y}_{3} - \hat{y}_{13} \geq 1 - M \cdot (1 - \alpha_{29})$$
$$\hat{y}_{13} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{30} \qquad \hat{y}_{5} - \hat{y}_{13} \geq 1 - M \cdot (1 - \alpha_{30})$$

$$\hat{y}_{13} - \hat{y}_{7} \geq 1 - M \cdot \alpha_{31} \qquad \hat{y}_{7} - \hat{y}_{13} \geq 1 - M \cdot (1 - \alpha_{31})$$
$$\hat{y}_{13} - \hat{y}_{11} \geq 1 - M \cdot \alpha_{32} \qquad \hat{y}_{11} - \hat{y}_{13} \geq 1 - M \cdot (1 - \alpha_{32})$$
$$\hat{y}_{13} - \hat{y}_{12} \geq 1 - M \cdot \alpha_{33} \qquad \hat{y}_{12} - \hat{y}_{13} \geq 1 - M \cdot (1 - \alpha_{33})$$
$$\hat{y}_{14} - \hat{y}_{1} \geq 1 - M \cdot \alpha_{34} \qquad \hat{y}_{1} - \hat{y}_{14} \geq 1 - M \cdot (1 - \alpha_{34})$$
$$\hat{y}_{14} - \hat{y}_{2} \geq 1 - M \cdot \alpha_{35} \qquad \hat{y}_{2} - \hat{y}_{14} \geq 1 - M \cdot (1 - \alpha_{35})$$
$$\hat{y}_{14} - \hat{y}_{5} \geq 1 - M \cdot \alpha_{36} \qquad \hat{y}_{5} - \hat{y}_{14} \geq 1 - M \cdot (1 - \alpha_{36})$$
$$\hat{y}_{14} - \hat{y}_{7} \geq 1 - M \cdot \alpha_{37} \qquad \hat{y}_{7} - \hat{y}_{14} \geq 1 - M \cdot (1 - \alpha_{37})$$
$$\hat{y}_{15} - \hat{y}_{2} \geq 1 - M \cdot \alpha_{38} \qquad \hat{y}_{2} - \hat{y}_{15} \geq 1 - M \cdot (1 - \alpha_{38})$$
$$\hat{y}_{15} - \hat{y}_{8} \geq 1 - M \cdot \alpha_{39} \qquad \hat{y}_{8} - \hat{y}_{15} \geq 1 - M \cdot (1 - \alpha_{39})$$
$$\hat{y}_{15} - \hat{y}_{12} \geq 1 - M \cdot \alpha_{40} \qquad \hat{y}_{12} - \hat{y}_{15} \geq 1 - M \cdot (1 - \alpha_{40})$$
$$\hat{y}_{1} - \hat{y}_{2} \geq 1 - M \cdot \beta_{1} \qquad \hat{y}_{2} - \hat{y}_{1} \geq 1 - M \cdot (1 - \beta_{1})$$
$$\hat{y}_{2} - \hat{y}_{3} \geq 1 - M \cdot \beta_{2} \qquad \hat{y}_{3} - \hat{y}_{2} \geq 1 - M \cdot (1 - \beta_{2})$$
$$\hat{y}_{2} - \hat{y}_{4} \geq 1 - M \cdot \beta_{3} \qquad \hat{y}_{4} - \hat{y}_{2} \geq 1 - M \cdot (1 - \beta_{3})$$
$$\hat{y}_{3} - \hat{y}_{4} \geq 1 - M \cdot \beta_{4} \qquad \hat{y}_{4} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{4})$$
$$\hat{y}_{4} - \hat{y}_{5} \geq 1 - M \cdot \beta_{5} \qquad \hat{y}_{5} - \hat{y}_{4} \geq 1 - M \cdot (1 - \beta_{5})$$
$$\hat{y}_{1} - \hat{y}_{6} \geq 1 - M \cdot \beta_{6} \qquad \hat{y}_{6} - \hat{y}_{1} \geq 1 - M \cdot (1 - \beta_{6})$$
$$\hat{y}_{3} - \hat{y}_{6} \geq 1 - M \cdot \beta_{7} \qquad \hat{y}_{6} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{7})$$
$$\hat{y}_{2} - \hat{y}_{7} \geq 1 - M \cdot \beta_{8} \qquad \hat{y}_{7} - \hat{y}_{2} \geq 1 - M \cdot (1 - \beta_{8})$$
