Ranking Social Emotions by Learning Listwise Preference

Qishen Wang[1] Ou Wu[2] Weiming Hu[2] Jinfeng Yang[1] Wanqing Li[3]

[1]Aeronautical Automation College
Civil Aviation University of China
Tianjin, China
(qswang_yjs09, jfyang)@cauc.edu.cn

[2]Institute of Automation
Chinese Academy of Sciences
Beijing, China
(wuou, wmhu)@nlpr.ia.ac.cn

[3]Information and Communication Technology RI
University of Wollongong
NSW 2522, Australia
wanqing@uow.edu.cn

Abstract—**Emotion modeling has received a great attention in recent years. This paper models the online social emotions that are the online users' emotional responds when they are exposed to news articles. Specifically, we rank social emotion labels for online documents. Unlike the existing method, referred to as Pair-LR, which learns pairwise preference and adopts binary classification, we address the problem of ranking social emotions by learning listwise preference. In particular, a novel approach, referred to as List-LR, is proposed to learn a ranking model for social emotion labels of online documents by minimizing the listwise loss defined on instances. Empirical experiments show that the proposed approach outperforms Pair-LR and is also competitive to other two start-of-the-art approaches for label ranking.**

Keywords-social emotions; label ranking; listwise preference;

I. INTRODUCTION

Emotion analyzing and modeling for online documents have attracted a considerable attention in natural language processing and machine learning. It is a powerful automatic tool in an opinion poll on what attitude people would take toward an event or a product. In previous literatures, most studies focus on classification of emotions from the perspectives of the writers. In recent years, preliminary studies on emotion analysis from the readers' perspectives have been conducted [1]. These studies focus on the online users' emotional responses when they are exposed to news articles. This kind of emotional responses is called social emotions [2]. Modeling social emotions has many potential applications including emotions-based document retrieval by analyzing and ranking social emotions of individuals triggered by online news events. In general, not all of the users would demonstrate the same emotional response to a given news article, but rather their responses are distributed over and can be better described by a set of emotion labels, such as happy, sadness, touched, angry, funny and boredom as shown in Fig. 1. Therefore, it is better to provide a ranking of emotions according to their popularity rather than associating a single emotion label with a document [3]. This is a typical label ranking problem, in which the task is to learn a mapping from an online document to a ranked list of social emotion labels, for instance, 'angry $\succ$ sadness $\succ$ boredom $\succ$ touched $\succ$ funny $\succ$ happy', where '$i \succ j$' represents label i ranks higher than label j.

Figure 1. An instance of online users' votes on social emotions. The emotions are happy, sadness, touched, angry, funny, boredom. (from left to right)

Study of social emotion ranking is still in its infancy and, to our best knowledge, there is only one study reported so far. Lin et al. [3] proposed a pairwise approach to rank readers' emotions. The approach formulates the problem of ranking as that of classification and learns a binary classifier over each pair of emotion labels. Essentially, it is a label ranking approach based on pairwise preference. We refer to this approach as Pair-LR in this paper. One advantage of the Pair-LR approach is that existing binary classification techniques can be used directly. However, the pairwise approach suffers from some problems. Firstly, the objective of learning pairwise preference only minimizes the pairwise loss that is defined as the cost of the mis-classified emotion pairs. While the fact that social emotion ranking is a predicative task on a list of emotion labels. So it's better to minimize the listwise loss defined as the cost of the mis-ranked emotion lists. Secondly, the imbalanced data distribution when constructing a binary classifier on each pair of emotion labels will result in the model biased toward the labels with more examples.

In this paper, we develop a novel learning algorithm that ranks social emotion labels by learning listwise preference to address the drawbacks of the Pair-LR approach. In our learning scenario, each online document and a list of preference over a predefined set of social emotion labels are used as a training unit. The crucial task is to construct a listwise loss function representing the difference between the list of preference output by a ranking model and the list of preference given as ground truth. In order to calculate the listwise loss function, a probability model representing the ranking lists of labels is proposed. By optimizing the listwise loss function, we learn a ranking model that induces a total order over the predefined set of social emotion labels. We detail the listwise label ranking in Section III.

