Knowledge-Based Systems 228 (2021) 107225

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Two-stage supervised ranking for emotion cause extraction

Bo Xu [a,c], Hongfei Lin [a,*], Yuan Lin [b], Kan Xu [a]

[a] School of Computer Science and Technology, Dalian University of Technology, Dalian, China
[b] WISE Lab., Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian, China
[c] State Key Lab. for Novel Software Technology, Nanjing University, P.R. China

ARTICLE INFO

Article history:
Received 10 December 2020
Received in revised form 7 April 2021
Accepted 9 June 2021
Available online 12 June 2021

Keywords:
Emotion cause extraction
Sentiment analysis
Ranking model
Natural language processing

ABSTRACT

Textual emotion analysis is a challenging research topic in the field of natural language processing (NLP), which plays an important role in related NLP tasks, such as opinion mining and personalized recommendation. Existing research on emotion analysis has focused mostly on detecting types of emotions, and has solved problems using classification-based methods. Recently, fine-grained emotion analysis has attracted the attention of researchers for probing the essential elements of emotions, such as the causes, experiencers and results of emotion events, which could help further elucidate textual emotions in more depth. In this paper, we focus on the task of emotion cause extraction, aiming to recognize the causes in sentences that provoke certain emotions. We propose a two-stage supervised ranking method for accurately extracting the emotion causes based on information retrieval techniques. In the first stage, we measure the complexity of provoked emotions using query performance predictors to distinguish the number of causes for each emotion in contexts. In the second stage, we incorporate the emotion complexity into learning an autoencoder-enhanced ranking model for accurately extracting the causal clauses. We also extract abundant emotion-level clause features for clause representations as the learning samples. We evaluate the proposed method on an existing dataset for emotion cause extraction and demonstrate that our method significantly outperforms the state-of-the-art baseline methods. The proposed method is effective in extracting textual emotion causes in sentences, which can greatly benefit in-depth emotion analysis for effective cognitive computing.

1. Introduction

Cognitive science is an iconic scientific research category that has emerged in recent years. It has attracted the attention of scientists all over the world as a cutting-edge interdisciplinary research field. Related research has focused on different cognitive processes and human behaviors involving perception, learning, reasoning and emotions. Emotions are among the most important cognitive processes for interpreting human behaviors.

In recent years, research has widely focused on human emotions through computational methods. Various computing methods have been proposed to identify the types of emotions from texts [1–8]. Textual emotion analysis has become a fundamental and challenging research topic in the field of natural language processing (NLP). Applications based on emotion analysis have been widely used in opinion mining and recommendation systems for user profiling and personalization. Existing studies on emotion analysis have mostly focused on detecting different types of emotions based on various machine learning classifiers. Recently, fine-grained emotion analysis beyond emotion classification has attracted much attention from NLP researchers for its use in investigating the essential elements of emotions, such as the causes, experiencers and results of emotion events. Extracting these elements of emotions is helpful for understanding emotions in depth. Emotion cause extraction, as one of the most important tasks in emotion analysis, has been considered in recent studies for interpreting how emotions are provoked and perceived. However, since emotion expressions are ambiguous and subtle, it is difficult to accurately extract the causes of provoked emotions.

Formally, in natural language, a textual emotion event is embedded in a certain sentence involving a provoked emotion. The sentence contains multiple clauses including the emotional clause and several other clauses closely related to the corresponding emotion. The essential elements of the emotion are contained in these nearby clauses. The goal of emotion cause extraction is to find the clause or clauses involving the emotion causes within the sentence of the provoked emotion.

To extract the clauses containing the emotion causes, early studies adopted the rule-based and classification-based methods to distinguish the causal clauses and the non-causal clauses [9–18]. However, information within emotional sentences has

* Corresponding author.
E-mail addresses: xubo@dlut.edu.cn (B. Xu), hflin@dlut.edu.cn (H. Lin).

https://doi.org/10.1016/j.knosys.2021.107225
0950-7051/

been partly overlooked by directly performing the clause-level classification because classifiers treat clauses as training samples regardless of the sentence or other emotion-level information. How to effectively incorporate the emotion information into cause extraction remains an unsolved problem. Supervised ranking methods, such as learning to rank in information retrieval, can be used to solve this problem.

Learning to rank is a series of ranking methods widely used in information retrieval that produces ranking lists of documents for given queries. Ranking models based on learning to rank are learned in a supervised manner. The training set for learning to rank consists of multiple subsets of documents with respect to different queries. The goal of training is to predict the ranking list of documents for each query. Since emotion cause extraction aims to extract the causal clauses for each emotional sentence, this work aims to tackle the task of emotion cause extraction using learning-to-rank methods from an information retrieval perspective, which transforms this task as a ranking problem. In the transformed problem, emotions are analogous to queries, and clauses are analogous to documents for training the ranking models.

To better explain the motivation of our work, we provide a toy example of emotion cause extraction in Fig. 1. The example sentence contains four clauses with respect to the provoked emotion 'happy'. The emotional sentence involves the cause of the emotion in the clause 'my favorite movie will be released'. The goal of this task is to extract this causal clause. To achieve the goal, we propose a two-stage supervised ranking method to tackle the problem. The proposed method aims to sort the causal clause at the top of the clause-ranking list, and extract the top-ranked clause as the causal clause for fine-grained emotion analysis.

The proposed two ranking-based stages in our method aim to accurately locate the causal clause or clauses in emotional sentences. To better explain the motivation of our two-stage ranking method, we summarize the reasons for adopting two stages as follows. Statistics in previous studies [19] have shown that more than 90% of emotional sentences contain one causal clause, while other emotional sentences contain two or more causal clauses. Existing work has mostly ignored the number of causal clauses when tackling the problem of emotion cause extraction because distinguishing the number of causal clauses is difficult for classical classifiers given data imbalance. However, determining the number of causal clauses is crucial for improving the accuracy of emotion cause extraction. In this work, we address the number of causal clauses as the complexity of the given emotional sentences by using information retrieval techniques. In information retrieval, large data imbalance of relevant and irrelevant documents exists and has been well dealt with using query performance predictors. Query performance predictors seek to rank candidate queries based on their latent retrieval performance. To measure the complexity of emotional sentences, we adopt query performance predictors to determine the number of emotional causal clauses as the first stage that ranks emotional sentences based on their complexity. Then, we adopt learning-to-rank methods to rank the candidate clauses as the second stage in consideration of the complexity of emotional sentences. We believe that the two-stage ranking method would contribute to accurate extraction of emotion causes.

Specifically, in the first stage, we measure the complexity of provoked emotions using query performance predictors to distinguish the number of causal clauses for each emotional sentence. In the second stage, we incorporate the emotion complexity into learning an autoencoder-enhanced ranking model for candidate clause ranking. We extract abundant emotion-related clause features for clause representations in the learning process. The main contributions of this work are listed as follows.

(1) We propose a two-stage supervised ranking model to improve the accuracy of emotion cause extraction by emotion-oriented clause ranking from an information retrieval perspective. The two stages of our method incorporate abundant emotional semantic information, and are beneficial to completely extract the causal clauses of certain provoked emotion for fine-grained emotion analysis.

(2) We adopt query performance predictors to measure the complexity of emotions in the first stage, which can help discriminate the number of causes of given emotional sentences. The complexity of emotions is then fed into the second stage of ranking. The second stage adopts an autoencoder-enhanced ranking model for effective clause ranking in consideration of emotional complexity.

(3) We conduct extensive experiments to examine the performance of our method in terms of various ranking features, pre-processing strategies and different approaches to learning-to-rank models. The experimental results demonstrate the effectiveness of our method in extracting emotional causes, which significantly outperforms other state-of-the-art models.

