IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 27, 2018, accepted January 7, 2019, date of publication January 24, 2019, date of current version February 12, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2894701*

# Extracting Emotion Causes Using Learning to Rank Methods From an Information Retrieval Perspective

**BO XU[1], HONGFEI LIN[1], YUAN LIN[2], YUFENG DIAO[1], LIANG YANG[1], AND KAN XU[1]**
[1]Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116023, China
[2]WISE Lab, School of Public Administration and Law, Dalian University of Technology, Dalian 116023, China

Corresponding author: Hongfei Lin (e-mail: hflin@dlut.edu.cn)

This work was supported in part by the Natural Science Foundation of China under Grant 61632011, Grant 61572102, Grant 61772103, Grant 61602078, Grant 61702080, and Grant 61572098, in part by the China Postdoctoral Science Foundation under Grant 2018M641691, in part by the Fundamental Research Funds for the Central Universities under Grant DUT18ZD102, and in part by the National Key Research Development Program of China under Grant 2016YFB1001103.

**ABSTRACT** Emotion cause extraction is a challenging task for the fine-grained emotion analysis. Even though a few studies have addressed the task using clause-level classification methods, most of them have partly ignored emotion-level context information. To comprehensively leverage the information, we propose a novel method based on learning to rank to identify emotion causes from an information retrieval perspective. Our method seeks to rank candidate clauses with respect to certain provoked emotions in analogy with query-level document ranking in information retrieval. To learn effective clause ranking models, we represent candidate clauses as feature vectors involving both emotion-independent features and emotion-dependent features. Emotion-independent features are extracted to capture the possibility that a clause is expected to provoke an emotion, and emotion-dependent features are extracted to capture the relevance between candidate cause clauses and their corresponding emotions. We investigate three approaches to learning to rank for emotion cause extraction in our method. We evaluate the performance of our method on an existing dataset for emotion cause extraction. The experimental results show that our method is effective in emotion cause extraction, significantly outperforming the state-of-the-art baseline methods in terms of the precision, recall, and F-measure.

**INDEX TERMS** Emotion analysis, emotion cause extraction, natural language processing, sentiment analysis, learning to rank.

## I. INTRODUCTION

Text-based emotion analysis has attracted plenty of attention in Natural Language Processing (NLP) for the purpose of opinion mining, personalized recommendation and content filtering. Most early research has focused on emotion classification by detecting different types of emotions [1]–[5]. On top of emotion classification, integral elements of emotions have also been addressed and extracted in recent studies to capture how emotions are provoked, expressed and perceived. Emotion cause has been considered as one of the most crucial emotion elements for in-depth emotion analysis. Emotion cause extraction aims to identify the reason or stimuli of an observed emotion. It is a challenging task due to the inherent ambiguity and subtlety of emotion expressions.

An emotion event in the context is always provoked by certain underlying causes and characterized by certain emotion words. The context involving the causes and the emotion words is then treated as a passage for emotion cause extraction. Emotion passages are segmented into clauses, and each clause is a candidate unit for emotion causes. The goal of emotion cause extraction is thus transformed into extracting the clauses containing the causes of provoked emotions.

Recent studies have focused on solving the problem using various methods [6]–[15]. Most of the existing studies on emotion cause extraction have adopted rule-based or classification-based methods to identify the emotion causes. However, they have partly overlooked emotion-level information conveying in the candidate clauses. Since multiple clauses for a provoked emotion are considered as the candidates for cause extraction, extraction methods should capture the relevance between each candidate clause and the observed emotions, and meanwhile measure the possibility

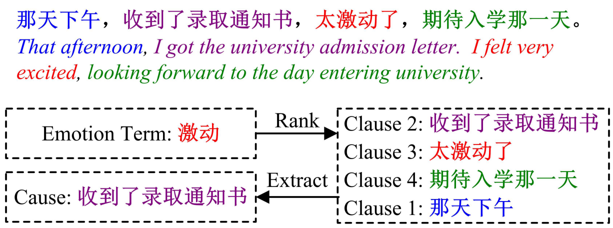



**FIGURE 1. Incorporating ranking mechanisms in extracting emotion causes.**

for each candidate clause to provoke an emotion. To this end, we transform the problem into refining the candidate clauses at emotion level to detect emotion causes, and tackle the problem using learning to rank methods.

In information retrieval (IR), learning to rank methods have been widely used to solve ranking problems. These methods incorporate ranking constraints into the loss functions of supervised machine learning methods for constructing effective ranking models. Ranking models produce query-level ranking lists of candidate documents for given queries. To extract causes for observed emotions, learning to rank methods are used to perform emotion cause-oriented clause-level ranking by considering multiple characteristics of candidate clauses. To help better understand the motivation of our work, we illustrate a toy example of emotion cause extraction based on ranking mechanisms in Fig. 1. In the example, there are four clauses in one passage with respect to a provoked emotion characterized by the emotion word 'excited'. The emotion passage contains both the emotion clause and the cause clause. The goal of the task is to extract the clause containing the emotion cause. We incorporate ranking mechanisms to rank the candidate clauses based on their possibility to provoke the emotion, aiming to sort the cause clause at the top of the ranking list for emotion cause extraction.

In this paper, we transform emotion cause extraction as a ranking problem, and deal with the problem using learning to rank methods from an information retrieval (IR) perspective. Our method treats provoked emotions as queries and candidate clauses in emotion passages as documents for emotion cause-oriented clause-level ranking. To train effective ranking models, we define and extract both emotion-independent and emotion-dependent ranking features of clauses in analogy with query-dependent and query-independent ranking features used in IR tasks. Emotion-independent features estimate the possibility of each clause provoking an emotion, and emotion-dependent features measure the relevance of each clause to its corresponding emotion. In model training, we investigate the pointwise, pairwise and listwise ranking constraints for building effective ranking models in emotion cause extraction, respectively. We summarize the contributions of our paper as follows.

1) We address the task of emotion cause extraction from an information retrieval perspective, and solve the problem using learning to rank methods. Our method performs emotion cause-oriented clause-level ranking by considering three types of ranking constraints for accurately extracting emotion causes.

2) We define and extract abundant features of clauses for comprehensive emotion cause-oriented representations, including the emotion-independent ranking features and emotion-dependent ranking features. We compare the effectiveness of different features for emotion cause extraction.

3) We conduct extensive experiments to examine the performance of our method in extracting emotion causes compared with the state-of-the-art baseline methods. Experimental results show that our method is effective in identifying the emotion causes for fine-grained emotion analysis.

The rest of the paper is organized as follows. Section 2 introduces the related work on emotion cause extraction and learning to rank; Section 3 details our model for learning to rank based emotion cause extraction; Section 4 provides our experiments and results analysis; Section 5 concludes the paper and provides our future work.