The remainder of the paper is organized as follows. Section II provides a brief overview of the related work. In Section III, we detail the proposed List-LR method. Experimental results

This work was carried out at the Institute of Automation, Chinese Academy of Sciences.

and discussions are presented in Section IV. Section V concludes this paper and gives the future work.

II. RELATED WORK

Since little study has been formally reported so far on social emotion ranking, in this section, we briefly review the related topic: label ranking. Label ranking is a relatively novel research branch in machine learning. A number of methods have been developed in previous literatures.

One class of the popular approaches for label ranking is to consider the problem of label ranking as that of classification and learn pairwise preference [4], which has been developed and successfully applied to social emotion ranking. In this learning scenario, for each pair of labels (l_i, l_j), a base learner M_{ij} that predicts whether $l_i \succ_x l_j$ or $l_j \succ_x l_i$ for an input instance x is trained. Thus, a total number of $m(m\text{-}1)/2$ learners are needed. Given an instance x, the learner decides whether $l_i \succ_x l_j$ or $l_j \succ_x l_i$. Specifically, M_{ij} can be implemented as a binary classifier that outputs 1 if $l_i \succ_x l_j$ and -1 otherwise. The final order over all labels is obtained by combining the preferences on each pair of labels. In particular, in social emotion ranking, Lin et al. [3] use Support Vector Machines (SVM) as classification model M_{ij} for each pair of emotion labels (i, j). Then a total order on the set of emotion labels is induced by combining the results of these individual SVM models. Johannes et al. [5] use a decision tree to learn a model for each pairwise preference. Fürnkranz [6] uses pairwise classification to reduce the problem of multiclass prediction to learning binary classifiers. Klaus et al. [4] give a systematic overview of label ranking by learning pairwise preference.

Recently, Cheng et al. [7] propose an instance-based label ranking method that adopts the Mallows model [8] to describe the probability of an instance's label order in conditioning to its neighbors and then pursues the order with the highest probability. Dekel et al. [9] introduce a log-linear algorithm for label ranking, in which the label order for each instance is represented by a preference graph and a loss function is defined on the preference graph. Then, a log-linear model is pursued using a boosting framework with the defined loss. A detailed survey on label ranking can be found in [10].

III. METHODOLOGIES

A. General Framwork

In this subsection, we give a general framework of social emotion ranking by learning listwise preference. In particular, a complete (total strict order) label ranking task is considered in this paper.

Under the framework, the core task is to learn a mapping from an online document d of an online document space D to ranking $\succ_d$ (total strict order) over a predefined set of social emotion labels $L = \{l(1), l(2), \ldots, l(m)\}$, where $l(i) \succ_d l(j)$ means that the emotion label $l(i)$ is deemed to be more relevant to document d and is ranked higher than the emotion label $l(j)$ that is considered to be less relevant. In training, each document d_i is associated with a list of emotion scores $y^i = (y^i_1, y^i_2, \ldots, y^i_m)$ over L, where d_i□D, i = 1,2, …, n, and n denotes the sizes of D; y^i_j denotes the relevance score of the emotion label $l(j)$ to the document d_i. A feature vector x^i can be created from the online document d_i. Then x^i and the list of emotion scores y^i corresponding to d_i form a 'train unit'. The emotion scores that we consider in this paper are referred to as the normalized users' votes on social emotion labels, which are detailed in Section IV.