The rest of the paper is organized as follows. Section 2 provides related work; Section 3 introduces the details of our method; Section 4 evaluates the proposed model and discusses the experimental results; and Section 5 concludes the paper and outlines future work.

2. Related work

Textual emotion analysis, as an important task in natural language processing (NLP), has been used in existing research mainly for detecting types of emotions [4–8,20–22]. For example, Zhou et al. [5] proposed a novel emotion distribution learning approach to measure the intensities of multiple emotions for emotion classification. Xu et al. [7] developed a coarse-to-fine analysis strategy to optimize emotion classification based on sentence-level textual similarity and sentence adjacency. Li et al. [8] proposed a graph inference method to classify emotional sentences based on the dependence of labels and contexts. These studies indicate that emotion analysis is highly useful for opinion mining and personalized recommendation on social media, while fine-grained emotion analysis can enhance the interpretation of emotions to gain further improvement in different NLP tasks.

To capture fine-grained emotional information, recent works have attempted to extract essential elements of emotions beyond emotion classification. These elements include the experiencer, causes and latent impacts of emotions, which depict detailed aspects of emotions in depth and contribute to a comprehensive understanding of textual emotions. Emotion causes refer to the reasons or stimulus that provoke certain emotions, which can significantly affect the trends of emotions [23–25]. Emotion cause extraction has therefore attracted much attention in recent studies, and has been dealt with based on linguistic rules and supervised classification methods. Rule-based methods seek to design useful linguistic rules to identify emotion cues in natural language expressions. For example, Lee et al. [13] detected the causes for provoked emotions by annotating emotion events and developing linguistic rules. Neviarouskaya et al. [26] employed syntactic and dependency parser and rules for the analysis of eight types of the emotion-cause linguistic relations to detect causal phrases. These rule-based methods can achieve the desired performance particularly on microblog texts [9,27,28]. To extract emotion causes from microblogs, Cheng et al. [17] integrated multiple-user structures in texts to detect emotion cause cues. However, the designed rules are always tailored for specific tasks with limited generalization ability. To enhance the generalization ability, supervised classification methods have been introduced to extract emotion causes with linguistic rule-based textual features. Chen et al. [14] proposed to use a maximum entropy classifier for emotion cause extraction. Their study

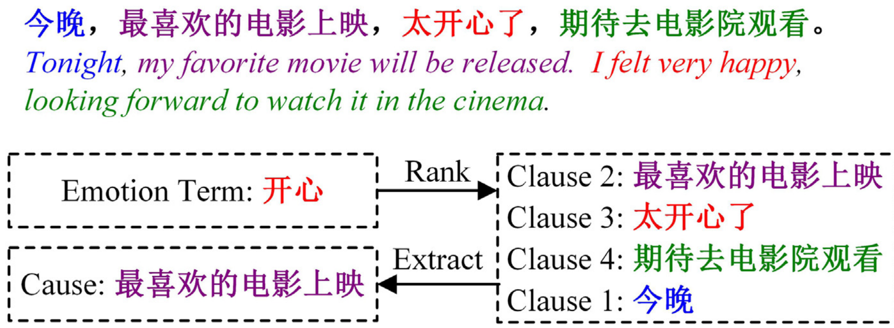

Fig. 1. A top example for ranking-based emotion cause extraction.

indicated that rule-based features exhibited better generalization ability than directly using the rules. Furthermore, other classifiers, such as support vector machines (SVMs) and conditional random fields (CRFs), have also been investigated in emotion cause extraction [10,15]. These classification methods aim to distinguish causal or noncausal clauses in emotional sentences. These methods may partly overlook emotion information with respect to the causal clauses because clauses for different emotions are treated equally in the classification process. To further incorporate emotion information, Gui et al. [11] proposed a novel SVM classifier with convolutional kernels extracted from syntactic trees of emotional sentences, which largely enhanced the extraction performance. In subsequent studies, emotion cause extraction was treated as a question answering problem, and was solved based on memory networks [12]. Recently, a new task, emotion-cause pair extraction, has been proposed and dealt with using different models [29–33]. Hu et al. [34] proposed a graph convolutional network over the inter-clause dependency to incorporate semantic and structural information for emotion cause extraction. Li et al. [35] proposed a multi-attention-based neural network model to consider the emotional context and interaction between emotional clause and candidate clauses.

Common sense knowledge, as a valuable resource in the understanding of natural language, has also been used for emotion cause extraction. For example, Russo et al. [16] enriched linguistic patterns of emotion causes with common sense knowledge for cause extraction from Italian newspaper articles. Ulkar-mehta et al. [18] inferred the textual causal relationships based on granularity relations. Existing studies on emotion cause extraction have deepened our understanding of emotions in natural language to some extent. However, there is still much room for improving the task. How to incorporate the relationship of clauses in emotional sentences into the learned model remains a great challenge. In fact, clauses in emotional sentences are closely correlated to each other and should be considered as a whole in model optimization. For two clauses from different sentences, their relationship is of little use in cause extraction. Improvement could be made by separating the training samples into smaller units for model selection. Therefore, we proposed the adoption of learning-to-rank methods for emotion cause extraction.

Learning-to-rank methods have been widely used in information retrieval (IR) tasks in the intersection of supervised machine learning and IR [29,36–40]. Ranking models have been constructed and applied in different IR and NLP tasks, such as machine translation [41], relation extraction [42], multidocument summarization [43] and emotion detection [44]. For example, Lin et al. [44] proposed two ranking approaches to detect the most popular emotions of readers based on pairwise ranking loss minimization and emotional distribution regression. These studies have shown that learning-to-rank methods are effective in ranking candidate items with query constraints. The query constraints can be analogous to the emotion constraints for emotion cause extraction task, namely, refining the causal clauses within emotional sentences. Therefore, we believe that learning-to-rank methods can deal with emotion cause extraction to further enhance its performance.

Our previous work has focused on computing methods for solving natural language processing tasks, including hot topic detection [45] and user attribute classification [46]. The difference between this work and previous works lies in the tasks and methods. In this work, we focus on emotion cause extraction for fine-grained emotion analysis, whereas our previous work focused on other NLP tasks. Since we address different tasks in these studies, we design different methods in consideration of task-specific characteristics. In this work, we propose a two-stage ranking method for emotion cause extraction, which differs from our previous methods.

3. Two-stage supervised ranking for emotion cause extraction

3.1. Problem formalization

We formalize the emotion cause extraction task in this section based on the definition in [11]. Emotion cause extraction seeks to locate the emotion-provoking clause in an emotional sentence for fine-grained emotion analysis. Emotional sentences are annotated with emotion words for emotion analysis and opinion mining. Unlike traditional emotion analysis that classifies emotion types, emotion cause extraction involves more in-depth emotional elements, making this task more challenging.

Formally, a set of N emotional sentences are annotated in advance for cause extraction, which can be denoted $E = \{e_1, e_2, \ldots, e_N\}$. In the set of sentences, a general emotional sentence e comprises M clauses denoted $C = \{c_1, c_2, \ldots, c_M\}$, which includes the emotional clause and the causal clauses. These clauses are annotated with ground truth labels $L = \{l_1, l_2, \ldots l_M\}$, indicating whether or not the clause can provoke the emotion in the emotional sentence. The goal of the task is to predict which clause or clauses provoke the emotion in each sentence without ground truth labels.

Inspired by relevance-based document retrieval in information retrieval tasks, we treat emotion cause extraction as a ranking problem, and we solve the task based on supervised ranking methods. In the proposed method, emotions are analogous to queries and candidate causal clauses are analogous to documents. Since learning-to-rank methods have achieved powerful performance for ranking the most relevant documents at the top of the ranking lists, we believe that ranking-based cause extraction will yield ideal results for emotion cause extraction. We illustrate the transformation problem in Fig. 2 to depict the motivation of our work.