$$\hat{y}_{4} - \hat{y}_{7} \geq 1 - M \cdot \beta_{9} \qquad \hat{y}_{7} - \hat{y}_{4} \geq 1 - M \cdot (1 - \beta_{9})$$
$$\hat{y}_{6} - \hat{y}_{7} \geq 1 - M \cdot \beta_{10} \qquad \hat{y}_{7} - \hat{y}_{6} \geq 1 - M \cdot (1 - \beta_{10})$$
$$\hat{y}_{3} - \hat{y}_{8} \geq 1 - M \cdot \beta_{11} \qquad \hat{y}_{8} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{11})$$
$$\hat{y}_{4} - \hat{y}_{9} \geq 1 - M \cdot \beta_{12} \qquad \hat{y}_{9} - \hat{y}_{4} \geq 1 - M \cdot (1 - \beta_{12})$$
$$\hat{y}_{1} - \hat{y}_{8} \geq 1 - M \cdot \beta_{13} \qquad \hat{y}_{8} - \hat{y}_{1} \geq 1 - M \cdot (1 - \beta_{13})$$
$$\hat{y}_{2} - \hat{y}_{8} \geq 1 - M \cdot \beta_{14} \qquad \hat{y}_{8} - \hat{y}_{2} \geq 1 - M \cdot (1 - \beta_{14})$$
$$\hat{y}_{6} - \hat{y}_{9} \geq 1 - M \cdot \beta_{15} \qquad \hat{y}_{9} - \hat{y}_{2} \geq 1 - M \cdot (1 - \beta_{15})$$
$$\hat{y}_{8} - \hat{y}_{9} \geq 1 - M \cdot \beta_{16} \qquad \hat{y}_{9} - \hat{y}_{8} \geq 1 - M \cdot (1 - \beta_{16})$$
$$\hat{y}_{2} - \hat{y}_{10} \geq 1 - M \cdot \beta_{17} \qquad \hat{y}_{10} - \hat{y}_{4} \geq 1 - M \cdot (1 - \beta_{17})$$
$$\hat{y}_{6} - \hat{y}_{10} \geq 1 - M \cdot \beta_{18} \qquad \hat{y}_{10} - \hat{y}_{6} \geq 1 - M \cdot (1 - \beta_{18})$$
$$\hat{y}_{5} - \hat{y}_{11} \geq 1 - M \cdot \beta_{19} \qquad \hat{y}_{11} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{19})$$
$$\hat{y}_{6} - \hat{y}_{11} \geq 1 - M \cdot \beta_{20} \qquad \hat{y}_{11} - \hat{y}_{6} \geq 1 - M \cdot (1 - \beta_{20})$$
$$\hat{y}_{7} - \hat{y}_{11} \geq 1 - M \cdot \beta_{21} \qquad \hat{y}_{11} - \hat{y}_{4} \geq 1 - M \cdot (1 - \beta_{21})$$
$$\hat{y}_{9} - \hat{y}_{11} \geq 1 - M \cdot \beta_{22} \qquad \hat{y}_{11} - \hat{y}_{9} \geq 1 - M \cdot (1 - \beta_{22})$$
$$\hat{y}_{10} - \hat{y}_{11} \geq 1 - M \cdot \beta_{23} \qquad \hat{y}_{11} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{23})$$
$$\hat{y}_{1} - \hat{y}_{12} \geq 1 - M \cdot \beta_{24} \qquad \hat{y}_{12} - \hat{y}_{1} \geq 1 - M \cdot (1 - \beta_{24})$$
$$\hat{y}_{2} - \hat{y}_{12} \geq 1 - M \cdot \beta_{25} \qquad \hat{y}_{12} - \hat{y}_{2} \geq 1 - M \cdot (1 - \beta_{25})$$
$$\hat{y}_{3} - \hat{y}_{12} \geq 1 - M \cdot \beta_{26} \qquad \hat{y}_{12} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{26})$$
$$\hat{y}_{6} - \hat{y}_{12} \geq 1 - M \cdot \beta_{27} \qquad \hat{y}_{12} - \hat{y}_{6} \geq 1 - M \cdot (1 - \beta_{27})$$