## II. RELATED WORK

We present related work on emotion cause detection and learning to rank in this section, and discuss how our study differs from the previous attempts.

Text-based emotion analysis has been studied for years in the area of natural language processing (NLP) [16]–[18]. Early researches have focused on identifying different categories of emotions [1]–[5]. For example, Li *et al.* [5] recasted sentence-level emotion classification as a factor graph inferring problem by considering the label dependence and the context dependence. Deyu *et al.* [2] incorporated the relations of different emotions into their learning algorithms in order to improve the accuracy of emotion classification. Xu *et al.* [4] proposed a novel approach using a coarse-to-fine analysis strategy for sentence-level emotion classification by considering the similarity and adjacency of sentences. However, classification-based emotion analysis may partly overlook the subtlety of human emotion, such as the provocation, evolution and results of emotions.

To capture fine-grained information for emotions, recent studies have sought to extract key elements for provoked emotions, such as the experiencer, cause and results of emotions. Emotion cause extraction is one of the most challenging extraction tasks, aiming to identify the cause or stimuli of an emotion. Emotion cause detection has been treated as an integral part of in-depth emotion analysis by most psychological theories [19]–[22].

Various approaches to emotion cause detection have been proposed in recent years, which can be categorized into two classes: the rule-based approach and the classification-based approach. The rule-based approach generalized linguistic rules to detect emotion-provoking cues in natural language. As the first attempt to automatic detection of emotion cause, Lee *et al.* [10] developed a rule-based system for emotion cause detection. They constructed an emotion dataset with cause event annotations, and generalized a series of linguistics rules based on the dataset. The rule-based approach was extended to microblogs in following studies [6], [23], [24]. The classification-based approach adopted supervised machine learning methods to detect emotion causes. For example, Chen *et al*. [11] introduced linguistic rule-based features to learn a max-entropy-based classifier to detect emotion causes. Support vector machines (SVMs) and conditional random fields (CRFs) were adopted to classify cause or non-cause text with extended rule-based features in existing studies [7], [12]. Gui *et al*. [8] employed an SVM based method for emotion cause detection, which extracted convolution kernels from syntactic trees to capture the cause information. In their subsequent study, they proposed a memory network-based question-answering approach to further enhance the performance of emotion cause extraction [9]. Russo *et al*. [13] incorporated common sense knowledge into linguistic patterns for emotion cause extraction in Italian newspaper articles. Cheng *et al*. [14] focused on emotion cause detection on Chinese microblog, and proposed to use multiple-user structures for extracting the causes. Mulkar-Mehta *et al*. [15] used granularity relations to discover and infer the presence of causal relations in text. However, these studies on emotion cause detection have partly overlooked the relationship between the provoked emotion and its context by directly classifying the candidate clauses. In fact, clauses within certain context of an emotion can be treated as candidates of causes, and refined based on their relevance to the corresponding emotion. To fully capture emotion-level information, we propose to rank candidate clauses for emotion cause extraction based on learning to rank from an information retrieval perspective. Learning to rank has been proved to be effective in constructing ranking models based on supervised machine learning methods in IR tasks [25]–[30]. The constructed ranking models have been widely applied in various natural language processing tasks. For example, Zhang *et al.* [31] proposed to choose decision rules in statistical machine translation from a listwise ranking perspective. Santos *et al*. [32] proposed a pairwise ranking method for relation classification using convolutional neural networks. Shen and Li [33] adopted learning to rank methods for multi-document summarization. These studies have shown that learning to rank methods are effective in ranking candidate items for a given query. Based on the definition of loss functions, learning to rank methods can be categorized into three approaches: the pointwise approach, pairwise and listwise approach [25]. Different approaches to learning to rank consider different ranking constraints for optimizing ranking models. The pointwise approach seeks to predict the relevance of each single document for a given query, the pairwise approach predicts the preference order of two documents, and the listwise approach considers the entire candidate list to optimize the ranking models.

Since the relationship of clauses within the context of a provoked emotion has not been fully considered in previous studies on emotion cause extraction, we transform emotion cause extraction into a supervised clause-level ranking problem, and solve the problem using different approaches to

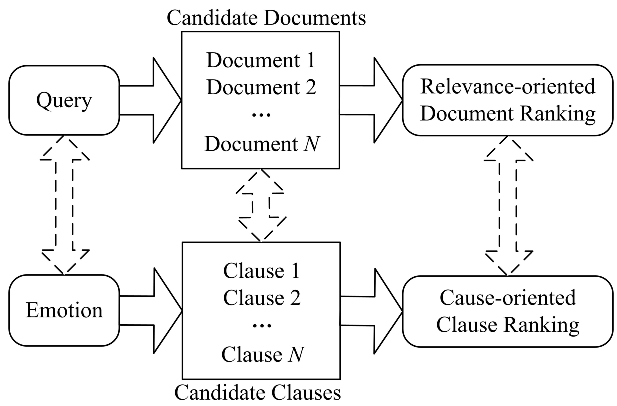



**FIGURE 2. Problem transformation from relevance-orient document ranking to cause-oriented clause ranking.**

learning to rank. In the transformation, we treat emotions as queries and clauses within the context of emotions as documents in analogy with query-oriented document ranking in IR tasks for emotion cause-oriented clause ranking.

## III. LEARNING TO RANK FOR EMOTION CAUSE EXTRACTION

### A. PROBLEM FORMALIZATION

In this section, we formalize the proposed framework based on learning to rank for emotion cause extraction. We follow the formal definition of emotion cause extraction by Gui *et al.* [8]. Text contents involving emotions are an indispensable resources for opinion mining and personalized recommendation. Generally, an emotion event in the context is always provoked by certain underlying causes and characterized by certain emotion words. A passage within the context of emotion words and emotion causes is used for emotion cause extraction. An emotion passage comprises several clauses, and each clause is a candidate unit for fine-grained emotion analysis. The emotion cause-oriented clauses are among the candidate clauses in an emotion passage. Emotion cause extraction seeks to extract the clauses containing the causes for a provoked emotion. For each emotion, the goal of the task is to identify which clauses contain the emotion cause. In fact, this is an emotion cause oriented clause-level extraction task. Previous studies addressing the task mainly treated it as a classification task, and solved based on various classifiers, which may partly overlook the emotion information. In our study, we transform the task as a ranking problem from an information retrieval (IR) perspective in analogy with relevance oriented document ranking used in IR tasks. We then tackle the problem using learning to rank methods for emotion cause oriented clause-level ranking. We illustrate the model transformation in Fig. 2 for easily understanding the motivation of our work.