In this study, we propose a two-stage supervised ranking model for emotion cause extraction. The first stage proposes an adaptive cause complexity predictor (ACCP) to address multi-cause emotion prediction based on query performance predictors.

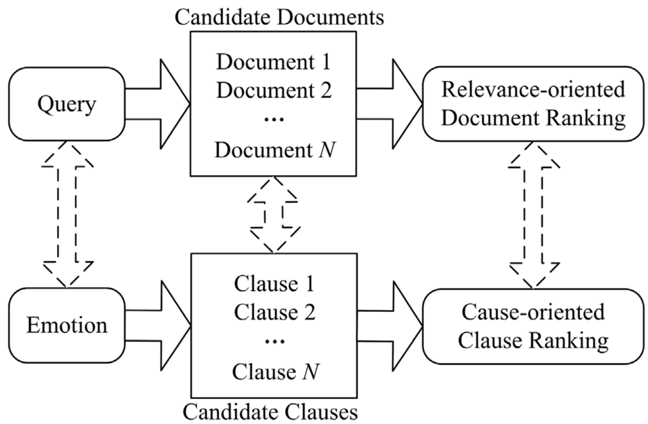

Fig. 2. Problem transformation from relevance-based document ranking to cause-based clause ranking.

Emotional sentences involving single and multiple causes are treated in different manners for clause-ranking model construction. In the second stage, we propose an emotion-constrained clause-ranking model based on autoencoders to enhance emotion cause extraction performance. Different types of clause features are investigated in this stage for constructing effective ranking models.

3.2. Adaptive Cause Complexity Predictor (ACCP)

People's emotional expressions are complex and changeable in daily life. Different factors may work together to stimulate certain emotions. In text-based emotion analysis, these factors are embedded in single or multiple clauses of emotional sentences as emotion causes. To accurately extract emotion causes, we seek to distinguish the complexity of certain emotions as the first step. We distinguish the complexity of emotions according to the number of the corresponding emotion causes. Namely, emotions are provoked by a single cause or multiple causes. The complexity of emotions is then considered in constructing the clause-ranking models to yield better extraction performance in the second stage.

Statistics related to emotion cause extraction [19] have indicated that more than 90% of emotional sentences contain only one causal clause, while other emotional sentences contain two or more causal clauses. The imbalanced distribution of causal clauses among emotional sentences has been partly ignored in previous studies. If the number of causal clauses can be fully considered in emotion cause extraction, the extraction performance may be further improved. To this end, we adopt query performance predictors in the first-stage ranking, and we refer to the proposed method as adaptive cause complexity predictor (ACCP). In information retrieval, relevant and irrelevant documents are largely imbalanced. Therefore, we believe that the data imbalance problem in emotion cause extraction can be dealt with using query performance predictors.

Query performance predictors stem from information retrieval tasks [47], which predict query performance in advance of query expansion and relevance feedback to enhance the effectiveness and efficiency of retrieval systems. Given that query performance predictors are powerful in measuring the query performance in IR tasks, we propose to incorporate query performance predictors into discriminating the complexity of emotions with a single cause or multiple causes. In constructing the ranking model, emotions with a single cause are ranked based on the top-one optimization strategy, and emotions with multiple causes are ranked based on the top-k optimization strategy.

Table 1
Emotional sentence features based on query performance predictors. In the table, e represents a given emotional sentence, t represents any term in e, S is the document containing the sentence e and other sentences, C is the entire corpus, N is the number of terms in the corpus, N_t is the number of term t in the corpus, and i represents any sentence in S.

Description	Formulation
Sentence length	$\|e\|$
Emotion entropy	$\sum_{t\in e} -P(t\|S)log_2P(t\|C)$
Emotion clarity 1	$\sum_{t\in e} -P(t\|S)log_2 \frac{P(t\|S)}{P(t\|C)}$
Emotion clarity 2	$\sum_{t\in e} -P(t\|e)log_2 \frac{P(t\|e)}{P(t\|C)}$
Sentence radius	$\frac{1}{\|S\|} \sum_{i\in S} \sum_{t\in i} P(t\|i)log \frac{P(t\|i)}{\frac{1}{\|S\|}\sum_{i\in SP(t\|i)}}$
Variance of inverse emotion Freq.	$var(idf(t\in e)), idf(t) = \frac{log_2(\|C\|+0.5)}{\|S\|log_2(\|C\|+1)}$
Max of inverse emotion Freq.	$maxidf(t\in e)$
Average of inverse term Freq.	$\frac{1}{\|e\|} log_2 \prod_{t\in e} \frac{N}{N_t}$
Sum of emotion collection similarity	$\sum_{t\in e}(1+log_2N_t)log(1+\frac{\|C\|}{\|S\|})$
Max of emotion collection similarity	$max_{t\in e}(1+log_2N_t)log(1+\frac{\|C\|}{\|S\|})$

We utilize support vector machines with Gaussian kernel for learning the query performance predictors. The adopted emotional sentence features based on query performance predictors are listed in Table 1, which are all widely used features for query performance predictors in existing studies [48–51]. We provide detailed definitions of these as follows.

Emotional sentence length counts the number of words in each sentence. Intuitively, long sentences may contain more complex emotions than short ones. Emotion entropy captures the divergence of distributions of certain emotions between the document containing the emotional sentence and the entire corpus. This feature is helpful for characterizing the frequency of certain emotions in text. Similarly, the features based on emotion clarity, which stem from query clarity, are used to compute the Kullback–Leibler divergence between the emotion and collection language models. Sentence radius is calculated to capture the average frequencies of terms in emotional sentences. Term frequency is an important indicator that reflects the complexity and ambiguity of emotional sentences. The variance and maximum of the inverse emotion frequency characterize the emotion complexity in a manner similar to sentence radius. These two features are designed based on inverse document frequency (IDF), which is widely used in information retrieval tasks for measuring the importance of terms. The average of the inverse term frequency measures the emotion complexity at the term level, counting the term distribution in the entire corpus. The sum and maximum of emotion collection similarity are used to capture the differences in term distribution between the document containing the emotional sentence and the entire corpus. These predictors have been proven to be effective in predicting the query performance in IR tasks. Therefore, we believe the modified features for measuring the emotion complexity can effectively distinguish the emotional sentences with different numbers of causal clauses.

3.3. Emotion-constrained Clause Ranking (ECCR)

We propose learning to rank candidate clauses within emotional sentences for cause extraction. The inputs of the ranking models are feature representations of clauses. To obtain effective ranking models for different emotions, we propose an autoencoder-enhanced ranking model. The proposed model seeks to optimize the ranking process for emotions simultaneously with a single cause and multiple causes. The learned model can predict the causal clauses of given emotional sentences for emotion cause extraction. Next, we provide details on clause feature definitions and the proposed ranking model.

Table 2
Definitions of core term categories.

ID	Category	Core terms
1	Causal conj.	因为(之所以)/because, 因(而)/due to, 由于/because of
2	Causal verbs	让/make, 令/cause, 使/let
3	Sensory verbs	想到/recall, 听到/hear, 看到/see, 感到/sense, 注意到/notice
4	Emotion terms	激动/excited, 快乐/happy, 愤怒/angry, 怨恨/resentful 惊讶/surprised, 恐惧/fearful, 悲伤/sad,
5	Negation	不/not, 没有/no
6	Family nouns	丈夫/husband, 妻子/wife, 儿子/son, 女儿/daughter, 父亲/father, 母亲/mother, 叔叔/uncle, 姑姑/aunt

3.3.1. Clause features for emotion cause extraction

Feature representations of clauses are used as the inputs for ranking model construction. High-quality features can greatly contribute to improving ranking performance. Two types of clause features are defined for emotion cause extraction in this work, namely, emotion-independent and emotion-dependent features. These two types of features are stemmed from the query-independent and query-dependent features for document representations used in retrieval tasks [36]. Emotion-independent features are used to measure the clause importance in terms of linguistics and semantics in an emotional sentence. Emotion-independent features are based mainly on textual statistics to discriminate causal clauses and noncausal clauses. Emotion-dependent features are used to reflect the relationship between emotional clauses and candidate clauses in sentences. These features are designed to capture the relationship between causal clauses and their corresponding emotions. We introduce the detailed definitions of features below.

