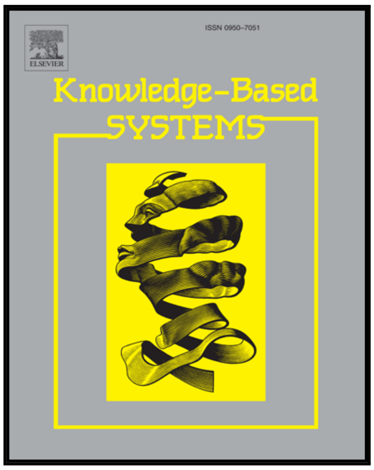
Accepted Manuscript



FREERL: Fusion Relation Embedded Representation Learning Framework for Aspect Extraction

Jian Liao, Suge Wang, Deyu Li, Xiaoli Li

PII: S0950-7051(17)30330-1
DOI: 10.1016/j.knosys.2017.07.015
Reference: KNOSYS 3979

To appear in: *Knowledge-Based Systems*

Received date: 5 February 2017
Revised date: 15 June 2017
Accepted date: 13 July 2017

Please cite this article as: Jian Liao, Suge Wang, Deyu Li, Xiaoli Li, FREERL: Fusion Relation Embedded Representation Learning Framework for Aspect Extraction, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.07.015

FREERL: Fusion Relation Embedded Representation Learning Framework for Aspect Extraction

Jian Liao[a], Suge Wang[a,b,*], Deyu Li[a,b], Xiaoli Li[c]

[a]*School of Computer & Information Technology, Shanxi University, China*
[b]*Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, China*
[c]*Institute for Infocomm Research, A*STAR, Singapore*

Abstract

Opinion object-attribute extraction is one of the fundamental tasks of fine-grained sentiment analysis. It is accomplished by identifying opinion aspect entities (including object entities and attribute entities) and then aligning object entities to attribute entities. Recent studies on knowledge graphs have shown that by adding the embeddings of semantic structures between opinion aspect entities, structure-based learning models can achieve better performance in link-prediction than traditional methods. The studies, however, focused only on learning semantic structures between aspect entities, did not take language expression features into account. In this paper, we propose the Fusion RElation Embedded Representation Learning (FREERL) framework, by which, one can fuse semantic structures and language expression features such as statistical co-occurrence or dependency syntax, into the embeddings of object entities and attribute entities. The obtained embeddings are then used to align object-attribute pairs and to predict new pairs in a zero-shot scenario. Experimental results on the datasets of COAE2014 and COAE2015 show that the best results in our framework achieve 12.1% and 32.1% improvements over the baselines, respectively.

Keywords: fusion learning, structure-based embedding, language expression feature, entity representation

*wsg@sxu.edu.cn

1. Introduction

In the past decade, opinion mining has become a hot research topic due to its potential to solve many challenging problems [11], such as affective computing[3], emotion recognition[24] and data mining in social network[32, 18]. According to the survey of [33], opinion mining can be studied at three levels of granularity, namely, the document level, the sentence level, and the aspect level. In the aspect level, opinion mining focuses on extracting aspect entities and analyzing their polarities. This task is similar to the semantic role labeling problem[20], the words or phrases in sentences need to be identified in the sense of what is the object and what is the attribute, but there are some differences between them. Although opinion aspect is generally regarded as a whole expression [25, 12], it is composed of opinion object and opinion attribute, and thus provides more detailed and hierarchical information for sentiment analysis.

For most fine-grained sentiment analysis tasks, opinion aspect can be briefly represented as a triplet <object, attribute, polarity>. This representation is more suitable for short texts in social media such as Twitter and Weibo, which are generally produced with only a few sentences and thus suffer from sparsity in the bag-of-words model. In such short texts, opinion aspects are usually nouns or noun phrases. For example, the weibo text "新出的 ipad4 电池续航简直坑爹。" (The new Apple ipad4's battery is cheating.) is represented as a triplet<ipad4, 电池续航 (battery), negative>, and the complete opinion target is "ipad4 电池续航"(ipad4's battery), where "ipad4" is the object and "电池续航"(battery) is the attribute. It is obvious that the triple can be also represented as a quadruple <object, attribute, expression, polarity>, i.e., <ipad4, 电池续航 (battery), 坑爹 (cheating), negative>, where the opinion expression is usually omitted because most opinion mining tasks are only interested in the polarities instead of the opinion expressions. From the perspective, it is more appropriate to regard opinion mining as a two-stage process, namely, extracting the pairs <object, attribute> and then analyzing the polarities of expressions. The two-stage process will produce more distinguishing and unambiguous representations of opinions, especially in the circumstance where more than one expression can be extracted from one opinion target. Another benefit of two-stage process is that it is good at dealing with implicit opinion expressions whose sentiment polarity cannot be simply derived from explicit sentiment words. These implicit opinion expressions are usually affected by their corresponding targets, which need to be pre-extracted [11]. In this paper, we mainly focus on the extraction of the object-attribute pairs.

Previous studies have shown the effectiveness of either language expression features[11] or semantic structure in opinion mining, but few of them considered both simultaneously. The traditional structure-based learning methods, such as TransE[6] modeled these relationships between entities by interpreting them as translations operating on the low-dimensional embeddings of entities. However, this type of methods model only the pair of entities and ignores the information that the background text contains. Thus, structure-based learning could suffer from under-fitting of the unknown new entities in a zero-shot scenario, where supervised learning has to be performed with not enough labeled examples available for all classes[19]. On the other hand, language expression feature based methods can fully make use of the texts, effectively cover the shortage of structure-based learning, but could fail to identify the implicit relation, especially for those entities that have a low frequency. For that reason, these two relations that complement each other can make a good combination and significantly improve the performance for many tasks.