Formally, we are given a set of $N$ emotions $E = \{e_1, e_2, \ldots, e_N\}$ at the training time. To simplify notations, we drop the emotion index, and refer to a general emotion $e$. Each emotion $e$ is associated with a set of $M$ candidate clauses

$C = \{c_1, c_2, \ldots, c_M\}$. The clauses are labeled with emotion cause-oriented labels, denoted as $L = \{l_1, l_2, \ldots l_M\}$. Each label is an integer indicating whether the clause can provoke the corresponding emotion. Each clause $c_i$ is represented as a feature vector $F = \{f_{i1}, f_{i2} \ldots f_{ik}\}$, reflecting the importance and preference of the clause for the given emotion. The goal is then transformed as learning a ranking-based extraction model, which can produce a ranking list of clauses with maximal agreement with the ideal list based on $L$ for emotion cause extraction. In our framework, we mainly address two crucial issues in learning to rank based emotion cause extraction. One is how to extract useful ranking features and the other is how to construct effective extraction models. We then detail these two issues within our framework in the following subsections.

## B. EMOTION CAUSE-ORIENTED RANKING FEATURES ON CLAUSES

Ranking features in emotion cause extraction are used to comprehensively represent the emotion-related clauses. The used features should not only indicate the importance of a certain clause in emotion context, but also reflect the relationship between clauses and the given emotion. To this end, we extract two types of clause-level ranking features for candidate clauses: emotion-independent features and emotion-dependent features.

### 1) EMOTION-INDEPENDENT FEATURES

Emotion-independent features are extracted to capture the importance of each clause in the context of a given emotion, reflecting the possibility that the clause provokes an emotion. We extract emotion-independent features in analogy with query-independent document features used in information retrieval tasks. To extract useful clause features for emotion cause extraction, we take consideration of linguistic and semantic information in clauses from three perspectives: the length of clauses, the part-of-speech (POS) tagging and the feature terms.

We first treat the length of each clause as one type of feature. Intuitively, long clauses can be more likely to provoke an emotion, and short clause may contain less semantic information for provoking an emotion. We count the number of characters and the number of words as the clause length in our study, respectively, because Chinese characters and words convey semantic information at different levels. These two features may contribute differently to the construction of ranking models.

To capture the linguistic information in clauses, we extract clause features based on POS tagging. POS tagging has been widely used in NLP tasks, particularly in the linguistic-based emotion classification task. In the feature extraction, we assign POS tags for words in clauses, and count the number and ratio of words in different tags as different features. We use the proportion of different part-of-speech in a clause to capture linguistic characteristics of clauses. Specifically, we consider the number and ratio of nouns, verbs, adjectives and adverbs in each clause as different ranking features for cause-oriented clause feature representations.

**TABLE 1. Definitions of categories of feature terms.**

| ID | Category | Feature terms |
|---|---|---|
| 1 | Causal conj. | 因为/because, 因/due to, 由于/because of, etc. |
| 2 | Causal verbs | 让/make, 令/cause, 使/let |
| 3 | Sensory verbs | 想到/think of, 听到/hear, 看到/see, 感到/sense, etc. |
| 4 | Emotion terms | 激动/excited, 快乐/happy, 愤怒/angry, 惊讶/surprised, 恐惧/fearful, etc. |
| 5 | Negation | 不/not, 没有/no |
| 6 | Family pron. | 丈夫/husband, 妻子/wife, 儿子/son, 女儿/daughter, 父亲/father, 母亲/mother, etc. |

In addition, previous research [10] has shown that different linguistic cues are highly collocated with emotion events. We therefore believe clauses containing the cue word may have more possibility to provoke the given emotion. Cue words in our study mainly refer to the words with respect to emotion cause. We therefore choose six groups of cue words as features terms, and then extract clause features based on the number and ratio of feature terms in each clause.

We provide the selected feature terms used in our study in Table 1. These feature terms include causal conjunctions, sensory verbs, emotion terms, causal verbs, negations and family-related pronouns. Causal conjunctions are an important indicator for reflecting whether a clause contains a cause or not. Similarly, causal verbs can introduce the cause in a clause, which may be helpful in emotion cause identification. Sensory verbs describe human reactions to external stimuli, which may explain the reasons of emotions in turn. Emotion terms indicate whether a clause is related to an emotion, because emotion clauses are more like to correlate with other emotions. Negations have been widely used in emotion analysis, which may record the emotion change in clauses. We observe that emotions are always provoked related to certain persons, such as family members, we treat family-related pronouns as feature terms in emotion cause extractions. Based on the above definitions, we extract clause features based the number and ratio of each group of features terms in clauses as different features.

We define and extract the above-mentioned emotion-independent features to represent clauses based on their own importance. To further consider the relations between clauses and their corresponding emotions, we next extract emotion-dependent features by fully considering the given emotions.

### 2) EMOTION-DEPENDENT FEATURES

Emotion-dependent features are used to depict the relationship between a certain emotion and its candidate clauses, which is analogous to the query-dependent document features used in IR tasks. We define and extract this type of ranking features from three perspectives: the relative position, the word embedding-based similarity and the topic model-based similarity.

We first extract a type of emotion-dependent features based on the relative position between a candidate clause and the given emotion clause. Existing research has shown that clauses next to a provoked emotion are more likely to cause the emotion [8]. Therefore, we take the relative position of candidate clauses to a given emotion as an important ranking feature. For example, we count the relative position of the emotion clause as 0, and the previous clause and the next clause to the emotion clause as 1 in terms of the distance between them.

Classic retrieval models used in IR are mostly based on scoring functions, which measure the similarity between a given query and its candidate documents. These models have been successfully used as document features in learning to rank. To measure the similarity of an emotion and its candidate clauses for emotion cause extraction, we adopt two kinds of approaches for clause feature extraction based on word embedding and topic models.

Word embedding can represent words in low-dimensional continuous vectors, capturing abundant semantic and linguistic information of words. It has been proved to be effective in many NLP tasks. In this work, we use word embedding to represent words in clauses as distributed vectors for measuring the similarity between an emotion and its candidate clauses. Specifically, we first compute the similarity between the emotion word and each word in a candidate clause. We then record the average, maximum and minimum similarities of words in the clause as different ranking features, respectively. These features can capture the similarity between candidate clauses and the emotion. In addition, we use word embedding to represent candidate clauses as vectors for directly measuring the similarity between candidate clauses and the emotion clause based on the averaged word embedding in each clause.

Topic models are a type of statistical model for discovering latent topics occurring in document collections, which have been widely used for detecting hidden semantic structures in context. To use topic models for clause representations in emotion cause extraction, we construct topic models based on three levels, including the clause level, the emotion level and the document level. The clause-level model treats each clause as a learning unit, the emotion-level model treats each passage within single emotion context as a learning unit, and the document-level model treats each document containing multiple emotions as a learning unit. Each level takes different granularities of emotion context into consideration to learn distinguished topic models from different perspectives.