We extract emotion-independent features based on clause length, part-of-speech (POS) and core terms. Clause length closely correlates to the amount of information in a certain clause. Intuitively, long clauses are more likely to contain useful information of emotions, while short clauses tend to contain less information. To consider the length of clauses, we extract features based on clause length in terms of characters and words, respectively. POS tagging is an important method used in NLP tasks for capturing linguistic information of texts. We consider the POS tags of words in each clause, and extract features based on the number and proportion of different POS tags in clauses. Moreover, we consider core terms within clauses. Core terms refer to the cause-related terms, which can be taken as a strong indicator for causal clauses [13]. Lee et al. [13] identified seven groups of linguistic cues for rule-based cause extraction. In a similar manner, we identify six groups of core terms in this work for Chinese text processing. We summarize the core terms in Table 2, including causal conjunctions, sensory verbs, emotion terms, causal verbs, negations and family-related nouns. Causal conjunctions, causal verbs and sensory verbs have been identified as effective linguistic cues for emotion cause extraction in English [13]. Therefore, the core terms in Chinese can help capture the causal information. Emotion terms are the most important information in emotional sentences. Different emotion terms may correspond to different types of causal information. Negations are an important factor for emotion analysis [52] because negative sentences often have a certain emotional orientation. Family nouns are always used in Chinese dialog among family members. Due to the intimacy among family members, dialogs often involve certain emotions. These emotion-independent features are extracted to reflect clause importance by the linguistics and semantics of sentences.

We extract emotion-dependent features based on relative position, word embedding and topic model. Relative position refers to the closeness of each clause to the emotional clause in a

Table 3
Detailed definitions of clause features.

ID	Feature definition	Feature template
1	Character-level clause length	#characters
2	Word-level clause length	#words
3	The number of nouns	#nouns
4	The number of verbs	#verbs
5	The number of adjectives	#adjectives
7	The number of adverbs	#adverbs
8	The ratio of nouns	#nouns/#words
9	The ratio of verbs	#verbs/#words
10	The ratio of adjectives	#adjectives/#words
11	The ratio of adverbs	#adverbs/#words
12	The number of causal conjunctions	#causal conjunctions
13	The number of causal verbs	#causal verbs
14	The number of sensory verbs	#sensory verbs
15	The number of emotion terms	#emotion terms
16	The number of negations	#negations
17	The number of family nouns	#family nouns
18	The ratio of causal conjunctions	#causal conjunctions/#words
19	The ratio of causal verbs	#causal verbs/#words
20	The ratio of sensory verbs	#sensory verbs/#words
21	The ratio of emotion terms	#emotion terms/#words
22	The ratio of negations	#negations/#words
23	The ratio of family nouns	#family nouns/#words
24	Relative position between clauses	$distance(c, e)$
25	The average word similarity	$average(similarity(c_w, e_w))$
26	The maximum word similarity	$maximum(similarity(c_w, e_w))$
27	The minimum word similarity	$minimum(similarity(c_w, e_w))$
28	The average clause similarity	$similarity(c, e)$
29	The LDA clause-level similarity	$clauseLDA(c, e)$
30	The LDA emotion-level similarity	$emotionLDA(c, e)$
31	The LDA document-level similarity	$documentLDA(c, e)$
32	The LSI clause-level similarity	$clauseLSI(c, e)$
33	The LSI emotion-level similarity	$emotionLSI(c, e)$
34	The LSI document-level similarity	$documentLSI(c, e)$

sentence. Studies have shown that causal clauses are more likely to be located next to the emotional clause [11]. Word embeddings have been proven effective in capturing the semantic information of words in NLP tasks. We adopt word embeddings to represent candidate clauses for emotion cause extraction. The clause representations are fed into the computation of similarity between candidate clauses and the emotional clause. We consider the maximum, average and minimum of similarities as clause features. In addition, we adopt topic models for computing the similarity among clauses. Two topic models, latent dirichlet allocation (LDA) [53] and latent semantic indexing (LSI) [54], are investigated for feature extraction. Topic models are learned in terms of clauses, sentences and documents as different features. These emotion-dependent features are extracted to reflect the relationship between emotional clauses and candidate clauses.

The feature representations of clauses are then treated as the inputs for model training. Detailed definitions of clause features are shown in Table 3 to help understand the definitions of clause-ranking features.

Since the clause features are extracted for cause extraction from different perspectives, there is no significant overlap among them. For example, the features based on core terms are identified to capture the emotional causal information, and the word embedding-based features are designed to measure the similarity between terms in emotional sentences. Even though they both consider the distribution of terms in sentences, they capture different information for model construction. Compared with the emotion-level features defined in Table 1, the clause features are more fine-grained for cause-oriented clause ranking. After feature extraction, we normalize the feature values to improve the robustness of the learned models. Normalization can be analogous to the query-level normalization used in learning-to-rank tasks 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; and min_e and max_e represent the minimum and maximum of the feature values with respect to the corresponding emotional sentence, respectively.

3.3.2. Autoencoder-enhanced ranking model for emotion cause extraction

In this section, we introduce an autoencoder-enhanced ranking model for emotion cause extraction, which was originally proposed to improve document ranking in our previous work [55]. The modified model for emotion cause extraction enriches the feature representations of clauses within emotional sentences using autoencoders, aiming to simultaneously improve the extraction accuracy for emotions with single cause and multiple causes. Details about the proposed model are described below.

Autoencoders, as the building blocks of neural networks, have been widely used to generate effective features in machine learning tasks. A general autoencoder includes one input layer, one hidden layer and one output layer, which encodes inputs into low-dimensional hidden representations and decodes the representations as outputs. The performance of autoencoders is evaluated based on the reconstruction ability, namely, the deviations between the input and output representations. Different loss functions can be used to measure the deviations. We formulate the general loss function as follows and propose incorporating the ranking constraints into the loss function for emotion cause extraction.

$$loss = \sum_{i=1}^{n} \|x_i - \hat{x}_i\|_2^2 \tag{2}$$

where n is the number of training emotional sentences. The Euclidean distance is adopted to measure the deviations between the input representations and the outputted representations. The learning process aims to minimize the loss for optimal parameters. To adapt autoencoders for emotion cause extraction, we consider the feature importance for training autoencoders based on Bregman divergence, and we modify the loss function as follows.

$$loss = \sum_{i=1}^{n} \theta^T (x_i - \hat{x}_i)^2 \tag{3}$$

where θ is a pretrained feature weighting vector that can be learned based on a pretrained ranking model based on ListNet [38]. ListNet is a listwise learning-to-rank method that learns a linear ranking model. The model aims to obtain the weights of different features and employs a probabilistic ranking loss function based on permutation probability. The loss function can be defined as follows.

$$loss(y, z(f(x))) = \sum_{j=1}^{n(i)} p_y(x_j) log(p_{z(f(x))}(x_j)) \tag{4}$$

where y is the ground truth ranking list, $z(f(x))$ is the predicted ranking list with the scoring function $f(x)$, $P_y(x_j)$ is the permutation probability on y, and $P_{z(f_w)}(x_j)$ is the permutation probability on $z(f(x))$, and $n(i)$ is the number of clauses for the ith emotion.