In this paper, we divide relations between object and attribute entities into semantic structure-based relations and language expression feature based relations. We assume that there are strong relevance between opinion objects and attributes in both language expressions and semantic structures. A novel Fusion RElation Embedded Representation Learning (FREERL) framework is proposed, aiming at taking advantage of both semantic structures and language expression features.

The main contributions of this paper are summarized as follows:

1. We propose a novel fusion relation embedded representation learning framework, which embeds semantic structures and language expression features into the representations of opinion aspect entities. This framework is domain independent.
2. Compared with traditional structure-based learning models, the proposed method can be a better choice and a new baseline for the aspect extraction problem, especially in a zero-shot scenario.
3. We release the embeddings of opinion objects and their corresponding attributes with different types of language expression features embedded, trained from a gold-standard social media corpus, which can benefit other sentiment analysis tasks.

The remainder of this paper is structured as follows: Section 2 presents an introduction to the related work. Section 3 describes the proposed method. The experiments and analysis are discussed in Section 4. Finally, we conclude the paper in Section 5.

2. Related Work

Aspect Extraction

Many previous studies in aspect extraction designed some linguistic patterns and used rule-based methods to match and extract the probable aspects[22]. In the studies, the linguistic patterns have to be designed manually and selected carefully. To solve this problem, Liu et al.[14] proposed a novel method to select an effective set of rules. Appel et al.[1] presented a hybrid approach to estimate the semantic orientation polarity and its intensity by using semantic rules, a sentiment lexicon enhanced with the assistance of SentiWordNet, and fuzzy sets. The rule-based methods are difficult to expand and suffer from poor results from grammatical analysis of twitter or weibo text. Sequence models have been proven to be excellent models for learning and identifying the sequence of aspect expressions [26, 30]. The models, however, are statistically oriented and thus did not take the semantic and linguistical knowledge into account. Yang and Cardie[31] employed a semi-conditional random fields(CRFs) based segment-level sequence labeler which have rich phrase-level syntactic features, and proposed a joint model for opinion expression extraction and attribute classification. The CRFs based model, however, is linear and thus requires an elaborately designed template and a large number of features to ensure its performance. Topic models[7, 5] could only find some general/rough aspects and had difficulty in finding fine-grained or precise aspects. Poria et al.[23] integrated common-sense computing in the calculation of word distributions in the latent dirichlet allocation(LDA) algorithm, thus enabling the shift from syntax to semantics in aspect-based sentiment analysis. Wang et al.[27] proposed a novel restricted boltzmann machines(RBM) based model to simultaneously extract aspects of entities and relevant sentiment-bearing words. Poria et al.[21] used a deep convolutional neural networks(CNNs) combined with a set of linguistic patterns to tag each word in opinion sentences as either an aspect or non-aspect word.

The alignment of object and attribute is another core point of opinion mining and usually treated as an entity assignment problem. Zhao et al.[34]

imposed intra- and inter-sentence constraints and employed integer linear programming to resolve the conflicts arising during the alignment. Liu et al.[12] employed an alignment process to identify opinion relations between opinion targets and opinion words, and proposed a partially supervised alignment model based approach. They [13] also constructed a heterogeneous graph to model the relations and proposed a co-ranking algorithm to estimate the confidence of each candidate.

Structure-based Learning Models

Recently, representation learning has attracted a large amount of attention in natural language processing(NLP) research. Many neural-based representation learning methods were proposed to encode the semantics of entities and relations in low-dimensional embeddings. Among them, structure-based learning regards the structure relations r between objects O and attributes A, such as the implication or coordination, as a vector, and added the constraint $\boldsymbol{O} + \boldsymbol{r} \approx \boldsymbol{A}$ to the distributed representation of each entity in a true pair, where the characters in bold mean the vectors of each entity and relation. Bordes et al.[2] regarded the relations between entities as a translation embedding and proposed the TransE model to learn the representations of entities and relations. The basic idea of TransE model is that the distance between object and attribute entities in a true pair should be as similar as possible to the relation between the two entities. Based on Bordes's work, Wang et al.[28] added a hyperplane vector to project the entities to different spaces and to effectively increase the accuracy, and proposed the TransH model. Different from Wang's work, Lin et al.[10] added a projection matrix, instead of a hyperplane normal vector, to the TransE model for projecting entities from entity space to relation space. The new model, called TransR, performed better than the two former models. Nguyen et al.[17] proposed a new embedding model, STransE, to transfer the object and the attribute entities to different entity spaces by using different projection matrices $\boldsymbol{M}_{rO}$ and $\boldsymbol{M}_{rA}$. Lin et al.[9] categorized existing knowledge graph relations into entities' types(nationality, gender) and structure relations(parent-of, part-of), and proposed a new KR model to learn the semantic relations between the entities. The KR model, however, neglected the necessity and effectiveness of language expression, which has been proven to be a significant feature when addressing aspect extraction. Xie et al.[29] used the descriptions of entities as external knowledge for entity prediction, and trained the description and structure-based embedding for mutual promotion.

3. Methodology

3.1. Design of the Framework

We focus on the extraction of the opinion aspects in sentences S, which can be represented as a pair of <object, attribute>. Each noun or noun phrase n_{ik} in $s_i \in S$ is regarded as a candidate opinion aspect entity (objects or attributes), $i = |S|$ and k is the number of nouns or noun phrases in s_i. For example, we consider the following labeled sentence:

> 看到 [OBJ:Find5] 的 [NN: 照片] 我有点动摇了，如果 [NN: 米 2] 在 [OBJ:Find5] 发布之前就发售我按 [NP: 原计划] 买 [NN: 米 2]，如果晚了我要真考虑 [OBJ:Find5] 了，[OBJ:Find5] 那个 [ATT: 外观] 和 [ATT: 性能] 真地对 [NN: 发烧友] 有致命的 [NN: 吸引力]。
>
> (I am shocked when I see the [NN:picture] of [OBJ:Find5], I'll buy [NN:MI2] according to [NP:original plan] if it goes on sale before [OBJ:Find5]; otherwise, I'll consider [OBJ:Find5], the [ATT:exterior] and [ATT:performance] of [OBJ:Find5] have a fatal [NN:attraction] to the [NN:crazy fans].)