Specifically, we aim to learn a label ranking model F that consists of a set of the basic emotion score functions, i.e., $F = \{f_{l(1)}, f_{l(2)}, \ldots, f_{l(m)}\}$. It is one-to-one correspondence with the set of the predefined emotion labels $L = \{l(1), l(2), \ldots, l(m)\}$. Let X be a feature vector space, for $x^i \in X$, the function $f_{l(j)}$: $X \rightarrow R$ assigns a relevance score to the corresponding emotion label $l(j)$. When predicting, the emotion labels are to be ranked according to the scores. Given a feature vector x^i corresponding to d_i, the ranking model F outputs a list of emotion scores $s^i = (f_{l(1)}(x^i), f_{l(2)}(x^i), \ldots, f_{l(m)}(x^i))$ over L. Then the model F is obtained by minimizing the following total listwise losses:

$$\sum_{i=1}^{n} loss(y^i, s^i) = \sum_{i=1}^{n} loss((y^i_1, \ldots, y^i_m), (f_{l(1)}(x^i), \ldots, f_{l(m)}(x^i))) \quad (1)$$

where $loss$ represents the difference between the ranking list s^i predicted by ranking model and the ground truth ranking list y^i, which is called listwise loss function. In the following subsection, we introduce a probability model to calculate the loss function.

B. Probability Model

Probability models have been used to represent a ranking list in previous literatures, such as the Plackett-Luce (PL) model [11], which is a parameterized probability distribution over ranking lists of objects, and have been studied in information retrieval and machine learning [12, 13]. The PL model seems to be a more appropriate probability model for representing a ranked label list [12]. So in this paper, we also propose using a similar model for representing the ranking list of emotion label scores as a probability to calculate the listwise loss function in Eq. (1). We refer the model as label permutation probability model (*L-PM*).

we assume that there are m emotion labels to be ranked. An emotion ranker can assign a relevance score to each emotion label, and the ranker outputs a list of emotion label scores according to the scores. The list of scores is denoted as $s = (s_1, s_2, \ldots, s_m)$, where s_j is the score of the *j-th* emotion label. We use $\psi = (\psi(1), \psi(2), \ldots, \psi(m))$ to represent a permutation of the m emotion labels, where $\psi(k)$ is the emotion label at the position k in the permutation ψ. The set of all possible permutations consisting of m emotion labels is denoted as Ω_m. Given a permutation ψ□Ω_m, its probability corresponding to the list of emotion scores s is defined as

$$p(\psi \mid s) = \prod_{k=1}^{m} \frac{\exp(s_{\psi(k)})}{\exp(s_{\psi(k)}) + \exp(s_{\psi(k+1)}) + \ldots + \exp(s_{\psi(m)})} \quad (2)$$

where $s_{\psi(k)}$ is the score of the emotion label at the position k of the permutation ψ. Eq. (2) is referred to as an *L-PM* model.

C. Learning Algorithm

As stated in the section of III-A, the objective of learning is to create a label ranking model F that can form a list of emotion scores over L by assigning a relevance score to each emotion label, and then the emotion labels can be ranked according to the scores in prediction. We now describe how the model F is obtained.

In our method, a linear neural network model without the constant b is employed as the basic emotion score function, i.e., $f_{l(j)}(x^i) = \langle w_{l(j)}, x^i \rangle$, where each $w_{l(j)}$ is a parameter vector to be learned and $\langle \cdot, \cdot \rangle$ denotes inner product. Given a feature vector x^i corresponding to d_i, a list of emotion scores $s^i = (f_{l(1)}(x^i), f_{l(2)}(x^i), \ldots, f_{l(m)}(x^i)) = (\langle w_{l(1)}, x^i \rangle, \langle w_{l(2)}, x^i \rangle, \ldots, \langle w_{l(m)}, x \rangle)$ can be obtained by F. According to the *L-PM* model proposed in the section of III-B, for $\forall\ \psi \in \Omega_m$, the probability of the permutation ψ corresponding to s^i can be calculated as