As mentioned above, the training data for emotion cause extraction consist of subsets of clauses with respect to different emotional sentences. The learning target is to extract the causal clause or clauses from each sentence, namely, learn a ranking model to predict the ranking list of clauses in each emotional sentence. Therefore, we incorporate another item into the loss function of autoencoders in consideration of emotional sentences.

Specifically, the pretrained ListNet ranker yields two ranking lists of documents based on the inputs and outputs of an autoencoder, respectively. The difference between these two lists of clauses measures the reconstruction capability of the autoencoder at the level of emotional sentence. We therefore incorporate the difference into the loss function of the autoencoder to guide the learning process for more effective ranking of clauses. Namely, if there exists a large difference, we believe that the reconstruction loss should be increased because the deviation would tend to be large. We formalize this method in the loss function below.

$$loss = \sum_{q \in Q} \eta(l_{in}^q, l_{out}^q)(\sum_{i=1}^{n(q)} \theta^T (\hat{x}_i - x_i)^2) \tag{5}$$

where $\eta(l_{in}^q, l_{out}^q)$ measures the differences between two ranking lists based on the inputs and outputs of an autoencoder and $n(q)$ is the number of documents corresponding to the query q. We accumulate the losses across all the queries in the training set in iterations. To reduce the loss, parameters including $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{b}_1$, and $\mathbf{b}_2$ are optimized for learning effective hidden representations. The problem is then transformed to the computation of $\eta(l_{in}^q, l_{out}^q)$ for measuring the difference in two ranking lists of documents.

To incorporate the emotion complexity into learning effective ranking models, we propose to measure the difference of two clause lists in terms of ranking performance. In IR, ranking performance can be measured based on evaluation metrics for optimizing the top-ranked clauses. We utilize the top-one and top-two optimization strategies for emotional sentences with single cause and multiple causes, respectively. We formalize this method as follows.

$$\eta(l_{in}^q, l_{out}^q) = \frac{|Eval_{in} - Eval_{out}|}{Eval_{in}} \tag{6}$$

where $Eval_{in}$ and $Eval_{out}$ are the ranking performances of the original inputs and the reconstructed outputs of an autoencoder evaluated by IR evaluation measures, denoted $Eval$. To accurately extract causal clause for different emotional sentences, we adopt Precision@1 and Precision@2 in our experiments for emotions with a single cause and multiple causes, respectively. We estimate the performance by applying the ListNet-based ranker on the corresponding features. Based on the equation, we incorporate emotion-level constraints into the loss function of the modified autoencoders for learning more effective clause-ranking models. The modified autoencoder can capture the latent information of the inputs through enhanced reconstruction capability at the emotion level and produce more compact and effective feature representations of clauses as complements of original features for improving the ranking performance. Finally, we employ ListNet to learn clause-ranking models using the enriched features for effectively extracting the emotion causes.

Table 4
Statistical information of the used dataset from the previous work [19].

Item	Number
Sentences	2,105
Clauses	11,799
Emotion causes	2,167
Sentences with 1 cause	2,046
Sentences with 2 causes	56
Sentences with 3 causes	3

4. Experiments and analysis

4.1. Experimental settings

In this section, we evaluate our method using a publicly accessible corpus for emotion cause extraction [11,19], which has been applied in existing studies of emotion cause extraction. The corpus was built from SINA News by extracting 15,687 sentences using keyword matching from raw data and refined in terms of emotional information. In the annotation process, two annotators manually annotated the emotion categories and the causes in the W3C Emotion Marup Language (EML) format. The kappa value on clause-level annotation is 0.9287, which means that this annotation is very reliable. More details on corpus construction and annotation can be found in [19]. The used corpus, which contains 2,105 sentences in total, stems from the extracted emotional sentences. Each sentence is characterized by a single emotional keyword and is separated into several clauses that contain the causal clauses of the emotion. The statistics of the corpus are provided in Table 4 obtained from the previous work by Gui et al. [19]. The table shows that the number of emotion causes is very imbalanced among the emotional sentences. Our method aims to address the problem of data imbalance using query performance predictors.

We adopt the word2vec tool to obtain the word embeddings using the original documents from the Sina News. Default configurations are set in our experiments. Part-of-speech taggings are obtained using the NLTK tool. The extraction performance is measured in terms of precision, recall and F-measure tailored for emotion cause extraction, which have been used in related studies [11,12,27,28]. In the tailored evaluation measures, if an extracted clause involves the annotated emotion cause, we consider the extraction to be correct. We formulate these evaluation measures as follows. The precision measure counts the ratio of correct causes out of the identified causes, and the recall measure counts the ratio of correct causes out of all the annotated causes.

$$Precision = \frac{\#Correct\ Cause}{\#Identified\ Cause} \tag{7}$$

$$Recall = \frac{\#Correct\ Cause}{\#Annotated\ Cause} \tag{8}$$

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

We conduct fivefold cross validations to obtain the average performance of ranking methods trained in different folds. Like the standard division of dataset for learning-to-rank tasks [37], we divide the emotional sentences into a training set, a validation set and a test set in a 3:1:1 by emotion IDs. The training set is used for model training, the validation set for parameter tuning, and the test set for emotion cause extraction. We report the average extraction performance of all folds in our experiments.

4.2. Baseline models

In our experiments, we compare the proposed method with baseline methods in three categories: rule-based, classification-based and neural network-based methods.

(1) **Rule-based methods**. The rule-based system (RB) was initially proposed by Lee et al. [13] and was enhanced using common-sense knowledge (CB) [16]. We compare with these two methods based on the emotion cognition lexicon [56] as the common-sense knowledge base.

(2) **Classification-based methods**. SVM classifiers with linguistic rule-based clause features (RB+CB+ML) are another baseline model based on machine learning (ML) [14]. The classification-based models adopted SVM classifiers based on diverse clause features including unigram, bigram, trigrams and word embedding of clauses [11,27,57]. We also compare our method with the state-of-the-art multikernel classifiers for emotion cause extraction [11].

(3) **Neural network-based methods**. These baselines include the convolutional neural network (CNN) [58] and the multiple-slot deep memory network (ConvMS-Memnet) [12], which has achieved the best performance in the emotion cause extraction tasks.

We compare the reported performance of baseline models from existing work [12] with the performance obtained in our experiments. More details about the baseline methods can be found by referring to the corresponding studies. Furthermore, we compare the proposed method with different learning-to-rank methods in one-stage and two-stage ranking scenarios. The methods include the pointwise regression method [59] and SVM with ranking features (denoted Regression and SVM-LTR), the pairwise RankBoost method [60] and the listwise LambdaMART method [61]. Ranking models are trained based on different combinations of the defined features to examine the feature importance for emotion cause extraction. We also examine the influence of stopword removal and clause feature normalization in our experiments.

4.3. Results and discussions

In this section, we first evaluate the overall performance of the proposed method and report the experimental results compared with the baseline methods. We then further analyze the performance of our method by conducting experiments based on different combinations of clause features to examine the feature importance in the task. We also examine the influence of preprocessing steps, including stopword removal and feature normalization, for emotion cause extraction in our method.

4.3.1. Overall performance on emotion cause extraction

The overall performance of our method and the performance of the baseline models are shown in Table 5. The table shows that the RB model achieves relatively high precision and low recall, while the CB model achieves relatively low precision and high recall. This result indicates that linguistic rules suffer from the generalization capability for emotion cause extraction, even though the models based on linguistic rules can yield high performance in terms of certain evaluation metrics. We also observe that balanced performance can be achieved by combining RB and CB. The performance is further enhanced by classification using linguistic rule-based features [14].