We use two approaches to topic models in our study, namely, the Latent Dirichlet Allocation model (LDA) [34] and the Latent Semantic Indexing model (LSI) [35]. LDA posits that each document is generated as a mixture of topics, and each topic is determined based on a mixture of words. LSI aims to map vector space representations of documents to low-dimensional latent semantic space for effective topic modeling. We finally use the learned model to represent candidate clauses and the emotion clause to measure the similarities between them as different clause features. We adopt cosine similarity to measure the distances between topic representations.

We represent candidate clauses with respect to given emotions based on the emotion-independent and emotion-dependent features. We then use the clause feature representations as inputs for learning effective emotion cause extraction models. We extract 34 features in total. For all the defined ranking features, we conduct emotion-level feature normalization to obtain comparable feature values in analogy with query-level normalization used in IR tasks. Query-level normalization has been widely used in learning to rank for constructing robust ranking models. Specifically, we normalize each feature as follows.

$$f_{new} = \frac{f_{old} - min_e}{max_e - min_e} \tag{1}$$

where $f_{old}$ is the raw feature value, $f_{new}$ is the normalized feature value. $min_e$ and $max_e$ represent the minimum and maximum of the feature values with respect to the corresponding emotion, respectively.

### C. RANKING MODELS FOR EMOTION CAUSE EXTRACTION

We introduce learning to rank models for emotion cause extraction in this section. We investigate three approaches to learning to rank in this task, including the pointwise, pairwise and listwise approach.

The pointwise approach directly employs existing machine learning methods to solve ranking task. The pointwise extraction model seeks to predict the exact relevance of each candidate clauses to a given emotion. The ranking loss is accumulated based on the difference between the predicted score and the ground truth labels, and iteratively reduced by model optimization. The loss function of the pointwise method Regression [36] can be formalized as follows.

$$loss(f(x_i), y_i) = \sum_i (f(x_i) - y_i)^2 \tag{2}$$

where $f$ is the ranking model, $f(x_i)$ is the predicted score of the clause $x_i$ by the model, and $y_i$ is the ground truth label of the clause. From the equation, we can see that the total loss for pointwise method is the loss summation over all the clauses.

The pairwise approach defines the ranking loss based on preference orders of each two clauses, and optimizes its model by considering the number of wrongly classified clause pairs. Take RankBoost [37] as an example. RankBoost combines preferences based on the boosting approach to machine learning. It uses the object pairs with preferences as instances in its training procedure, and operates in rounds by combining weak learners, each of which is weakly correlated with the target ranking model. The final ranking model of RankBoost is an ensemble of all the weak learners.

$$loss(f(x_i), f(x_j), y_{i,j}) = \sum_{i,j} e^{-y_{i,j}\cdot(f(x_i)-f(x_j)))} \tag{3}$$

Equation (9) is the loss function of RankBoost, where $f$ is the ranking model, $f(x_i)$ and $f(x_j)$ is the predicted scores of the clause $x_i$ and the clause $x_j$, and $y_{i,j}$ is the preference between these two clauses based on the ground truth labels.

From the equation, we can see that the total loss for pairwise methods is the loss summation over all the clause pairs.

The listwise approach learns the ranking model by directly fitting the predicted ranking list and the ideal ranking list of clauses. Listwise approach makes the most of ranking information for model optimization. LambdaMART [38], as a listwise method, is as an ensemble of tree-based rankers, implements LambdaRank [39] using multiple additive regression tree [40]. LambdaMART uses MART with specific appropriate gradients and the Newton step to find the minimum of the loss function, and then computes outputted values of leaf nodes in each regression tree. LambdaMART introduces parameter $\lambda$ as a replacement of the loss function gradient. The $\lambda$ for a given clause in the ranking list gets contributions from all other clauses for the same emotion as follows.

$$\lambda_i = \sum_{j:(i,j)\in I} \lambda_{i,j} - \sum_{j:(j,i)\in I} \lambda_{i,j} \tag{4}$$

LambdaMART modifies the gradient of the loss with the variation of ranking performance through swapping the rank positions of two clauses, where $\lambda_{i,j}$ is the ranking loss by swapping the positions of the clause $i$ and the clause $j$. LambdaMART uses $\lambda$ as the gradient of loss function and uses boosted regression trees as its model to decrease ranking loss in iterations.

We illustrate our learning to rank framework for emotion cause extraction in Fig. 3. In the framework, we first represent candidate clauses for provoked emotions as feature vectors. The vector representations of clauses involve different types of emotion-independent features and emotion-dependent features according to our definitions. We then treat these representations as inputs of learning to rank for constructing emotion cause-oriented clause ranking models. We investigate three approaches to learning to rank for the emotion cause extraction task. It has been proved that more than 97% emotions corresponds to a single cause [8]. We therefore treat the top-one ranked clause for each emotion as the emotion cause in our evaluation, because learning to rank can well deal with top-one ranking optimization. We remain the identification of multi-cause emotions in our future work.

## IV. EXPERIMENTS AND ANALYSIS

In this section, we examine evaluate the proposed method for emotion cause extraction. We first introduce the experimental setup, and then report the experimental results for analysis and discussion.

### A. EXPERIMENTAL SETUP

We use the publicly available dataset by Gui *et al.* [8] in our experiments. The dataset is designed for emotion cause extraction, containing 2,105 emotion-oriented passages from SINA city news. Each passage contains only one provoked emotion characterized by one keyword and at least one emotion causes. It has been ensured that the emotion and the causes are relevant. The passages are segmented into clauses manually for the identification of emotion causes. The goal of this task is to identify which clauses in each passage provoke the emotion. Details of the dataset are shown in Table 2. The table indicates that 97.2% of the emotions has only one emotion cause, and 2.6% and 0.2% of the emotions has two and three emotion causes, respectively. Since most emotions have one cause, we treat top-one ranked clause for each emotion as the target emotion cause. We then evaluate the performance of emotion cause extraction based on the extracted top-ranked causes. We pretrain word embeddings using the word2vec tool and the Sina news data with default configurations. We use the NLTK tool for part-of-speech tagging in our experiments.