All the nouns (labeled as NN) and noun phrases (labeled as NP) along with the opinion aspect entities (labeled as OBJ/ATT) in the sentence are considered as candidate opinion aspect entities and are paired off. A true pair means that there is an implicit semantic relation such as "has-a" relation between the object entity and its correspond attribute, while the entities in a false pair are unrelated to each other. For example, <Find5, 外观 (exterior)> and <Find5, 性能 (performance)> are two true object-attribute pairs since *exterior* and *performance* are two general evaluation indexes of a mobile phone, while <Find5, 发烧友 (crazy fans)> and < 照片 (picture), 米 2 (MI2)> are two false ones[1].

According to the analysis of social media texts, two basic assumptions are proposed firstly, as follows:

1. There is an implicit semantic relation between an opinion object and its corresponding attribute, and the relation can be learned as a distributed representation. For example, the pair<*mobile phone*, *screen*> imply

[1]Here "Find5" and "MI2" are two popular mobile phone products in the Chinese mainland market, just like "iPhone7" or "galaxy s6".

a "*has-a*" relation between *mobile phone* and *screen*, and this implicit relation can be represented and learned as a semantic embedding.

2. Opinion objects and their corresponding attributes should share stronger relevance than others in statistics or in syntactic connections, and vice versa.

According to the basic assumptions, the FREERL framework is proposed for opinion object-attribute extraction. The framework, which is shown as Figure 1, aims at ranking the true pairs higher than the false ones in a sentence by making fully use of the information of semantic structure and language expression feature between entities.

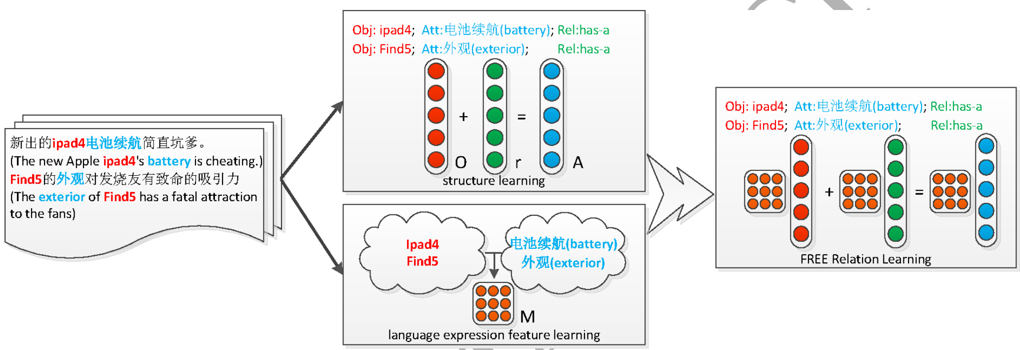


Figure 1: The FREERL Framework.

In Figure 1, from the perspective of semantic structure-based learning, for the opinion aspect entities in true pairs, we are more likely to learn a stable relation vector $\boldsymbol{r}$ that can represent the semantic structure and measure the semantic relevance between opinion object O and its corresponding attribute A, where the character in bold means the vector or matrix of its corresponding variable(similarly hereinafter). Taking the sentences in Figure 1 for examples, we can learn a stable relation "has-a" which satisfy $\boldsymbol{O}_{ipad4} + \boldsymbol{r}_{has-a} \approx \boldsymbol{A}_{battery}$ and $\boldsymbol{O}_{Find5} + \boldsymbol{r}_{has-a} \approx \boldsymbol{A}_{exterior}$. The hyponymy between entities can also be distinguished by semantic structure-based learning model automatically. For example, in the pair <ipad4, 电池续航 (battery)>, an ipad4 has a battery and they satisfy $\boldsymbol{O}_{ipad4} + \boldsymbol{r}_{has-a} \approx \boldsymbol{A}_{battery}$, but there is a large difference between $\boldsymbol{A}_{battery} + \boldsymbol{r}_{has-a}$ and $\boldsymbol{O}_{ipad4}$. In terms of language expression feature, the relevance of objects and attributes in true pairs, which are learned as the matrix $\boldsymbol{M}$, are much higher than those in false pairs according to the second assumption. For example, the *Jaccard* values of true pairs<ipad4, 电池续航 (battery)> and <Find5, 外观 (exterior)> are 0.614 and 0.377, higher than false pairs like <ipad4, 三星 (Sumsang)> and <Find5, 发烧友 (crazy fans)> whose *Jaccard* values are 0.235 and 0.167, respectively. Then, the language expression feature matrix $\boldsymbol{M}$ is fused with the structure learning score function as weight and embedded in the representation of opinion object $\boldsymbol{O}$, attribute $\boldsymbol{A}$, and implicit relation $\boldsymbol{r}$ by the framework.

3.2. Fusion Relation Embedded Representation Learning

The proposed framework aims at ranking the true pairs higher than the false ones in a sentence. Thus, the objective function is defined by Eq. (1):

$$P(A|O,\theta) = \frac{score(O,A,\theta)}{\sum_{\hat{A}\in s_i} score(O,\hat{A},\theta)} \tag{1}$$

where O and A represent the object and attribute words or phrases in a true object-attribute pair, respectively, $\hat{A}$ represents all the candidate attribute words or phrases in sentence $s_i \in S$.