We observe that different classifiers yield diversified performance with different features. Specifically, textual features based on unigrams, bigrams and trigrams yield comparable results with the word representation-based features for SVM classifiers. Features based on linguistic rules and common sense knowledge

Table 5
Comparison with baseline models.

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
Multikernel	0.6588	0.6927	0.6752
ConvMS-Memnet	0.7076	0.6838	0.6955
Regression	0.7506	0.7291	0.7397
SVM-LTR	0.7672	0.7453	0.7561
RankBoost	0.7720	0.7499	0.7608
LambdaMART	0.7910	0.7683	0.7795
Proposed model	**0.8076***	**0.7845***	**0.7959***

*Indicate significant improvements over the ConvMS-Memnet model.

achieve better performance than the former two types of features. This finding indicates that accurate emotion cause extraction requires more linguistic information and common senses than statistical information. Among all the classifiers, the multikernel SVM achieves the best performance, which reflects the usefulness of syntactic information of sentences for emotion cause extraction.

Furthermore, the CNN model outperforms most classifiers except the multikernel SVM. The ConvMS-Memnet model yields better performance by F-measure and precision than other baseline models, which used memory slots to incorporate the context information of emotions for cause extraction. The performance of baseline models indicates that linguistic and contextual information highly contributes to emotion cause extraction. Therefore, we incorporate these two types of information in our methods for the construction of ranking models.

We further compare the proposed autoencoder-enhanced ranking model with other learning-to-rank methods. The comparison shows that all the ranking methods outperform other baseline methods. For the four existing learning-to-rank models, the pairwise RankBoost model outperforms the pointwise Regression and SVM-LTR models, and the listwise LambdaMART model achieves the best performance. One possible explanation for this finding is that the listwise model encodes more ranking information in its loss function than the pairwise and pointwise models, and ranking information is highly useful in obtaining the context information of emotions for cause extraction. The experimental results show that learning-to-rank methods are useful for emotion cause extraction. The proposed method achieves the best performance among all the baseline models. To further analyze the proposed method, we examine the importance of different ranking features in the following subsections.

We also compare the proposed method and other learning-to-rank methods using one-stage and two-stage ranking models, respectively, where one-stage ranking models refer to directly training the ranking models without using adaptive cause complexity predictors. The comparisons are shown in Fig. 3. The results show that the two-stage ranking significantly outperforms the one-stage ranking among all the ranking methods ($p < 0.01$). The comparison further demonstrates the effectiveness of adaptive cause complexity predictors in the proposed model. Compared with one-stage ranking based on learning-to-rank methods, the proposed two-stage method captures more causal information of emotions by considering the number of causal clauses in model construction. Therefore, this method produces more accurate performance for emotion cause extraction. By analyzing the classification errors, we notice that our model has difficulty in extracting the correct causal clauses for complex emotional sentences, particularly for sentences containing long-distance dependency relations. It remains challenging to properly model the complex relations of clauses in emotional sentences.

Table 6
Evaluation of feature importance. The column 'Ratio' shows the proportion of performance decline according to the F-measure compared with the model containing all the defined features.

Feature	Precision	Recall	F-measure	Ratio
All-FT	0.7625	0.7407	0.7515	−5.57%
All-WE	0.7838	0.7614	0.7725	−2.91%
All-TM	0.7601	0.7383	0.7491	−5.88%
All-POS	0.7743	0.7522	0.7631	−4.12%
All-Position	0.6746	0.6553	0.6648	−16.47%
All-Length	0.7720	0.7499	0.7608	−4.41%

Table 7
Cause position of each emotion.

Position	Number	Percentage
Previous 3 clauses	37	1.71%
Previous 2 clauses	167	7.71%
Previous 1 clauses	1180	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%

4.3.2. Evaluation of ranking features

To examine the feature effectiveness for constructing ranking models, we remove one type of feature each time, and we use the remaining features to train the proposed ranking model. The evaluation results related to the learned models are reported in Table 6. In the table,'All-FT', 'All-WE', 'All-TM', 'All-POS', 'All-Position' and 'All-Length' represent the feature sets without core term, word embedding, topic model, POS tagging, relative position, and clause length features, respectively. 'All' represents all the defined features.

The results show that removal of the relative position features leads to a sharp decrease in ranking performance. This phenomenon indicates that this type of feature is indispensable in the emotion causal clause because causal clauses are always adjacent to emotional clauses according to the distribution of cause positions of the used dataset, as shown in Table 7. The statistics show that more than 85% of emotion causes are next to the emotional clause, which helps explain why relative position features greatly impact the ranking performance.

We also observe that features based on topic models and core terms affect the ranking performance to large extents because the defined features based on topic models learn the latent topics of emotion events at different levels for modeling the correlation among clauses. Core terms encode abundant causal indicators of clauses in certain sentences for emotion provocation. The adopted core terms and topic models jointly contribute to improving the ranking performance, which we discuss in depth in the following sections. We also find that the word embedding-, POS- and length-based features lead to comparable performance for learning effective clause-ranking models. The best performance is achieved using all the defined features.

4.3.3. Evaluation of core term-based features

To evaluate the importance of different types of core terms, we add one group of core terms for model construction each time, and we compare the ranking performance of different models. Our method adopts six groups of core terms in total. We compare and report their performance in Table 8. The experimental results

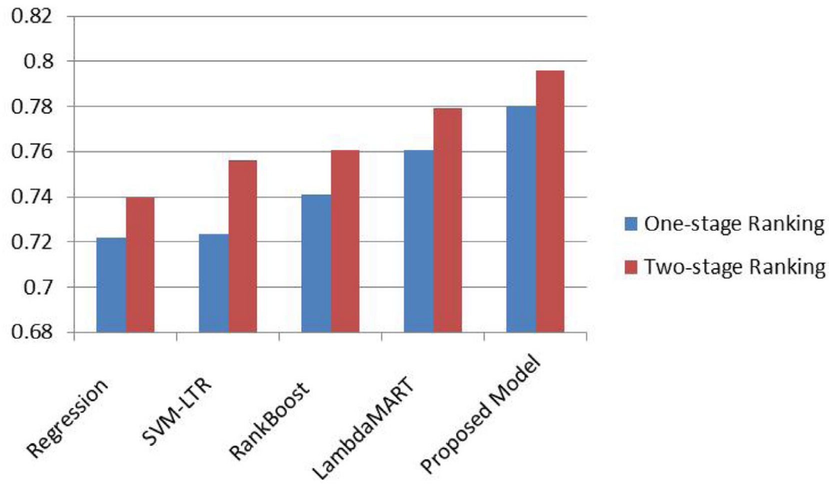

Fig. 3. Comparisons between one-stage ranking and two-stage ranking models.

Table 8
Evaluation of the influence of core terms. The column 'Ratio' shows the proportion of performance increase compared with the model with the All-FT feature set according to the F-measure.

Core term	Precision	Recall	F-measure	Ratio
Causal conj.	0.7981	0.7753	0.7865	+4.65%
Causal verbs	0.7838	0.7614	0.7725	+2.79%
Sensory verbs	0.7910	0.7683	0.7795	+3.73%
Emotion terms	0.7862	0.7637	0.7748	+3.10%
Negation	0.7920	0.7693	0.7805	+3.86%
Family-related pron.	0.7767	0.7545	0.7654	+1.85%

Table 9
Evaluation of the influence of topic models. The column 'Ratio' shows the proportion of performance increase compared with the model with the All-TM feature set according to the F-measure.