$$P(\psi \mid s^i) = P((\psi(1), \psi(2), \ldots, \psi(m)) \mid (f_1(x^i), f_2(x^i), \ldots, f_m(x^i)))$$
$$= P((\psi(1), \psi(2), \ldots, \psi(m)) \mid (f_{\psi(1)}(x^i), f_{\psi(2)}(x^i), \ldots, f_{\psi(m)}(x^i)))$$
$$= \prod_{k=1}^{m} \frac{\exp(f_{\psi(k)}(x^i))}{\exp(f_{\psi(k)}(x^i)) + \exp(f_{\psi(k+1)}(x^i)) +, \ldots, \exp(f_{\psi(m)}(x^i))}$$
$$= \prod_{k=1}^{m} \frac{\exp(\langle w_{\psi(k)}, x^i \rangle)}{\exp(\langle w_{\psi(k)}, x^i \rangle) + \exp(\langle w_{\psi(k+1)}, x^i \rangle) +, \ldots, + \exp(\langle w_{\psi(m)}, x^i \rangle))}$$

Similarly, the probability of the permutation ψ corresponding to the ground truth list of emotion scores y^i is calculated as

$$P(\psi \mid y^i) = P((\psi(1), \psi(2), \ldots, \psi(m) \mid (y_1^i, y_2^i, \ldots, y_m^i)))$$
$$= P((\psi(1), \psi(2), \ldots, \psi(m)) \mid (y_{\psi(1)}^i, y_{\psi(2)}^i, \ldots, y_{\psi(m)}^i))$$
$$= \prod_{k=1}^{m} \frac{\exp(y_{\psi(k)}^i)}{\exp(y_{\psi(k)}^i) + \exp(y_{\psi(k+1)}^i) +, \ldots, + \exp(y_{\psi(m)}^i)}$$

Given two lists of emotion score s^i outputted by the ranking model and y^i the ground truth, corresponding two probability distributions over $\mathbf{\Omega}_m$ can be obtained. Then we measure the difference between the two probability distributions. To make the measurement be a metric, we adopt Cross Entropy. The listwise loss function is then written as

$$loss(y^i, s^i) = -\sum_{\psi \in \Omega_m} P(\psi \mid y^i) \log(P(\psi \mid s^i))$$

The total losses in Eq. (1) become as follows

$$\sum_{i=1}^{n} loss(y^i, s^i) = \sum_{i=1}^{n} (-\sum_{\psi \in \Omega_m} P(\psi \mid y^i) \log(P(\psi \mid s^i))) \tag{3}$$
$$= -\sum_{i=1}^{n} \sum_{\psi \in \Omega_m} (\prod_{k=1}^{m} \frac{\exp(y_{\psi(k)}^i)}{\exp(y_{\psi(k)}^i) + \exp(y_{\psi(k+1)}^i) +, \ldots, + \exp(y_{\psi(m)}^i)}$$
$$\times \log(\prod_{k=1}^{m} \frac{\exp(\langle w_{\psi(k)}, x^i \rangle)}{\exp(\langle w_{\psi(k)}, x^i \rangle) + \exp(\langle w_{\psi(k+1)}, x^i \rangle) +, \ldots, + \exp(\langle w_{\psi(m)}, x^i \rangle))}))$$

The model's parameters $\{w_{l(1)}, w_{l(2)}, \ldots, w_{l(m)}\}$ are obtained by minimizing the above total losses. To this end, first the gradient of the listwise loss function with respect to each parameter $w_{l(j)}$ is calculated as

$$\Delta w_{l(j)} = \frac{\partial loss(y^i, s^i)}{\partial w_{l(j)}} = -\sum_{\forall \psi \in \Omega_m} \frac{P(\psi \mid y^i)}{P(\psi \mid s^i)} \frac{\partial P(\psi \mid s^i)}{\partial w_{\psi(k)}}, \tag{4}$$
$$subject \quad to: j = 1, 2, \ldots, m; \quad l(j) = \psi(k)$$

Then, based on (4), we utilize a Gradient Descent method to optimize the total losses defined in Eq. (3). The learning algorithm, named as List-LR, is described in Algorithm 1.