Semantic structures and language expression features are then fused into the representations of the object and its corresponding attribute in a sentence by the score function

$$score(O,A,\theta) = s_l(O,A)s_s(\boldsymbol{O},\boldsymbol{A},\theta) \tag{2}$$

where s_l and s_s represent the relevancies of language expression and semantic structure respectively, and the characters in bold represent the vectors or matrices. The functions s_l and s_s are defined by

$$s_l(O,A) = \boldsymbol{X}_O^T \boldsymbol{M}_l \boldsymbol{X}_A \tag{3}$$

$$s_s(\boldsymbol{O},\boldsymbol{A},\theta) = e^{g(\boldsymbol{O},\boldsymbol{A},\theta)} \tag{4}$$

where

- $\boldsymbol{M}_l \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{E}|}$ is the language expression feature matrix which can be learned to model the language expression relevance of object and attribute entities, $\mathcal{E}$ is the set of the candidate aspect entities. Each element in $\boldsymbol{M}_l$ represents the feature value of an entity pair. In this paper, we categorized the features into two classes, namely statistical correlation(word co-occurrence for instance) and syntactic dependency relation. The definition and learning process of $\boldsymbol{M}_l$ are detailed in Section 3.4;

- $\boldsymbol{X}_O \in \mathbb{R}^{|\mathcal{E}|\times 1}$ and $\boldsymbol{X}_A \in \mathbb{R}^{|\mathcal{E}|\times 1}$ are the one-hot represented index vectors of entities O and A. For example, $\boldsymbol{X}_O$ is a $|\mathcal{E}|$ dimensional vector in which the element at the index of entity O is 1 and the rest elements are 0, like $[0,0,0,...,1,0,0]_{|\mathcal{E}|\times 1}$. $\boldsymbol{X}_O^T \boldsymbol{M}_l \boldsymbol{X}_A$ is used to locate the index of the element in matrix $\boldsymbol{M}_l$ and get the corresponding feature value of pair $< O, A >$.

- $\boldsymbol{O}$, $\boldsymbol{A}$ are the distributed representations of aspect entities O and A, which are obtained by $\boldsymbol{O} = \boldsymbol{M}_e \boldsymbol{X}_O$ and $\boldsymbol{A} = \boldsymbol{M}_e \boldsymbol{X}_A$, $O, A \in \mathcal{E}$, respectively, $\boldsymbol{M}_e \in \mathbb{R}^{n\times|\mathcal{E}|}$ is a shared entity embedding matrix to project the one-hot representation of aspect entities in set $\mathcal{E}$ into a n-dimensional continuous space;

- $g(\boldsymbol{O}, \boldsymbol{A}, \theta)$ measures the relevance of semantic structure between entities. We introduce 4 popular structure-based learning models into our framework, TransE, TransH, TransR and STransE. The models are defined as the following Eq*s*. (5)-(8).

$$g_{TransE}(\boldsymbol{O}, \boldsymbol{A}, \theta) = -\|\boldsymbol{O} + \boldsymbol{r} - \boldsymbol{A}\|_{L2} + b \tag{5}$$

where b is the bias and $\theta = \{\boldsymbol{r}, b\}$.

$$g_{TransH}(\boldsymbol{O}, \boldsymbol{A}, \theta) = -\|\boldsymbol{O} - \boldsymbol{w}_r^T \boldsymbol{O} \boldsymbol{w}_r + \boldsymbol{r} - (\boldsymbol{A} - \boldsymbol{w}_r^T \boldsymbol{A} \boldsymbol{w}_r)\|_{L2} + b \tag{6}$$

where $\boldsymbol{w_r}$ is the semantic hyperplane's normal vector of r and $\theta = \{\boldsymbol{w}_r, \boldsymbol{r}, b\}$.

$$g_{TransR}(\boldsymbol{O}, \boldsymbol{A}, \theta) = -\|\boldsymbol{M}_r \boldsymbol{O} + \boldsymbol{r} - \boldsymbol{M}_r \boldsymbol{A}\|_{L2} + b \tag{7}$$

where $\boldsymbol{M}_r$ is the translating matrix of relation r and $\theta = \{\boldsymbol{M}_r, \boldsymbol{r}, b\}$.

$$g_{STransE}(\boldsymbol{O}, \boldsymbol{A}, \theta) = -\|\boldsymbol{M}_{rO} \boldsymbol{O} + \boldsymbol{r} - \boldsymbol{M}_{rA} \boldsymbol{A}\|_{L2} + b \tag{8}$$

where $\boldsymbol{M}_{rO}$ and $\boldsymbol{M}_{rA}$ are the semantic project matrices of the object and attribute entities, respectively, and $\theta = \{\boldsymbol{M}_{rO}, \boldsymbol{M}_{rA}, \boldsymbol{r}, b\}$.

Also note that the exponential function in Eq. (4) is used to map the range of $g(\cdot)$ into $(0, 1]$.