Topic model	Precision	Recall	F-measure	Ratio
LSI-clause	0.7553	0.7337	0.7444	−0.63%
LSI-emotion	0.7981	0.7753	0.7865	+4.99%
LSI-document	**0.8029**	**0.7799**	**0.7912**	**+5.62%**
LDA-clause	0.7672	0.7453	0.7561	+0.93%
LDA-emotion	0.7910	0.7683	0.7795	+4.06%
LDA-document	0.7933	0.7707	0.7818	+4.37%

Table 10
Influence of stopword removal and query-level normalization.

Normalization	Stopword	Precision	Recall	F-measure
No	No	0.7767	0.7545	0.7654
Yes	No	0.7791	0.7568	0.7678
No	Yes	0.7648	0.7430	0.7537
Yes	Yes	**0.8076**	**0.7845**	**0.7959**

show that different groups of terms jointly contribute to improving the ranking performance, and the causal conjunctions are the most effective for emotion cause extraction. Negation terms and sensory verbs also play important roles in the task. The results imply that these groups of core terms can accurately reflect the possibility that a clause provokes the target emotion.

4.3.4. Evaluation of topic model-based features

To evaluate the effectiveness of features based on topic models, we add one topic model-based feature for model construction each time, and we compare the ranking performance of different models. We adopt LSI and LDA as the topic models for extracting clause-level, emotion-level and document-level features. We compare these six features and report the performances in Table 9. The performances in the table are obtained based on the proposed ranking model with one topic model feature and all the other types of defined features.

The experimental results indicate that LDA yields comparable results to LSI. Emotion-level topic features outperform clause-level features, and document-level features achieve the best performance in emotion cause extraction. One possible explanation for this finding is that documents involving multiple emotions are more intact than emotional sentences and clauses. As a result, the learned topics based on documents are more effective in computing the similarity of clauses for learning effective ranking models.

4.3.5. Impact of stopword removal and feature normalization

Based on preliminary experiments, we observe that stopword removal and feature normalization have a great influence on the ranking performance for emotion cause extraction. We therefore performed a set of experiments to investigate the effectiveness of these two factors in the proposed method. In the experiments, we construct four ranking models based on these two factors and provide the experimental results in Table 10. In the table, 'stopword' represents stopword removal in feature extraction, and 'normalization' represents feature normalization in clause representations.

The experimental results imply that both stopword removal and feature normalization affect the ranking performance of emotion cause extraction. Feature normalization plays a more important role than stopword removal because stopwords can add noise in the similarity computation between clauses, and feature normalization, stemming from learning to rank in IR, contributes to constructing more robust and effective models. Therefore, stopword removal and feature normalization, when used jointly, can improve the performance of emotion cause extraction.

5. Conclusions

In this work, we proposed a novel two-stage ranking method for emotion cause extraction by transforming the task as a learning-to-rank problem from an information retrieval perspective. The proposed method is different from existing studies based on classification-based methods. In the two stages, the proposed method integrates query performance predictors for emotion complexity prediction in the first stage, and incorporates the emotion complexity into an autoencoder-enhanced ranking model in the second stage for emotion cause extraction. We also define and extract a large set of clause features for clause representations with respect to the provoked emotions. The proposed method is advantageous in accurately extracting the causal clauses within the emotional sentence, in which contextual information is fully considered in the model construction. Experimental results demonstrate the effectiveness of the proposed method in different aspects. Since emotion cause extraction is an incipient and challenging task in NLP, our future work will evaluate our method on other cause-related tasks, such as extracting the causes of certain events in text or emotion-cause pairs. We will also explore to extract other essential elements of emotions, such as the emotional results, for in-depth emotion analysis in our future work.

CRediT authorship contribution statement

Bo Xu: Conceptualization, Methodology, Writing - original draft. **Hongfei Lin:** Formal analysis, Funding acquisition. **Yuan Lin:** Writing - review & editing, Investigation. **Kan Xu:** Writing - review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by grant from the Natural Science Foundation of China (No. 62006034, 62076046), the Ministry of Education Humanities and Social Science, China Project (No. 19YJCZH199), the State Key Laboratory of Novel Software Technology, China (No. KFKT2021B07), and the Fundamental Research Funds for the Central Universities, China (No. DUT21RC(3)015).

References

[1] J. Li, Z. Zhang, H. He, Hierarchical convolutional neural networks for EEG-based emotion recognition, Cogn. Comput. 10 (2) (2018) 368–380.
[2] S. Feng, Y. Wang, K. Song, D. Wang, G. Yu, Detecting multiple coexisting emotions in microblogs with convolutional neural networks, Cogn. Comput. 10 (1) (2018) 136–155.
[3] M.Z. Asghar, A. Khan, A. Bibi, F.M. Kundi, H. Ahmad, Sentence-level emotion detection framework using rule-based classification, Cogn. Comput. 9 (6) (2017) 868–894.
[4] Y.-C. Chang, C.-H. Chu, C.C. Chen, W.-L. Hsu, Linguistic template extraction for recognizing reader-emotion, Int. J. Comput. Linguist. Chin. Lang. Process. 21 (1) (2016) 29–50.
[5] D. Zhou, X. Zhang, Y. Zhou, Q. Zhao, X. Geng, Emotion distribution learning from texts, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 638–647.
[6] W. Gao, S. Li, S.Y.M. Lee, G. Zhou, C.-R. Huang, Joint learning on sentiment and emotion classification, in: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, ACM, 2013, pp. 1505–1508.
[7] J. Xu, R. Xu, Q. Lu, X. Wang, Coarse-to-fine sentence-level emotion classification based on the intra-sentence features and sentential context, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM, 2012, pp. 2455–2458.
[8] S. Li, L. Huang, R. Wang, G. Zhou, Sentence-level emotion classification with label and context dependence, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 1045–1053.
[9] K. Gao, H. Xu, J. Wang, A rule-based approach to emotion cause detection for Chinese micro-blogs, Expert Syst. Appl. 42 (9) (2015) 4517–4528.
[10] L. Gui, L. Yuan, R. Xu, B. Liu, Q. Lu, Y. Zhou, Emotion cause detection with linguistic construction in Chinese Weibo text, in: Natural Language Processing and Chinese Computing, Springer, 2014, pp. 457–464.
[11] L. Gui, D. Wu, R. Xu, Q. Lu, Y. Zhou, Event-driven emotion cause extraction with corpus construction, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 1639–1649.
[12] L. Gui, J. Hu, Y. He, R. Xu, Q. Lu, J. Du, A question answering approach to emotion cause extraction, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 1593–1602.
[13] S.Y.M. Lee, Y. Chen, C.-R. Huang, A text-driven rule-based system for emotion cause detection, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Association for Computational Linguistics, 2010, pp. 45–53.
[14] Y. Chen, S.Y.M. Lee, S. Li, C.-R. Huang, Emotion cause detection with linguistic constructions, in: Proceedings of the 23rd International Conference on Computational Linguistics, Association for Computational Linguistics, 2010, pp. 179–187.
[15] D. Ghazi, D. Inkpen, S. Szpakowicz, Detecting emotion stimuli in emotion-bearing sentences, in: International Conference on Intelligent Text Processing and Computational Linguistics, Springer, 2015, pp. 152–165.
[16] I. Russo, T. Caselli, F. Rubino, E. Boldrini, P. Martínez-Barco, Emocause: an easy-adaptable approach to emotion cause contexts, in: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Association for Computational Linguistics, 2011, pp. 153–160.
[17] X. Cheng, Y. Chen, B. Cheng, S. Li, G. Zhou, An emotion cause corpus for Chinese microblogs with multiple-user structures, ACM Trans. Asian Low-Resour. Lang. Inf. Process. (TALLIP) 17 (1) (2017) 6.
[18] R. Mulkar-Mehta, C. Welty, J.R. Hoobs, E. Hovy, Using granularity concepts for discovering causal relations, in: Proceedings of the FLAIRS Conference, 2011.
[19] L. Gui, R. Xu, Q. Lu, D. Wu, Y. Zhou, Emotion cause extraction, a challenging task with corpus construction, in: Social Media Processing. SMP 2016. Communications in Computer and Information Science, 2016, pp. 98–109.
[20] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015.
[21] P. Ekman, Expression and the nature of emotion, Approaches Emotion 3 (1984) 19–344.
[22] D.D. Franks, On the origins of human emotions: A sociological inquiry into the evolution of human affect, Contemp. Sociol. 30 (5) (2001) 483.
[23] W. James, What is an emotion?, Mind 9 (34) (1884) 188–205.
[24] R. Plutchik, Emotion: A Psychoevolutionary Synthesis, Harpercollins College Division, 1980.
[25] A. Wierzbicka, Emotions Across Languages and Cultures: Diversity and Universals, Cambridge University Press, 1999.
[26] A. Neviarouskaya, M. Aono, Extracting causes of emotions from text, in: International Joint Conference on Natural Language Processing, 2013, pp. 932–936.
[27] W. Li, H. Xu, Text-based emotion classification using emotion cause extraction, Expert Syst. Appl. 41 (4) (2014) 1742–1749.
[28] K. Gao, H. Xu, J. Wang, Emotion cause detection for Chinese micro-blogs based on ECOCC model, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2015, pp. 3–14.
[29] F. Xia, T.-Y. Liu, J. Wang, W. Zhang, H. Li, Listwise approach to learning to rank: theory and algorithm, in: Proceedings of the 25th International Conference on Machine Learning, ACM, 2008, pp. 1192–1199.
[30] Z. Ding, R. Xia, J. Yu, End-to-end emotion-cause pair extraction based on sliding window multi-label learning, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020, Association for Computational Linguistics, 2020, pp. 3574–3583.
[31] C. Fan, C. Yuan, J. Du, L. Gui, M. Yang, R. Xu, Transition-based directed graph construction for emotion-cause pair extraction, in: D. Jurafsky, J. Chai, N. Schluter, J.R. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 3707–3717.
[32] P. Wei, J. Zhao, W. Mao, Effective inter-clause modeling for end-to-end emotion-cause pair extraction, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 3171–3181.
[33] Z. Ding, R. Xia, J. Yu, ECPE-2D: emotion-cause pair extraction based on joint two-dimensional representation, interaction and prediction, in: D. Jurafsky, J. Chai, N. Schluter, J.R. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, Association for Computational Linguistics, 2020, pp. 3161–3170.
[34] G. Hu, G. Lu, Y. Zhao, FSS-GCN: a graph convolutional networks with fusion of semantic and structure for emotion cause analysis, Knowl.-Based Syst. 212 (2021) 106584.
[35] X. Li, S. Feng, D. Wang, Y. Zhang, Context-aware emotion cause analysis with multi-attention-based neural network, Knowl.-Based Syst. 174 (2019) 205–218.
[36] T.-Y. Liu, et al., Learning to rank for information retrieval, Found. Trends Inf. Retr. 3 (3) (2009) 225–331.