$$\hat{y}_{9} - \hat{y}_{12} \geq 1 - M \cdot \beta_{28} \qquad \hat{y}_{12} - \hat{y}_{9} \geq 1 - M \cdot (1 - \beta_{28})$$
$$\hat{y}_{10} - \hat{y}_{12} \geq 1 - M \cdot \beta_{29} \qquad \hat{y}_{12} - \hat{y}_{10} \geq 1 - M \cdot (1 - \beta_{29})$$
$$\hat{y}_{1} - \hat{y}_{13} \geq 1 - M \cdot \beta_{30} \qquad \hat{y}_{13} - \hat{y}_{1} \geq 1 - M \cdot (1 - \beta_{30})$$
$$\hat{y}_{4} - \hat{y}_{13} \geq 1 - M \cdot \beta_{31} \qquad \hat{y}_{13} - \hat{y}_{4} \geq 1 - M \cdot (1 - \beta_{31})$$
$$\hat{y}_{8} - \hat{y}_{13} \geq 1 - M \cdot \beta_{32} \qquad \hat{y}_{13} - \hat{y}_{8} \geq 1 - M \cdot (1 - \beta_{32})$$
$$\hat{y}_{9} - \hat{y}_{13} \geq 1 - M \cdot \beta_{33} \qquad \hat{y}_{13} - \hat{y}_{9} \geq 1 - M \cdot (1 - \beta_{33})$$
$$\hat{y}_{10} - \hat{y}_{13} \geq 1 - M \cdot \beta_{34} \qquad \hat{y}_{13} - \hat{y}_{10} \geq 1 - M \cdot (1 - \beta_{34})$$
$$\hat{y}_{3} - \hat{y}_{14} \geq 1 - M \cdot \beta_{35} \qquad \hat{y}_{14} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{35})$$
$$\hat{y}_{6} - \hat{y}_{14} \geq 1 - M \cdot \beta_{36} \qquad \hat{y}_{14} - \hat{y}_{6} \geq 1 - M \cdot (1 - \beta_{36})$$
$$\hat{y}_{8} - \hat{y}_{14} \geq 1 - M \cdot \beta_{37} \qquad \hat{y}_{14} - \hat{y}_{8} \geq 1 - M \cdot (1 - \beta_{37})$$
$$\hat{y}_{9} - \hat{y}_{14} \geq 1 - M \cdot \beta_{38} \qquad \hat{y}_{14} - \hat{y}_{9} \geq 1 - M \cdot (1 - \beta_{38})$$
$$\hat{y}_{10} - \hat{y}_{14} \geq 1 - M \cdot \beta_{39} \qquad \hat{y}_{14} - \hat{y}_{10} \geq 1 - M \cdot (1 - \beta_{39})$$
$$\hat{y}_{13} - \hat{y}_{14} \geq 1 - M \cdot \beta_{40} \qquad \hat{y}_{14} - \hat{y}_{13} \geq 1 - M \cdot (1 - \beta_{40})$$
$$\hat{y}_{3} - \hat{y}_{15} \geq 1 - M \cdot \beta_{41} \qquad \hat{y}_{15} - \hat{y}_{3} \geq 1 - M \cdot (1 - \beta_{41})$$
$$\hat{y}_{4} - \hat{y}_{15} \geq 1 - M \cdot \beta_{42} \qquad \hat{y}_{15} - \hat{y}_{4} \geq 1 - M \cdot (1 - \beta_{42})$$
$$\hat{y}_{6} - \hat{y}_{15} \geq 1 - M \cdot \beta_{43} \qquad \hat{y}_{15} - \hat{y}_{6} \geq 1 - M \cdot (1 - \beta_{43})$$
$$\hat{y}_{7} - \hat{y}_{15} \geq 1 - M \cdot \beta_{44} \qquad \hat{y}_{15} - \hat{y}_{7} \geq 1 - M \cdot (1 - \beta_{44})$$
$$\hat{y}_{9} - \hat{y}_{15} \geq 1 - M \cdot \beta_{45} \qquad \hat{y}_{15} - \hat{y}_{9} \geq 1 - M \cdot (1 - \beta_{45})$$
$$\hat{y}_{10} - \hat{y}_{15} \geq 1 - M \cdot \beta_{46} \qquad \hat{y}_{15} - \hat{y}_{10} \geq 1 - M \cdot (1 - \beta_{46})$$