Algorithm 1: List-LR

Input: Learning rate ζ and number of iterations T.
Training data (x^i, y^i), i=1, 2, …, n.
Initialized model parameters $\{w_{l(1)}, w_{l(2)}, \ldots, w_{l(m)}\}$.
Optimization:
For t=1 to T
For i=1 to n
Input x^i to the ranking model and generate s^i. Eq. (1) can be compute by L-PM model. Then compute gradient $\Delta w_{l(j)}$ using Eq.(4).
Update $w_{l(j)} = w_{l(j)} - \zeta \times \Delta w_{l(j)}$
End for
End for
Output: $\{w_{l(1)}, w_{l(2)}, \ldots, w_{l(m)}\}$

D. Differences from ListNet

The proposed method is mainly inspired by ListNet [13] used in information retrieval to rank documents. However, as they are designed for different learning scenarios, the former is for label ranking and the latter is for object ranking, there are distinct differences between them. Firstly, the training units are different. In our method, a training unit consists of an instance x and a list of scores over predefined labels, while in ListNet, the train unit consists of a list of instances and a list of scores corresponding to the list of instances. Secondly, the objectives of learning are different. In our method, we aim to learn a model that predict an order over the predefined set of labels for each instance, while in ListNet, the aim is to predict an order over the set of instances. Thirdly, the composition of the ranking models is different. The model in our method consists of a set of functions, while in ListNet, the model is only a single ranking function. In other words, the ranking model in our method is associated with m parameter vectors ($w_{l(1)}$, $w_{l(2)}$, …, $w_{l(m)}$), while in ListNet, the ranking model is associated with only one parameter vector w.

IV. EXPERIMENTS

In this section, we report the experiments conducted to evaluate the performance of proposed List-LR. The List-LR is empirically compared with the existing social emotion ranking method Pair-LR [3]. Additionally, to further evaluate the performance of the proposed method, we also compare the List-LR with the other state-of-the-art label ranking algorithms proposed by [7] and [9], which are referred to as IB-Mal and Lin-LL respectively.

A. Experimental Setup

We collected 3,031 Chinese news articles dated between May 26, 2010 and December 31, 2010 from the web site http://www.sohu.com/. Besides the headline and content of the news articles, each collected sample has the associated online users' voting on emotions to show how they respond to the news article emotionally. We normalized the number of the voting for each emotion label by the total number of votes for a document as the corresponding emotion score. Then for each sample, a list of scores over the predefined set of six emotion labels was formed. The six labels are happy, sadness, touched, angry, funny and boredom, respectively. The users' votes imply an order over the six emotion labels. For example, Fig. 1