**TABLE 2.** Statistics of dataset.

| Item | Number |
|---|---|
| Passages | 2,105 |
| Clauses | 11,799 |
| Emotion Causes | 2,167 |
| Passage with 1 Cause | 2,046 |
| Passage with 2 Causes | 56 |
| Passage with 3 Causes | 3 |

We adopt the commonly accepted evaluation metrics proposed by Lee *et al.* [10] to examine the performance of cause extraction. These evaluation metrics have been widely used in evaluation the performance of emotion cause extraction [8], [9], [23], [24]. As mentioned above, we treat the top-ranked clause in each emotion-oriented passage as the cause clause for the computation of the precision, recall and F-measure. In these metrics, if an identified cause covers the annotated cause, the extraction is considered correct. The precision (P), recall (R) and F-measure (F) for emotion cause extraction are defined as follows.

$$Precision = \frac{\sum_{correct\ cause} 1}{\sum_{identified\ cause} 1} \tag{5}$$

$$Recall = \frac{\sum_{correct\ cause} 1}{\sum_{annotated\ cause} 1} \tag{6}$$

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{7}$$

To obtain the average performance of learning to rank methods, we perform five-fold cross validations to train the clause-ranking models. Specifically, we divide the emotion passages into a training set, a validation set and a test set in a ratio of 3:1:1 by emotion IDs, following the standard partition strategy used in LETOR [26], the widely used learning to rank datasets. We use the training set to train ranking models, the validation set to select the parameters for different ranking models, and the test set to predict on new emotions. We report the experimental results based on the average performance on all folds for fair comparisons.

### B. BASELINE MODELS

We compare our models with three categories of baseline models, including the rule-based models, the classification-based models and the neural network-based models.

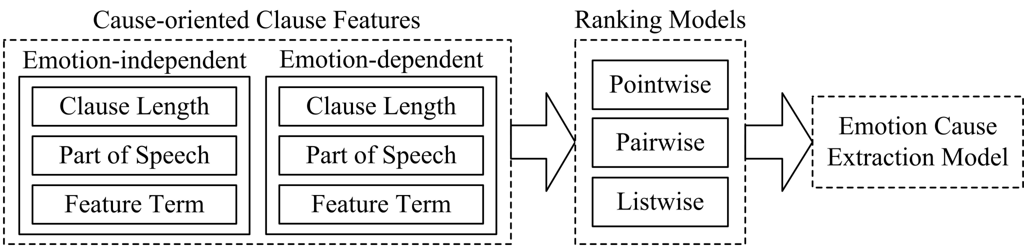



**FIGURE 3.** learning to rank framework for emotion cause extraction.

For the rule-based models, we compare with the rule-based system (RB) proposed by Lee *et al.* [10] and the common-sense-based method (CB) by Russo *et al.* [13]. We use the Chinese emotion cognition lexicon [41] as the common-sense knowledge base. Furthermore, we extract rule-based features from the common-sense knowledge based for SVM-based classification, denoted as RB+CB+ML.

For the classification-based models, we first use an SVM classifier with unigram, bigram and trigrams features for cause extraction [8], [23]. Moreover, we employ word representations by Mikolov *et al.* [42] as features for classification. We further compare with the state-of-the-art method using the multi-kernel SVM for emotion cause identification [8].

For the neural network-based models, we first use the convolutional neural network (CNN) for clause-level classification [43]. We then compare with the state-of-the-art model using convolutional multiple-slot deep memory network (ConvMS-Memnet) [9] based on the best performance reported in their work.

For the proposed ranking-based models, we evaluate the performance using the pointwise model Regression [36] and SVM (denoted as Regression and SVM-LTR), the pairwise model RankBoost [37] and the listwise model LambdaMART [38], respectively, with the defined features. Furthermore, we compare the ranking models using different ranking features to examine the effectiveness of different features. We then examine the influence of stopword removal and emotion-level feature normalization in learning ranking models.

### C. RESULTS AND DISCUSSIONS

In this section, we first compare our models with the baseline models, and provide analysis on the comparison. We then examine the effectiveness of different types of defined features, and report the experimental results using different features. We particularly investigate feature terms and topic models for constructing emotion cause extraction models. We also examine the influence of stopword removal and emotion-level normalization in the proposed method. We finally provide further discussions on our models.

#### 1) COMPARISON WITH EXISTING METHODS

We first compare our models with the existing baseline models, and report the comparisons in Table 3. From the table, we observe that for the rule-based models, RB yields high precision with relatively low recall, and CB obtains high recall with relatively low precision. This exhibits the advantages of linguistic rules in emotion cause extraction, and meanwhile indicates the limitation of the rule-based models in improving the overall performance. The performance in terms of different evaluation metrics can be balanced when combining RB and CB, and further enhanced using machine learning based classifiers with linguistic rule based and common sense based features.

**TABLE 3.** Comparison with existing models. An asterisk indicates significant improvements over the ConvMS-Memnet model.

| Method | Precision | Recall | F-measure |
|---|---|---|---|
| RB | 0.6747 | 0.4287 | 0.5243 |
| CB | 0.2672 | **0.7130** | 0.3887 |
| RB+CB | 0.5435 | 0.5307 | 0.5370 |
| RB+CB+ML | 0.5921 | 0.5307 | 0.5597 |
| SVM | 0.4200 | 0.4375 | 0.4285 |
| Word2vec | 0.4301 | 0.4233 | 0.4136 |
| CNN | 0.6215 | 0.5944 | 0.6076 |
| Multi-kernel | 0.6588 | 0.6927 | 0.6752 |
| ConvMS-Memnet | **0.7076** | 0.6838 | **0.6955** |
| Regression | 0.7331* | 0.7176* | 0.7219* |
| SVM-LTR | 0.7340* | 0.7130* | 0.7233* |
| RankBoost | 0.7520* | 0.7305* | 0.7411* |
| LambdaMART | **0.7720*** | **0.7499*** | **0.7608*** |

For the classification-based models, the SVM based classifier with unigram, bigram and trigram features achieves comparable performance with the classifier with word representation based features. The classifier with linguistic rule based and common sense based features outperforms the former two classifiers. The experimental results show that linguistic rules and common senses are more useful than other statistical textual features for emotion cause extraction. The multi-kernel based SVM classifier outperforms other SVM based models in terms of all the evaluation measures. This is because the multi-kernel SVM classifier uses lexical information at the terminal nodes of syntactic trees to improve the kernel function with a synonym based improvement.