3.3. Optimization and Implementation Details

During the learning process, it is time consuming to directly compute the exponential function in Eq. (1) when the numbers of opinion objects and attributes are very large in a social media dataset. Thus, a negative sampling-based method[16] is introduced to approximate the objective function for optimization. The new approximate objective function aims to maximize the score of every true object-attribute pair while minimizing the scores of all false ones, and is defined by

$$\begin{aligned}\mathcal{L}(A|O,\theta) \propto \sum_{i} \sum_{(O,A)\in S^{i}_{(O,A)}} &\{\log[1-\sigma(s_l(O,A)g(\boldsymbol{O},\boldsymbol{A},\theta))]+ \\ &\sum_{\tilde{A}\in S^{i}_{(O,\tilde{A})}} \log \sigma[s_l(O,\tilde{A})g(\boldsymbol{O},\tilde{\boldsymbol{A}},\theta)]\}\end{aligned} \tag{9}$$

where $\sigma(x) = 1/(1+\exp(-x))$ is the sigmoid function, $S^{i}_{(O,A)}$ is the set of true opinion objects-attribute pairs in sentence s_i, and $S^{i}_{(O,\tilde{A})}$ is the set of false pairs in s_i. If s_i contains more than one object-attribute pair, then calculate the loss for each pair. For each single true pair(O_j, A_j) in s_i, we denote the loss as

$$\mathcal{L}_j = \log[1-\sigma(s_l(O_j,A_j)g(\boldsymbol{O}_j,\boldsymbol{A}_j,\theta))]+ \sum_{\tilde{A}_j\in S^{i}_{(O,\tilde{A}_j)}} \log \sigma[s_l(O_j,\tilde{A}_j)g(\boldsymbol{O}_j,\tilde{\boldsymbol{A}}_j,\theta)] \tag{10}$$

The objective function can be minimized by a gradient descent algorithm instead of stochastic gradient descent, because the former can achieve a globally optimal solution and the size of experiment data is appropriate. For brevity and legibility, we denote $s_l(O_j, A_j)$ and $s_l(O_j, \tilde{A}_j)$ as s_l and $\tilde{s}_l$, and denote $g(\boldsymbol{O}_j, \boldsymbol{A}_j, \theta)$ and $g(\boldsymbol{O}_j, \tilde{\boldsymbol{A}}_j, \theta)$ as g and $\tilde{g}$. The gradients during iterations for each variable and parameter are given by

$$\frac{\partial \mathcal{L}_j}{\partial O_j} = -s_l\sigma(s_l g)\frac{\partial g}{\partial O_j} + \sum_{\tilde{A}\in S_{i,\tilde{A}_j}} (1-\tilde{s}_l\sigma(\tilde{s}_l\tilde{g}))\frac{\partial \tilde{g}}{\partial O_j} \tag{11}$$

$$\frac{\partial \mathcal{L}_j}{\partial A_j} = -s_l\sigma(s_l g)\frac{\partial g}{\partial A_j} \tag{12}$$

$$\frac{\partial \mathcal{L}_j}{\partial \tilde{A}_j} = (1-\tilde{s}_l\sigma(\tilde{s}_l\tilde{g}))\frac{\partial \tilde{g}}{\partial \tilde{A}_j}, \tilde{A}_j \in S_{i,\tilde{A}_j} \tag{13}$$

$$\frac{\partial \mathcal{L}_j}{\partial \theta} = -s_l \sigma(s_l g) \frac{\partial g}{\partial \theta} + \sum_{\tilde{A} \in S_{i,\tilde{A}_j}} (1 - \tilde{s}_l \sigma(\tilde{s}_l \tilde{g})) \frac{\partial \tilde{g}}{\partial \theta} \tag{14}$$

Traditional gradient descent algorithm will terminate the iteration when the difference between the current loss and the last one reaches a small threshold. In this case, the current iteration may not obtain the optimal solution. An improvement can be made by adding a delay process to record and update the best result during each iteration. When the calculation of the optimization function is near the globally optimal solution, we record the current model as a temporary best solution and continue the next iteration. If the training algorithm can achieve a better loss value than the recorded model within the max-waiting iterations, we update the temporary best solution as the new optimal one and reset the iteration wait counter. Otherwise, the current recorded model can be regarded as the globally optimal solution. The main process of training is detailed in Algorithm 1, where $MinLoss$ is the current minimal loss calculated by the approximation function (Eq. (9)) and $WaitLoop$ is the maximal loop time after the optimization function reaches a new optimum solution.

3.4. Language Expression Features Learning

Language expression features play a significant role in a large amount of NLP studies because they are capable of effectively modeling such common linguistic phenomena as word co-occurrences and rhetoric. The language expression features can be categorized into two classes, namely statistical correlation and syntactic dependency relation. For example, a word co-occurrence can be considered as the statistical correlation of entities at the word level; the syntactic structure information between opinion objects and attributes can be considered as syntactic dependent relation. In this paper, three different statistical and one syntactic dependency based measurements are designed to learn the feature matrix $\boldsymbol{M}_l$. Moreover, we combined the syntactic dependency based feature with each statistical based feature as new language expression features. Each element in $\boldsymbol{M}_l$ represents the feature value of an entity pair, and $\boldsymbol{M}_l^{ij}$ is the corresponding element of entity pair$< i, j >$ in matrix $\boldsymbol{M}_l$.

$\boldsymbol{\chi^2}$ *Measurement.* The χ^2 statistic assumes that two random variables follow a chi-square distribution with one degree of freedom, and is used to measure

Algorithm 1 Algorithm of FREERL

Input: Set of labeled sentences S_{train}, pre-trained entity representation matrix $\boldsymbol{M}_e$, pre-trained language expression feature matrix $\boldsymbol{M}_l$.
Output: Fusion relation embedded entity representation matrix $\boldsymbol{M}'_e$, relation embedding $\boldsymbol{r}$, parameter set of structure-based learning θ.

```
Initial: MinLoss = +∞, MaxLoop = m, MaxWaitLoop = n
for i = 0 to MaxLoop do
  if WaitLoop > MaxWaitLoop then
    End.
  end if
  Set loss_sum = 0
  for all sentences s ∈ S_train do
    Calculate loss_s by Eq. (9), loss_sum+ = loss_s
    Update the embedding of each entity in M_e and the parameters θ
      by the gradient
  end for
  if loss_sum < MinLoss then
    Save M_e as M'_e and θ. Set MinLoss = loss_sum, WaitLoop = 0
  else
    WaitLoop + +
  end if
end for
```

the correlation between them. Consider Eq. (15)

$$\boldsymbol{M}_l^{ij} = N\frac{(P(i,j) - P(\tilde{i},\tilde{j}))^2}{P(i)(1-P(i))P(j)(1-P(j))}, i, j, \tilde{i}, \tilde{j} \in \mathcal{E} \tag{15}$$

where the characters having a tilde mean that the corresponding entities are absent in a sentence and N represents the number of sentences in the corpus.