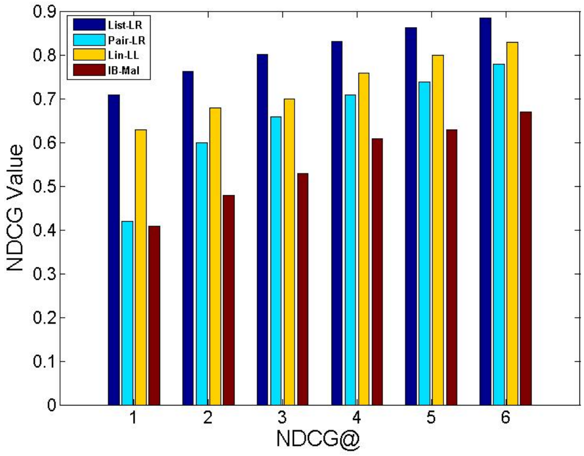

Figure 2. Ranking accuracies using 200-dimensional features

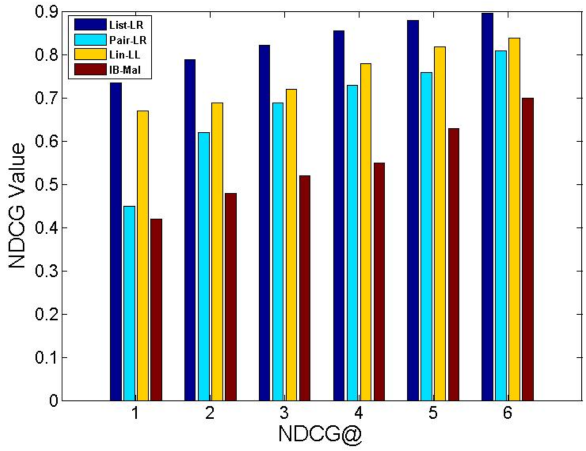

Figure 3. Ranking accuracies using 500-dimensional features

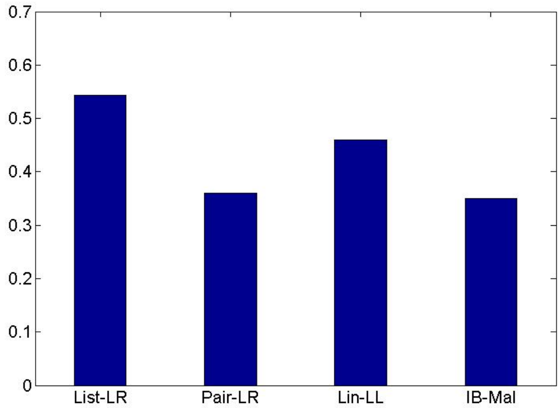

Figure 4. Ranking accuracies in terms of *MAP* using 200-dimensional features

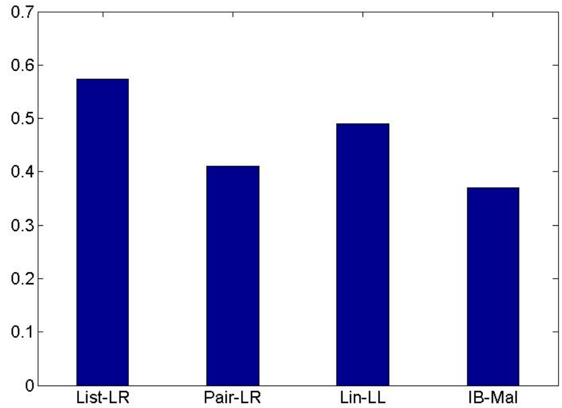

Figure 5. Ranking accuracies in terms of *MAP* using 500-dimensional features

shows an example of users' votes when reading the news article addressed by the URL[1] for which the order is ψ: 'funny $\succ$ angry $\succ$ boredom $\succ$ happy $\succ$ sadness $\succ$ touched'.

As for the features, we extracted Chinese words formed by a Chinese segmentation tool from the title and content of the news articles. Then 200-dimensional features and 500-dimensional features were selected based on the document frequency (*DF*) respectively. Stop words were removed in the data.

In the experiments, we randomly divided the examples into three subsets and performed 3-fold cross-validation trials. The parameters were set based on cross-validations that were performed solely on training set. The average results over three trials are reported.

B. *Evaluation Metrics for Label Ranking*

We use *NDCG@k* and *MAP* to evaluate the quality of a ranked list of emotion labels based on the relevance between the document and the emotion labels.

NDCG@k which stands for the normalized discounted cumulative gain at the position *k* in a ranked list is used to measure the quality of a ranked list when more than two levels of relevance are taken into account. Given a predicted ranked list of labels, the *NDCG@k* is defined as

$$NDCG@k = Z_k \sum_{j=1}^{k} \frac{2^{r_j}-1}{\log_2(1+j)}$$

where r_j is the relevance score of the label at the position j in a ranked list. Z_k is a normalizing factor that ensures that a perfect ranked list has an *NDCG@k* value of 1. The *NDCG@k* values reported in this paper are the averaged results over the whole testing set.