$$\hat{y}_{11} - \hat{y}_{15} \geq 1 - M \cdot \beta_{47} \qquad \hat{y}_{15} - \hat{y}_{11} \geq 1 - M \cdot (1 - \beta_{47})$$
$$\hat{y}_{13} - \hat{y}_{15} \geq 1 - M \cdot \beta_{48} \qquad \hat{y}_{15} - \hat{y}_{13} \geq 1 - M \cdot (1 - \beta_{48})$$
$$\hat{y}_{14} - \hat{y}_{15} \geq 1 - M \cdot \beta_{49} \qquad \hat{y}_{15} - \hat{y}_{14} \geq 1 - M \cdot (1 - \beta_{49})$$
$$\hat{y}_{1} - \hat{y}_{5} \leq M \cdot \gamma_{1} \qquad \hat{y}_{5} - \hat{y}_{1} \leq M \cdot \gamma_{1}$$
$$\hat{y}_{2} - \hat{y}_{6} \leq M \cdot \gamma_{2} \qquad \hat{y}_{6} - \hat{y}_{2} \leq M \cdot \gamma_{2}$$
$$\hat{y}_{7} - \hat{y}_{8} \leq M \cdot \gamma_{3} \qquad \hat{y}_{8} - \hat{y}_{7} \leq M \cdot \gamma_{3}$$
$$\hat{y}_{3} - \hat{y}_{9} \leq M \cdot \gamma_{4} \qquad \hat{y}_{9} - \hat{y}_{3} \leq M \cdot \gamma_{4}$$
$$\hat{y}_{3} - \hat{y}_{10} \leq M \cdot \gamma_{5} \qquad \hat{y}_{10} - \hat{y}_{3} \leq M \cdot \gamma_{5}$$
$$\hat{y}_{9} - \hat{y}_{10} \leq M \cdot \gamma_{6} \qquad \hat{y}_{10} - \hat{y}_{9} \leq M \cdot \gamma_{6}$$
$$\hat{y}_{4} - \hat{y}_{11} \leq M \cdot \gamma_{7} \qquad \hat{y}_{11} - \hat{y}_{4} \leq M \cdot \gamma_{7}$$
$$\hat{y}_{4} - \hat{y}_{12} \leq M \cdot \gamma_{8} \qquad \hat{y}_{12} - \hat{y}_{4} \leq M \cdot \gamma_{8}$$
$$\hat{y}_{11} - \hat{y}_{12} \leq M \cdot \gamma_{9} \qquad \hat{y}_{12} - \hat{y}_{11} \leq M \cdot \gamma_{9}$$
$$\hat{y}_{2} - \hat{y}_{13} \leq M \cdot \gamma_{10} \qquad \hat{y}_{13} - \hat{y}_{2} \leq M \cdot \gamma_{10}$$
$$\hat{y}_{6} - \hat{y}_{13} \leq M \cdot \gamma_{11} \qquad \hat{y}_{13} - \hat{y}_{6} \leq M \cdot \gamma_{11}$$
$$\hat{y}_{4} - \hat{y}_{14} \leq M \cdot \gamma_{12} \qquad \hat{y}_{14} - \hat{y}_{4} \leq M \cdot \gamma_{12}$$
$$\hat{y}_{11} - \hat{y}_{14} \leq M \cdot \gamma_{13} \qquad \hat{y}_{14} - \hat{y}_{11} \leq M \cdot \gamma_{13}$$
$$\hat{y}_{12} - \hat{y}_{14} \leq M \cdot \gamma_{14} \qquad \hat{y}_{14} - \hat{y}_{12} \leq M \cdot \gamma_{14}$$
$$\hat{y}_{1} - \hat{y}_{15} \leq M \cdot \gamma_{15} \qquad \hat{y}_{15} - \hat{y}_{1} \leq M \cdot \gamma_{15}$$
$$\hat{y}_{5} - \hat{y}_{15} \leq M \cdot \gamma_{16} \qquad \hat{y}_{15} - \hat{y}_{5} \leq M \cdot \gamma_{16}$$
$$\alpha_1, \alpha_2, \ldots, \alpha_{40}, \beta_1, \beta_2, \ldots, \beta_{49}, \gamma_1, \gamma_2, \ldots, \gamma_{16} \in \{0,1\}$$
$$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_{15} \in \{1,2,3,4,5\}$$