Jaccard Coefficient. The Jaccard coefficient is used to compare the similarity and diversity of the sample sets and defined by

$$\boldsymbol{M}_l^{ij} = \frac{F(i,j)}{F(i) + F(j) - F(i,j)}, i, j \in \mathcal{E} \tag{16}$$

where $F(\cdot)$ means the number of sentences that contain the entities.

Pointwise Mutual Information. The pointwise mutual information (PMI) is a measure of the mutual dependence between two random variables and defined by

$$\boldsymbol{M}_l^{ij} = \log \frac{P(i,j)}{P(i)P(j)}; i, j \in \mathcal{E} \tag{17}$$

where $P(\cdot)$ is the co-occurrence probability or frequency of the entities at the sentence level.

Syntactic Dependency. Given a sentence s_i, its syntactic dependent structure can be analyzed by using the Language Technology Platform Cloud(LTP)[2] [4]. Figure 2 illustrates the result of a sentence processed by LTP.

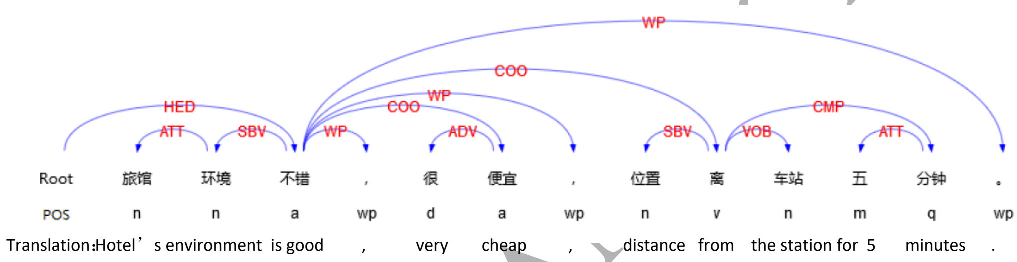


Figure 2: An example of syntactic dependency structure.

In Figure 2, an edge "环境 (enviroment) $\xrightarrow{ATT}$ 旅馆 (hotel)" refers to the latter " 旅馆" (hotel) governs the former " 环境"(enviroment) with the syntactic dependency type ATT, which means an attribute-subject structure[3].

To quantitatively analyze the strength of syntactic dependency between entities, we use the probability of each dependent type to measure the syntactic dependent strength and define

$$\boldsymbol{M}_l^{ij} = \begin{cases} \max\limits_{d_t} P(d_t|(i,j)) + 1, (i,j) \in S_{(O,A)} \\ 1, otherwise \end{cases} \tag{18}$$

where $S_{(O,A)}$ is the set of true opinion objects-attribute pairs in the corpus

[2]http://www.ltp-cloud.com/

[3]Note that the '*attribute*' here is a concept in syntactics, instead of a opinion aspect entity.

and $P(d_t|(i,j))$ is defined by

$$P(d_t|(i,j)) = \frac{\sum_{(i,j)\in S_{(O,A)}} \#_{d_t,(i,j)}}{\sum_{(i,j)\in S_{(O,A)}} \#_{(i,j)}} \tag{19}$$

where $\#_{d_t,(i,j)}$ counts the occurrence of true pair (i,j) with the dependent type d_t and $\#_{(i,j)}$ represents the frequency of true pair (i,j). Note that the added 1 in Eq. (18) is used to ensure that the values of incorrect pairs do not affect the fusion result in Eq. (2).

The basic idea of syntactic dependent strength is that, if there is a direct link from the object entity to its corresponding attribute entity in the syntactic dependency tree, and then the dependent type can be regarded as the syntactic dependent relation between the entities. Based on the assumption that the more frequent the entities co-occurred with a dependent type d_t, the stronger the relevance between them, the probability of this type is considered as the syntactic dependent strength.

Feature Combinations. The statistical based and syntactic dependency based features can also be combined as language expression features by using

$$\boldsymbol{M}_l^{ij} = \boldsymbol{M}_s^{ij} + \boldsymbol{M}_{DP}^{ij} - 1 \tag{20}$$

where $\boldsymbol{M}_s$ is the statistics-based feature matrix calculated by Eq*s*. (15)-(17), respectively, and $\boldsymbol{M}_{DP}$ is the syntactic dependency based feature matrix calculated by Eq. (18).

4. Experiment

4.1. Data Set

We adopted the datasets of COAE2014 (Chinese Opinion Analysis Evaluation) and COAE2015 for the fine-grained sentiment analysis task in the experiments [8][4]. The datasets were collected from Weibo, one of the most popular social media in China, and segmented at the sentence level. The COAE2014 dataset contains 3 domains, namely, the mobile phone, insurance, and jadeite, and the COAE2015 dataset covers a wide range of domains, such as news, sports, finance, and food. We reorganized the labeled

[4]The datasets are available at http://115.24.12.5

opinion objects and attributes into pairs <*object*, *attribute*>. We focus on the identification and alignment of the opinion aspects, and thus, we use the subset of the labeled dataset for the experiment by filtering the pairs whose object or attribute are the default, and obtained 4123 labeled pairs for the COAE2014 dataset and 10331 labeled pairs for the COAE2015 dataset.