For the neural network-based models, the CNN based model outperforms most classification based models except the multi-kernel model, and the ConvMS-Memnet model achieves the best performance in terms of F-measure and the precision among all the baseline models. the ConvMS-Memnet model considers the context of emotions using multiple memory slots, and achieves the best performance

**TABLE 4. Evaluation on feature importance. The column 'Ratio' shows the proportion of performance decline compared to the model with all the defined features by F-measure.**

| Feature | Precision | Recall | F-measure | Ratio |
|---|---|---|---|---|
| All-FT | 0.7525 | 0.7310 | 0.7416 | -2.52% |
| All-WE | 0.7629 | 0.7411 | 0.7519 | -1.17% |
| All-TM | 0.7363 | 0.7153 | 0.7257 | -4.61% |
| All-POS | 0.7591 | 0.7374 | 0.7481 | -1.67% |
| All-Position | 0.6456 | 0.6271 | 0.6362 | -16.38% |
| All-Length | 0.7577 | 0.7360 | 0.7467 | -1.85% |

**TABLE 5. Cause position of each emotion.**

| Position | Number | Percentage |
|---|---|---|
| Previous 3 clauses | 37 | 1.71% |
| Previous 2 clauses | 167 | 7.71% |
| Previous 1 clauses | 1,180 | 54.45% |
| In the same clauses | 511 | 23.58% |
| Next 1 clauses | 162 | 7.47% |
| Next 2 clauses | 48 | 2.22% |
| Next 3 clauses | 11 | 0.51% |
| other | 42 | 1.94% |

compared to existing methods. The performance of baseline models implies that the context and linguistic information of emotions is highly useful for emotion cause extraction. We consider these two aspects of information in our models by constructing cause-oriented clause-level ranking models with linguistics-based ranking features.

The proposed ranking models further improve the performance of the baseline models as shown in the table. The improvement is very significant with p-value less than 0.01 in t-test. For the three ranking models, the pairwise model RankBoost achieves better performance than the pointwise model SVM, and the listwise model LambdaMART outperforms the other two models. This trend is consistent with ranking performance in other IR tasks. One possible explanation for this finding is that the listwise ranking model captures more ranking information to optimize the intermediate model than the pairwise and pointwise models. Ranking information in emotion cause extraction is useful to capture the context information for given emotions, and the clause features comprehensively model the candidate clauses for building effective ranking models. The experimental results show that learning to rank models are effective in emotion cause extraction compared to other baseline extraction models.

To gain more insights into our model, we conduct further experiments to evaluate the proposed models. We examine the effectiveness of different kinds of ranking features, particularly focusing on the features based on features terms and topic models. We also compare the performance of ranking models with different preprocessing steps to examine the influence of stopword removal and emotion-level feature normalization in constructing ranking models.

### 2) EVALUATION ON RANKING FEATURES

We first examine the effectiveness of different ranking features by removing one category of features from the entire feature set, and using the remaining features to construct the ranking models based on LambdaMART. We report the experimental results in Table 4, where 'All-FT', 'All-WE', 'All-TM', 'All-POS', 'All-Position' and 'All-Length' refer to the feature set without feature term, word embedding, topic model, POS tagging, relative position, clause length based features, respectively. 'All' refers to the entire feature set.

The table shows that when removing the relative position based features, the performance decreases sharply, which implies that the relative position-based ranking features are important for identifying the cause clause. We further explore the reasons why relative position based features can largely affect the extraction performance. We provide the distribution of emotion cause positions in Table 5. The table shows that more than 85% emotion causes adjoin the emotion clause. We therefore believe that relative position plays a very important role in emotion cause extraction, and the relative position based ranking features are rational and necessary.

The topic model-based features and feature term-based features also play an important role in emotion clause ranking. Topic models capture the latent topics at different levels, which contributes much to modeling the relationship between candidate clauses and emotion clause. Feature terms are important causal indicators in clauses, and thus clauses containing different features terms are more likely to provoke emotions. Since we define different groups of feature terms and adopt different topic models as learning features, we will further examine the effectiveness of different features in the following subsections.

In addition, the word embedding-based, POS-based and length-based features yield comparable performance to each other, contributing to more effective ranking models for emotion cause extraction. The ranking model using all the defined features achieves the best performance, which demonstrates the effectiveness of different features.

### 3) EVALUATION ON FEATURE TERMS

We define and extract six groups of feature terms as ranking features based on linguistic cues. In this section, we combine each group of feature terms with other types of features for model training, and report the performance in Table 6. From the table, we observe that each group of feature terms is useful in improving the performance of cause extraction. Among different groups of feature terms, causal conjunctions are the most effective for emotion cause extraction. Negation terms and family-related pronouns are the second most effective in the task. This finding implies that different feature terms appearing in clauses can indicate the possibility that a clause is related to emotion causes. Different groups of features terms can jointly contribute to the improvement of emotion cause extraction performance.

### 4) EVALUATION ON TOPIC MODELS

We use two kinds of topic models, LSI and LDA, to generate topic model-based ranking features. To train topic models,

**TABLE 6.** Evaluation on the influence of feature terms. The column 'Ratio' shows the proportion of performance increase compared to the model with All-FT feature set by F-measure.

| Feature term | Precision | Recall | F-measure | Ratio |
|---|---|---|---|---|
| Causal conj. | 0.7672 | 0.7453 | 0.7561 | +1.96% |
| Causal verbs | 0.7568 | 0.7351 | 0.7458 | +0.57% |
| Sensory verbs | 0.7606 | 0.7388 | 0.7495 | +1.07% |
| Emotion terms | 0.7572 | 0.7356 | 0.7463 | +0.63% |
| Negation | 0.7644 | 0.7425 | 0.7533 | +1.58% |
| Family-related pron. | 0.7629 | 0.7411 | 0.7519 | +1.39% |

**TABLE 7.** Evaluation on the influence of topic models. The column 'Ratio' shows the proportion of performance increase compared to the model with All-TM feature set by F-measure.

| Topic model | Precision | Recall | F-measure | Ratio |
|---|---|---|---|---|
| LSI-clause | 0.7587 | 0.7370 | 0.7477 | +3.03% |
| LSI-emotion | 0.7663 | 0.7443 | 0.7551 | +4.05% |
| LSI-document | 0.7682 | 0.7462 | 0.7570 | +4.31% |
| LDA-clause | 0.7601 | 0.7383 | 0.7491 | +3.22% |
| LDA-emotion | 0.7615 | 0.7397 | 0.7505 | +3.42% |
| LDA-document | 0.7667 | 0.7448 | 0.7556 | +4.12% |

we adopt the clause-level, emotion-level and document-level units for obtaining diversified models, respectively. These models are then used to represent clauses for extracting different ranking features. In this section, we compare the effectiveness of different topic models for ranking. We report the experimental results in Table 7, where the performance of each row is obtained using the LambdaMART-based ranking model with one topic model-based feature and other types of features.

The table shows that the LDA-based features achieve comparable performance with the LSI-based features. Emotion-level topic modeling outperforms clause-level topic modeling, and document-level topic modeling is more effective than the other two levels for emotion cause extraction. The reason may be that documents with multiple emotions are more intact than emotion-oriented passages or clauses for extracting the latent topics. As a result, the extracted topics tend to be more useful to measure the similarity of candidate clauses and emotion clauses for effective ranking.