This optimization problem can be solved by some optimization modeling software, such as Lingo, CPLEX, etc. Accordingly, the predicted rating values for the 15 instances can be obtained. They are:

$$\hat{y}_1 = 1, \hat{y}_2 = 5, \hat{y}_3 = 4, \hat{y}_4 = 3, \hat{y}_5 = 1,$$
$$\hat{y}_6 = 5, \hat{y}_7 = 2, \hat{y}_8 = 2, \hat{y}_9 = 4, \hat{y}_{10} = 4,$$
$$\hat{y}_{11} = 3, \hat{y}_{12} = 3, \hat{y}_{13} = 5, \hat{y}_{14} = 3, \hat{y}_{15} = 1$$

Comparing with the ground truth values of these 15 instances, the predicted values of d_2, d_3, d_6, d_9, d_{10}, and d_{13} are errors.

6 Conclusions

Pairwise approaches are widely utilized in the field of information retrieval and data mining. However, these approaches are not conventionally evaluated by standard evaluation metrics. What is desired is to transform pairwise based results to the original ratings.

In this research, an integer linear programming model is formulated for the problem. The objective of this model is to find how to assign the rating values for all instances, which makes the assignment to minimize the total number of violations for the predicted pairwise based results. Finally, an example with 15 instances is presented in order to illustrate the details about the proposed model.

Acknowledgments

The authors acknowledge The Hong Kong Polytechnic University for the financial support of the project (G-YL27). The work described in this paper was also supported by a grant from the National Nature Science Foundation of Chinese (project no. NSFC 71271185/G011201).

References

[1] Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In ICML'05. ACM, 2005.

[2] Yoav Freund, Raj Lyer, Robert E. Schapire, and Yoram Singer. An efficient boosting algorithm for combining preferences. The Journal of Machine Learning Research, 4:933--969, 2003.

[3] Ralf Herbrich, Thore Graepel, and Klaus Obermayer. Large margin rank boundaries for ordinal regression, page 115–132. MIT Press, Cambridge, MA, 2000.

[4] Jian Jin, Ping Ji, and Ying Liu. Product Characteristic Weighting for Designer from Online Reviews: An Ordinal Classification Approach. In EDBT'12 workshop: BEWEB. ACM, 2012

[5] Thorsten Joachims. Optimizing search engines using clickthrough data. In KDD'02, pages 133--142, 2002.

[6] Thorsten Joachims, Hang Li, Tie-Yan Liu, and Chengxiang Zhai. Learning to rank for information retrieval (LR4IR 2007). SIGIRForum, 41:58--62, December 2007.

[7] Tie-Yan Liu. Learning to Rank for Information Retrieval, Foundations and Trends in Information Retrieval. Now Publishers inc, Hanover, MA, 2009.

[8] Leonardo Rigutini, Tiziano Papini, Marco Maggini, and Franco Scarselli. Sortnet: Learning to rank by a neural preference function. IEEE Transactions on Neural Networks, 22:1368--1380, 2011.

[9] Tao Qin, Xu-Dong Zhang, De-Sheng Wang, Tie-Yan Liu, Wei Lai, and Hang Li. Ranking with multiple hyperplanes. In SIGIR'07, pages 279--286, 2007.