[37] T.-Y. Liu, J. Xu, T. Qin, W. Xiong, H. Li, Letor: Benchmark dataset for research on learning to rank for information retrieval, in: Proceedings of SIGIR 2007 Workshop on Learning to Rank for Information Retrieval, Vol. 310, ACM Amsterdam, The Netherlands, 2007.

[38] Z. Cao, T. Qin, T.-Y. Liu, M.-F. Tsai, H. Li, Learning to rank: from pairwise approach to listwise approach, in: Proceedings of the 24th International Conference on Machine Learning, ACM, 2007, pp. 129–136.

[39] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang, H.-W. Hon, Adapting ranking SVM to document retrieval, in: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2006, pp. 186–193.

[40] J. Xu, T.-Y. Liu, M. Lu, H. Li, W.-Y. Ma, Directly optimizing evaluation measures in learning to rank, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2008, pp. 107–114.

[41] M. Zhang, Y. Liu, H. Luan, M. Sun, Listwise ranking functions for statistical machine translation, IEEE/ACM Trans. Audio Speech Lang. Process. (TASLP) 24 (8) (2016) 1464–1472.

[42] C.N.d. Santos, B. Xiang, B. Zhou, Classifying relations by ranking with convolutional neural networks, 2015, arXiv preprint arXiv:1504.06580.

[43] C. Shen, T. Li, Learning to rank for query-focused multi-document summarization, in: Data Mining (ICDM), 2011 IEEE 11th International Conference on, IEEE, 2011, pp. 626–634.

[44] H.Y. Lin, H.H. Chen, Ranking reader emotions using pairwise loss minimization and emotional distribution regression, in: Conference on Empirical Methods in Natural Language Processing, 2008.

[45] L. Yang, H. Lin, Y. Lin, S. Liu, Detection and extraction of hot topics on Chinese microblogs, Cogn. Comput. 8 (4) (2016) 577–586.

[46] Y. Li, L. Yang, B. Xu, J. Wang, H. Lin, Improving user attribute classification with text and social network attention, Cogn. Comput. (2019) http://dx.doi.org/10.1007/s12559-019-9624-y.

[47] S. Cronen-Townsend, Y. Zhou, W.B. Croft, Predicting query performance, in: International ACM SIGIR Conference on Research and Development in Information Retrieval, 2002, pp. 299–306.

[48] Y. Lv, C. Zhai, Adaptive relevance feedback in information retrieval, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, ACM, 2009, pp. 255–264.

[49] B. He, I. Ounis, Query performance prediction, Inf. Syst. 31 (7) (2006) 585–594.

[50] C. Hauff, D. Hiemstra, F. de Jong, A survey of pre-retrieval query performance predictors, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, ACM, 2008, pp. 1419–1420.

[51] Y. Zhao, F. Scholer, Y. Tsegay, Effective pre-retrieval query performance prediction using similarity and variability evidence, in: European Conference on Information Retrieval, Springer, 2008, pp. 52–64.

[52] E.A. Kensinger, S. Corkin, Memory enhancement for emotional words: are emotional words more vividly remembered than neutral words?, Mem. Cogn. 31 (8) (2003) 1169–1180.

[53] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2002) 601–608.

[54] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, R.A. Harshman, Indexing by latent semantic analysis, J. Amer. Soc. Inf. Sci. 41 (6) (1990) 391–407.

[55] B. Xu, H. Lin, Y. Lin, K. Xu, Learning to rank with query-level semi-supervised autoencoders, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017, 2017, pp. 2395–2398.

[56] J. Xu, R. Xu, Y. Zheng, Q. Lu, K.-F. Wong, X. Wang, Chinese emotion lexicon developing via multi-lingual lexical resources integration, in: International Conference on Intelligent Text Processing and Computational Linguistics, Springer, 2013, pp. 174–182.

[57] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, 2013, arXiv preprint arXiv:1301.3781.

[58] Y. Kim, Convolutional neural networks for sentence classification, Empir. Methods Nat. Lang. Process. (2014) 1746–1751.

[59] M. Lease, J. Allan, W.B. Croft, Regression rank: Learning to meet the opportunity of descriptive queries, in: European Conference on Information Retrieval, Springer, 2009, pp. 90–101.

[60] Y. Freund, R.D. Iyer, R.E. Schapire, Y. Singer, An efficient boosting algorithm for combining preferences, J. Mach. Learn. Res. 4 (2003) 933–969.

[61] C.J. Burges, From ranknet to lambdarank to lambdamart: An overview, Learning 11 (23–581) (2010) 81.