#### 5) INFLUENCE OF STOPWORDS AND EMOTION-LEVEL FEATURE NORMALIZATION

In our experiments, we observe that stopwords removal and emotion-level feature normalization can affect the cause extraction performance. Therefore, we evaluate their influence in this section by training the ranking models using different combinations of them. We report the experimental results in Table 8, where 'stopword' represents removing the stopwords when extracting ranking features except the feature term-based features, and 'Norm.' represents the emotion-level feature normalization.

The table shows that both stopword removal and emotion-level feature normalization affect the performance of cause extraction. Emotion-level feature normalization yields larger improvement than stopword removal. One possible explanation for this finding is that stopwords may add noises when

**TABLE 8.** Influence of stopwords removal and query-level normalization.

| Norm. | Stopword | Precision | Recall | F-measure |
|---|---|---|---|---|
| No | No | 0.7568 | 0.7351 | 0.7458 |
| Yes | No | 0.7648 | 0.7430 | 0.7537 |
| No | Yes | 0.7615 | 0.7397 | 0.7505 |
| Yes | Yes | **0.7720** | **0.7499** | **0.7608** |

measuring the similarity between candidate clauses and emotion clauses. Similar to the query-level feature normalization used in IR, emotion-level feature normalization is helpful in constructing robust and effective ranking models for emotion cause extraction. Therefore, these two factors jointly contribute to the clause ranking performance.

### D. DISCUSSIONS

We have examined the performance of the proposed models in comparison with the baseline models. In this section, we provide further analysis and discussions particularly on the advantages and disadvantages of our methods for further optimization. Unlike the existing methods for emotion cause extraction, our method seeks to rank candidate clauses for certain provoked emotions from an information retrieval perspective. The learned extraction model based on learning to rank integrates emotion cause-oriented clause information comprehensively, contributing much to the ranking models. In model training, learning to rank methods fully capture context information of clauses, and rank the cause clause at the top of the ranking list. Experiments show that clause features based on relative position, feature terms and topic models are the most useful for clause feature representations, capturing linguistic information and latent topic information. In addition, our method still has room for further optimization. In clause ranking, we treat the top-one ranked clause for each emotion as the cause, ignoring the emotions with multiple causes. This may limit the performance of our models. In the future, we will design useful strategies to determine whether an emotion has one cause or multiple causes for the improvement of extraction performance. We will also examine the effectiveness of our method on other cause extraction datasets. Since the manual annotation of emotion cause data is expensive, we will explore effective ways for automatic annotations.

## V. CONCLUSIONS

In this paper, we propose a novel clause ranking method to tackle the problem of emotion cause extraction using learning to rank. We first transform emotion cause extraction as a supervised ranking problem from an information retrieval perspective. We then define and extract a large amount of emotion-independent and emotion-dependent clause features for emotion cause-oriented clause representations. We investigate three approaches to learning to rank for constructing clause-level ranking models. We evaluate our models using a publicly available dataset for emotion cause extraction. Experimental results show that the proposed models is effective in identifying emotion causes. Our models outperform the state-of-the-art models in terms of the precision, recall and F-measure. In the future, we will construct more powerful ranking models by developing effective ranking features for emotion cause extraction. We will attempt to optimize the ranking-based extraction model by considering more semantic and linguistic information of emotions for fine-grained emotion analysis.

## REFERENCES

[1] Y.-C. Chang, C.-H. Chu, C. C. Chen, and W.-L. Hsu, ''Linguistic template extraction for recognizing reader-emotion,'' *Int. J. Comput. Linguistics Chin. Lang. Process.*, vol. 21, no. 1, pp. 29–50, 2016.

[2] Z. Deyu, X. Zhang, Y. Zhou, Q. Zhao, and X. Geng, ''Emotion distribution learning from texts,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 638–647.

[3] W. Gao, S. Li, S. Y. M. Lee, G. Zhou, and C.-R. Huang, ''Joint learning on sentiment and emotion classification,'' in *Proc. 22nd ACM Int. Conf. Conf. Inf. Knowl. Manage.*, 2013, pp. 1505–1508.

[4] J. Xu, R. Xu, Q. Lu, and X. Wang, ''Coarse-to-fine sentence-level emotion classification based on the intra-sentence features and sentential context,'' in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 2455–2458.

[5] S. Li, L. Huang, R. Wang, and G. Zhou, ''Sentence-level emotion classification with label and context dependence,'' in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, 2015, pp. 1045–1053.

[6] K. Gao, H. Xu, and J. Wang, ''A rule-based approach to emotion cause detection for Chinese micro-blogs,'' *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4517–4528, 2015.

[7] L. Gui, L. Yuan, R. Xu, B. Liu, Q. Lu, and Y. Zhou, ''Emotion cause detection with linguistic construction in Chinese weibo text,'' in *Natural Language Processing and Chinese Computing*. Berlin, Germany: Springer, 2014, pp. 457–464.

[8] L. Gui, D. Wu, R. Xu, Q. Lu, and Y. Zhou, ''Event-driven emotion cause extraction with corpus construction,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1639–1649.

[9] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, and J. Du, ''A question answering approach to emotion cause extraction,'' in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1593–1602.

[10] S. Y. M. Lee, Y. Chen, and C.-R. Huang, ''A text-driven rule-based system for emotion cause detection,'' in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text*, 2010, pp. 45–53.

[11] Y. Chen, S. Y. M. Lee, S. Li, and C.-R. Huang, ''Emotion cause detection with linguistic constructions,'' in *Proc. 23rd Int. Conf. Comput. Linguistics*, 2010, pp. 179–187.

[12] D. Ghazi, D. Inkpen, and S. Szpakowicz, ''Detecting emotion stimuli in emotion-bearing sentences,'' in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*. Berlin, Germany: Springer, 2015, pp. 152–165.

[13] I. Russo, T. Caselli, F. Rubino, E. Boldrini, and P. Martínez-Barco, ''EMOCause: An easy-adaptable approach to emotion cause contexts,'' in *Proc. 2nd Workshop Comput. Approaches Subjectivity Sentiment Anal.*, 2011, pp. 153–160.

[14] X. Cheng, Y. Chen, B. Cheng, S. Li, and G. Zhou, ''An emotion cause corpus for chinese microblogs with multiple-user structures,'' *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 17, no. 1, p. 6, 2017.

[15] R. Mulkar-Mehta, C. Welty, J. R. Hoobs, and E. Hovy, ''Using granularity concepts for discovering causal relations,'' in *Proc. FLAIRS Conf.*, 2011, pp. 1–6.

[16] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[17] P. Ekman, ''Expression and the nature of emotion,'' *Approaches Emotion*, vol. 3, pp. 319–344, Jan. 1984.