MAP (the mean of average precision) is used to measure the quality of a ranked list when there are only two levels: relevant and irrelevant. In calculation of *MAP*, the top-2 labels in the ground truth ranked list are treated as relevant and others as irrelevant. The metric is defined as

$$MAP = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{1}{M_{rel}}\frac{M_{rel}(j)}{M_{rel}(j)+M_{irrel}(j)}I(j)$$

where for given an instance x and the corresponding predicted ranked list ψ, j denotes the position in ψ. M is the number of labels. M_{rel} denotes the total number of labels relevant to the instance. $M_{rel}(j)$ is the number of relevant labels in the top-j position of ψ and $M_{irrel}(j)$ is the number of irrelevant labels. $I(j)$ takes on 1 if the th-j label is relevant to x, otherwise 0. N is the total number of instances in the testing set.

C. *Results and Discussions*

Fig. 2 to Fig. 5 show the performances of the different approaches in terms of *NDCG@k* and *MAP* when 200-dimensional features and 500-dimensional features were used respectively.

We empirically compare the proposed List-LR method with the existing Pair-LR method. Fig. 2 to Fig. 5 show that

[1] http://news.sohu.com/20100730/n273868884.shtml

List-LR is better to rank social emotion labels with higher scores in the top of the list. The ranked lists predicted by the List-LR approach are closer to the ground truth. We analyze the reason why the List-LR approach outperforms the Pair-LR as follows: Firstly, the Pair-LR is a label ranking approach by learning pairwise preference. So it can only minimize the cost of the mis-classified emotion pairs rather than minimize the cost of the mis-ranked emotion lists. In the List-LR, we take an instance x and the corresponding lists of preference over the predefined set of labels as a learning unit, construct a listwise loss function on each instance, and minimize the cost of the mis-ranked lists of emotions. By leaning the listwise preference, the proposed approach is able to effectively capture the correlation among the labels. Secondly, for the Pair-LR, the imbalanced data distribution when constructing binary classifier on each pair of labels will result in the model biased toward the label with more examples. Table I shows that the distributions of the number of positive examples and negative examples over each emotion label pair. We can see that the distributions are skewed for some classifiers. However, the trouble does not exist in List-LR.

Experimental results as shown in Fig. 2 to Fig 5 also demonstrate that the proposed approach is better than Lin-LL and IB-Mal methods. The reasons are as follows: Firstly, compared with Lin-LL, the List-LR directly optimizes a listwise loss function defined on the instance, which directly measures the difference between the list outputted by model and the ground truth. While Lin-LL optimizes a loss function defined on the preference graph. The graph is decomposed into elementary subgraphs. Essentially, their loss function is still based on mis-ordered label pairs. So Lin-LL is also incapable of minimizing the cost of the mis-ordered label lists. Secondly, for IB-Mal, it is an instance-based method and cannot minimize the listwise loss obviously. And since social emotion ranking is a complete label ranking task, it shows the worst result among the competing approaches although performing well in incomplete ranking according to [7].

In all, the experimental results show that the proposed List-LR outperforms all other methods in terms of the both indicators *NDCG@k* and *MAP*. Although the complexity of List-LR is $O(m! \cdot n)$ and the calculation might be intractable in theory, it still practically feasible because the number of labels to be ranked for each instance is usually small in social emotion ranking. For example, the m value was set to 6 in our experiments. We implemented the proposed method on a computer with 2.53*GHz* and 2.00*GB* memory space, and each iteration is around 1.5 *min*. In Fig. 6, it is seen that *NDCG@*6 reaches its limit when the iterative number is up to about 20, which shows that the time complexity of the proposed approach for social emotion label ranking is tractable.

V. Conclusions

In this paper, we have proposed a novel label ranking method to rank social emotions, referred to as List-LR. In contrast to existing methods, List-LR learns a ranking model by learning listwise preference instead of pairwise preference. List-LR takes an instance x and the corresponding lists of preference over the predefined set of social emotion labels as a learning unit and minimizes listwise loss defined on instances instead of only pairwise loss. Results have shown that the proposed List-LR is effective and promising.