4.2. Experimental Tasks

We designed two tasks to evaluate the proposed framework.

Opinion object and attribute alignment. This task aims to identify the correct corresponding attributes on the condition that the opinion objects are already given in a sentence. The COAE2015 dataset is used for training, and the COAE2014 dataset is for testing. For each sentence in COAE2014, all the nouns or noun phrases are regarded as the candidate attributes to the given object. We rank all the candidate attributes by the scores in descending order after calculating by Eq. (2).

New object-attribute pair prediction in a zero-shot scenario. The new pair prediction task can be regarded as a zero-shot problem because either the object or the attribute entities could be unknown in the training set. The traditional structure-based learning model cannot solve the problem effectively. The COAE2014 dataset is used for training and the COAE2015 dataset is used for testing, because the domains of the former are covered by the latter. All the nouns or noun phrases are combined into the candidate pairs for prediction. Other settings of the experiment are the same as the alignment task in Section 4.2.

4.3. Baselines, Evaluation, Pre-Training and Parameter Configuration

Four basic structure-based learning models, TransE, TransH, TransR, and STransE are used as our baselines. Moreover, to verify the effectiveness of our framework, we trained all the original structure-based learning models without fusing the language expression features during the training process; this can be done by setting $s_l(O, A) = s_l(O, \tilde{A}) = 1$ in Eq. (9). We then combine original structure-based learning models with the language expression features for comparison.

The *precision@N* is adopted as the evaluation index for each task, which is defined by

$$precision@N = \frac{\sum_i^{|S|} correct@N}{|S|} \tag{21}$$

$$correct@N = \begin{cases} 1, If\ correct\ answer\ ranks\ within\ top\ N \\ 0, otherwise \end{cases} \tag{22}$$

where $|S|$ is the number of sentences.

In most studies on knowledge graphs [2], they set N to be 10% of the number of entities. In our experiment, since after filtering the function words, the average length of the sentences in the experiment dataset is 16.73, we set $N = 1$. Word2Vec [15] is adopted to pre-train all the words in the corpus to obtain the matrix $\boldsymbol{M}_e$. Compared with randomly generated word representations, pre-training can embed some hyponymy information into the word embeddings. The shared embedding matrix $\boldsymbol{M}_e$ was pre-learned from a large social media dataset that contains 10 million weibos for the new word identification subtask in COAE2014. The core model of word2Vec is skip-gram, and the dimensions of the embeddings are set to $\{50, 100, 150\}$. All the different language expression matrices $\boldsymbol{M}_l$ were pre-learned from the experiment corpus. We set the maximum loop times $MaxLoop = 2000$ and the maximum wait times after reaching a new optimum of $MaxWaitLoop = 100$. The bias b in structure-based learning model is set to 0. We repeated our experiments for 5 times over different random initializations and used the average performances with standard deviations as the results.

5. Results and analysis

5.1. Performance of the Opinion Object and Attribute Alignment Task

The evaluation of the opinion alignment task is shown in Table 1, where the numbers between parenthesis are the standard deviations and SLM is a short for the original structure-based learning model such as TransE.

Compared with original structure-based learning model in the first row, all the FREERL combinations except TransE+χ^2 and TransE+PMI(in the third part of Table 1, rows 9–15) have effectively improved the performances. For the statistics-based language expression features, the best combinations of FREERL for each structure-based learning model are TransE+Jaccard, TransH+Jaccard, TransR+Jaccard and STransE+χ^2, achieving 0.2%, 4.4%, 7.0%, and 7.5% improvements respectively. The FREERL+Jaccard and FREERL+χ^2, on average, increase the precision with 4.6% and 4.0% compared with the basic structure-based learning models, which perform better than the FREERL+PMI by only 2.1%. We checked the PMI-based matrix $\boldsymbol{M}_l^{PMI}$ and χ^2-based matrix $\boldsymbol{M}_l^{\chi^2}$, and found that the feature values of some

Table 1: Results of opinion object and attribute alignment.

models	TransE	TransH	TransR	STransE
SLM	0.286(0.021)	0.210(0.014)	0.197(0.027)	0.196(0.024)
SLM+χ^2	0.285(0.021)	0.209(0.013)	0.196(0.027)	0.196(0.023)
SLM+Jaccard	0.293(0.012)	0.235(0.010)	0.223(0.023)	0.220(0.019)
SLM+PMI	0.256(0.003)	0.233(0.004)	0.228(0.010)	0.227(0.008)
SLM+DP	0.387(0.010)	0.332(0.013)	0.288(0.037)	0.306(0.024)
SLM+DP+χ^2	0.396(0.013)	0.329(0.013)	0.317(0.026)	0.318(0.020)
SLM+DP+Jaccard	0.406(0.006)	0.360(0.010)	0.337(0.026)	0.343(0.019)
SLM+DP+PMI	0.373(0.002)	0.354(0.004)	0.350(0.009)	0.348(0.007)
FREERL+χ^2	0.276(0.003)	0.236(0.013)	0.264(0.016)	0.271(0.048)
FREERL+Jaccard	0.288(0.001)	0.254(0.008)	0.267(0.020)	0.262(0.016)
FREERL+PMI	0.248(0.001)	0.240(0.006)	0.242(0.006)	0.243(0.008)
FREERL+DP	0.398(0.001)	0.348(0.012)	0.313(0.021)	0.328(0.023)
FREERL+DP+χ^2	**0.407(0.002)**	0.346(0.013)	0.357(0.019)	0.362(0.029)
FREERL+DP+Jaccard	0.400(0.001)	**0.385(0.011)**	**0.362(0.019)**	**0.366(0.015)**
FREERL+DP+PMI	0.369(0.001)	0.360(0.002)	0.358(0.006)	0.351(0.005)

pairs are much smaller than the values in $\boldsymbol{M}_l^{Jaccard}$. For example, the true object-attribute pair < 扬声器 (loudspeaker), 效果 (performance)> is 0.005 in $\boldsymbol{M}_l^{PMI}$ and 0.250 in $\boldsymbol{M}_l^{\chi^2}$, whereas the value in $\boldsymbol{M}_l^{Jaccard}$ is 0.563. Ranking mistakes are more likely to occur when the fusion scores of the candidate pairs are similar, especially when the differences in their semantic structures are also small.