[18] B. V. Johnston and J. H. Turner, ''On the origins of human emotions: A sociological inquiry into the evolution of human emotions. Stanford University Press, California, 2000. $49.50 cloth, $17.95 paper,'' *Sociol. Focus*. vol. 34, no. 3, pp. 321–322, 2001.

[19] R. Descartes, *The Passions of the Soul*. Indianapolis, IN, USA, Hackett Publishing, 1989.

[20] W. James, ''What is an emotion?'' *Mind*, vol. 9, no. 34, pp. 188–205, 1884.

[21] R. Plutchik, *Emotion: A Psychoevolutionary Synthesis*. New York, NY, USA: Harpercollins College Division, 1980.

[22] A. Wierzbicka, *Emotions Across Languages and Cultures: Diversity and Universals*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[23] W. Li and H. Xu, ''Text-based emotion classification using emotion cause extraction,'' *Expert Syst. Appl.*, vol. 41, no. 4, pp. 1742–1749, 2014.

[24] K. Gao, H. Xu, and J. Wang, ''Emotion cause detection for Chinese micro-blogs based on ECOCC model,'' in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Berlin, Germany: Springer, 2015, pp. 3–14.

[25] T.-Y. Liu, ''Learning to rank for information retrieval,'' *Found. Trends Inf. Retrieval*, vol. 3, no. 3, pp. 225–331, 2009.

[26] T.-Y. Liu, J. Xu, T. Qin, W. Xiong, and H. Li, ''LETOR: Benchmark dataset for research on learning to rank for information retrieval,'' in *Proc. SIGIR Workshop Learn. Rank Inf. Retr.*, vol. 310, 2007, pp. 1–8.

[27] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, and H. Li, ''Learning to rank: From pairwise approach to listwise approach,'' in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 129–136.

[28] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang, and H.-W. Hon, ''Adapting ranking SVM to document retrieval,'' in *Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2006, pp. 186–193.

[29] F. Xia, T.-Y. Liu, J. Wang, W. Zhang, and H. Li, ''Listwise approach to learning to rank: Theory and algorithm,'' in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1192–1199.

[30] J. Xu, T.-Y. Liu, M. Lu, H. Li, and W.-Y. Ma, ''Directly optimizing evaluation measures in learning to rank,'' in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2008, pp. 107–114.

[31] M. Zhang, Y. Liu, H. Luan, and M. Sun, ''Listwise ranking functions for statistical machine translation,'' *IEEE/ACM Trans. Audio, Speech Language Process.*, vol. 24, no. 8, pp. 1464–1472, Aug. 2016.

[32] C. N. D. Santos, B. Xiang, and B. Zhou. (2015). ''Classifying relations by ranking with convolutional neural networks.'' [Online]. Available: https://arxiv.org/abs/1504.06580

[33] C. Shen and T. Li, ''Learning to rank for query-focused multi-document summarization,'' in *Proc. IEEE 11th Int. Conf. Data Mining (ICDM)*, Dec. 2011, pp. 626–634.

[34] D. M. Blei, A. Y. Ng, and M. I. Jordan, ''Latent Dirichlet allocation,'' *J. Mach. Learn. Res.*, vol. 3, pp. 601–608, Jan. 2002.

[35] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, ''Indexing by latent semantic analysis,'' *J. Amer. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, 1990.

[36] M. Lease, J. Allan, and W. B. Croft, ''Regression rank: Learning to meet the opportunity of descriptive queries,'' in *Proc. Eur. Conf. Inf. Retr.*, 2009, pp. 90–101.

[37] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, ''An efficient boosting algorithm for combining preferences,'' *J. Mach. Learn. Res.*, vol. 4, pp. 933–969, Dec. 2003.

[38] C. J. Burges, ''From ranknet to lambdarank to lambdamart: An overview,'' *Learning*, vol. 11, p. 81, Jun. 2010.

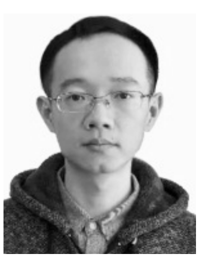
[39] C. J. Burges, R. Ragno, and Q. V. Le, ''Learning to rank with nonsmooth cost functions,'' in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 193–200.

[40] J. H. Friedman, ''Greedy function approximation: A gradient boosting machine,'' *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[41] J. Xu, R. Xu, Y. Zheng, Q. Lu, K.-F. Wong, and X. Wang, ''Chinese emotion lexicon developing via multi-lingual lexical resources integration,'' in *Proc. Int. Conf. Intell. Text Process. Comput. Linguistics*, 2013, pp. 174–182.

[42] T. Mikolov, K. Chen, G. Corrado, and J. Dean. (2013). ''Efficient estimation of word representations in vector space.'' [Online]. Available: https://arxiv.org/abs/1301.3781

[43] Y. Kim, ''Convolutional neural networks for sentence classification,'' in *Proc. Empirical Methods Natural Lang. Process.*, 2014, pp. 1746–1751.



**BO XU** received the B.Sc. and Ph.D. degrees from the Dalian University of Technology, China, in 2011 and 2018, respectively, where he is currently a Postdoctoral Research Associate with the Faculty of Electronic Information and Electrical Engineering. His current research interests include information retrieval and natural language processing.

**HONGFEI LIN** received the B.Sc. degree from Northeastern Normal University, in 1983, the M.Sc. degree from the Dalian University of Technology, in 1992, and the Ph.D. degree from the Northeastern University, in 2000. He is currently a Professor with the School of Computer Science and Technology, Dalian University of Technology, where he is also the Director of the Information Retrieval Laboratory. He has published more than 100 research papers in various journals, conferences, and books. His research interests include information retrieval, text mining for biomedical literatures, biomedical hypothesis generation, information extraction from huge biomedical resources, and learning-to-rank.

**YUAN LIN** received the B.Sc. and Ph.D. degrees from the Dalian University of Technology, China, in 2006 and 2012, respectively, where he is currently an Associate Professor with the School of Public Administration and Law. His current research interests include information retrieval and learning-to-rank.

**YUFENG DIAO** received the B.Sc. degree from Inner Mongolia University, in 2009, and the master's degree from the Dalian University of Technology, China, in 2012, where she is currently pursuing the Ph.D. degree with the School of Computer Science and Technology. Her current research interests include sentiment analysis and natural language processing.

**LIANG YANG** received the B.Sc. and Ph.D. degrees from the Dalian University of Technology, China, in 2009 and 2017, respectively, where he is currently a Lecturer with the School of Computer Science and Technology. His current research interests include sentiment analysis and natural language processing.

**KAN XU** received the B.Sc. and master's degrees from the Dalian University of Technology, China, in 2005 and 2009, respectively, where he is currently a Senior Engineer with the School of Computer Science and Technology. His current research interests include patent retrieval and learning-to-rank.