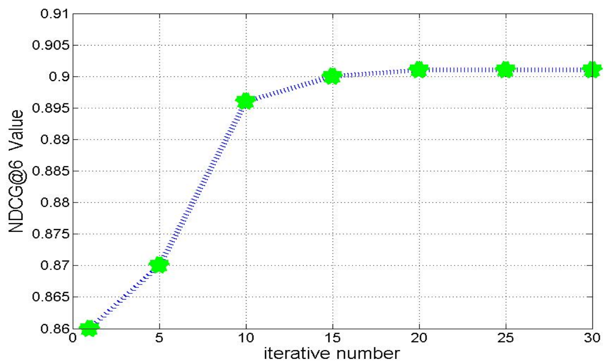

Figure 6. Relation of the iterative number and *NDCG@*6

TABLE I. Example distribution over emotion label pairs (Labels: $l_{(1)}$, $l_{(2)}$, …, $l_{(6)}$ denote happy, sadness, touched, angry, funny and boredom, respectively. The value is the ratio between the number of examples over each emotion label pair)

_	l(2)	l(3)	l(4)	l(5)	l(6)
l(1)	1 : 1.19	3.86 : 1	1 : 2.63	1 : 1.05	1.74 : 1
l(2)	_	4.17 : 1	1 : 2.73	1.05 : 1	1.97 : 1
l(3)	_	_	1 : 4.18	1 : 3.05	1 : 2.29
l(4)	_	_	_	2.59 : 1	3.62 : 1
l(5)	_	_	_	_	2.49 : 1

In the future, we plan to evaluate the proposed approach on a large scale corpus, and evaluate the proposed approach using different type of features. We will also seek for more reasonable metric to model the listwise loss function.

Acknowledgment

This work was supported in part by the NSFC (Grant No. 60825204, 60935002 and 60903147).

References

[1] K. Lin, C. Yang, and H. Chen, “What Emotion do News Articles Trigger in Their Readers?”, SIGIR, 2007, pp. 733-734.

[2] S. Bao,et. al, “Joint emotion-topic modeling for social affective text mining”, IEEE ICDM, 2009, pp. 699-704.

[3] K. Lin, and H. Chen, “Ranking reader emotions using pairwise loss minimization and emotional distribution regression”, EMNLP, 2008, pp. 136-144.

[4] K. Brinker, J. Fürnkranz, and E. Hullermeier, “Label Ranking by Learning Pairwise Preferences”, AI, 2008, pp. 1897-1916.

[5] J. Fürnkranz and E. Hullermeier, “Pairwise preference learning and ranking”, ECML, 2003, pp. 145-156.

[6] J. Fürnkranz, “Round robin classification”, JMLR, 2002, pp. 2:721-747.

[7] W. Cheng, E. H¨ullermeier, “A New Instance-Based Label Ranking Approach Using the Mallows Model”, ISNN(1), 2009, pp. 707-716.

[8] C. L. Mallow, “Non-null ranking models”, Biometrika, 44, 1957, pp. 114-130.

[9] O. Dekel, C. D. Manning, Y. Singer, “Log-linear models for label ranking”, NIPS 16, 2003.

[10] S. Vembu and T. Gartner, “Label Ranking Algorithms: A Survey”, Springer-Verlag, 2010.

[11] R. Plackett, “The analysis of permutations”, Applied Stat., 24, 1975, pp. 193-202.

[12] W. Cheng, K. Dembczynski, E. H¨ullermeier, “Label Ranking Methods based on the Plackett-Luce Model”, ICML, 2010, pp. 215-222.

[13] Z. Cao, T. Qin, T. Liu, M. Tsai, H. Li, “Learning to rank: From pairwise approach to listwise approach”, ICML, 2007, pp. 129-136.