Syntactic dependency based features (DP for short in Table 1) can remarkably increase the performances (12.5% improvement on average). It verifies that syntactic dependency features can provide abundant information by explicitly capturing the linguistic structure. We analyse the types of syntactic dependency features, and list the proportion of each dependent type in Table 2.

We can see that there are four direct syntactic dependent relations in our experiment corpus, *ATT* (attribute), *SBV* (subject-verb), *COO* (coordinate) and *VOB* (verb-object). The *ATT* relation is overwhelming among the dependent types and can be regarded as the most significant feature. We tested the performance by only using the syntactic dependency feature and

Table 2: Proportion of each dependant type.

dependant type	dependent strength
ATT	0.950
SBV	0.039
COO	0.009
VOB	0.002

the *precision*@1 is 0.198, much lower than the fusion models of FREERL+DP. This demonstrates that:1)the direct syntactic dependent relations, especially the ATT relation, are effective for their high precision; 2)most object entities and their corresponding attributes are linked indirectly and need a more elaborate method to measure the dependent strength between them.

Considering the language expression features, the best combination of FREERL are TransE+DP+χ^2, TransH+DP+Jaccard, TransR+DP+Jaccard and STransE+DP+Jaccard achieving 12.1%, 17.5%, 16.5%, and 17.0% improvements respectively. The FREERL+DP+Jaccard and FREERL+DP+χ^2, on the whole, increase the precision with 15.6% and 14.6%, showing better performance than the FREERL+DP+PMI, which improves by only 13.7%. In consideration of structure-based learning models, the TransE based FREERL combined with DP+χ^2 makes the best performance than the others. Because the semantic entities and the gap between entity domains are comparatively clear in our experiment data, few are ambiguous. At the same time, in the testing set, the types of most of the semantic structure relations between the objects and attributes are one-to-one, and many-to-many relations occur much less than in the training set. The more complex models such as TransH, TransR and STransE, which excel in addressing many-to-many relations, cannot contribute maximum advantage for the bias between the datasets. This circumstance leads to a reverse phenomenon when we exchange the training and testing set in another experiment, as discussed in Section 5.2.

Compared fusion models with the original models' combination, as shown in the second part of Table 1 (rows 2-8), we can see that most of the FREERL combinations achieved significant improvements. Some of the TransE based fusion models perform slightly lower than original structure-based learning models, probably because that the TransE based models suffer from modeling many-to-many relations, and fail to fully make use of the language expression features. However, the fusion models get more stable results with lower standard deviations. TransE+DP+χ^2, TransH+χ^2, TransR+χ^2 and STransE+χ^2 in FREERL framework outperform their corresponding combinations based on original SLMs with the improvements of 1.1%, 2.7%, 6.8%, and 7.5%. The TransH, TransR and STransE based FREERL combinations, which get the average improvements of 1.7%, 3.2%, and 3.2%, respectively, show that the fusion relation embedded learning is effective.

5.2. Performance of the New Object-Attribute Pair Prediction Task in a Zero-Shot Scenario

The performance of the new object-attribute pair prediction task in a zero-shot scenario is detailed in Table 3. Similar to the task of alignment task in Section 5.1, we use the average performances with standard deviations(between parenthesis) as the results, and SLM is a short for the original structure-based learning model such as TransE.

Table 3: Results of new object-attribute pair prediction in a zero-shot scenario.

models	**TransE**	**TransH**	**TransR**	**STransE**
SLM	0.076(0.001)	0.082(0.009)	0.069(0.012)	0.087(0.023)
SLM+χ^2	0.083(0.001)	0.095(0.010)	0.072(0.013)	0.080(0.015)
SLM+Jaccard	0.126(0.001)	0.145(0.009)	0.112(0.011)	0.125(0.013)
SLM+PMI	0.184(0.001)	0.186(0.004)	0.167(0.024)	0.181(0.007)
SLM+DP	0.257(0.003)	0.338(0.022)	0.183(0.036)	0.253(0.058)
SLM+DP+χ^2	0.365(0.001)	0.378(0.013)	0.311(0.047)	0.340(0.019)
SLM+DP+Jaccard	0.335(0.001)	0.397(0.014)	0.265(0.042)	0.317(0.027)
SLM+DP+PMI	0.359(0.001)	0.360(0.003)	0.349(0.009)	0.352(0.005)
FREERL+χ^2	0.055(0.001)	0.119(0.014)	0.075(0.016)	0.130(0.032)
FREERL+Jaccard	0.087(0.008)	0.172(0.009)	0.129(0.022)	0.163(0.018)
FREERL+PMI	0.154(0.013)	0.191(0.002)	0.181(0.010)	0.188(0.010)
FREERL+DP	0.285(0.004)	0.346(0.015)	0.278(0.050)	0.300(0.044)
FREERL+DP+χ^2	0.327(0.0004)	0.384(0.004)	0.333(0.015)	**0.377(0.016)**
FREERL+DP+Jaccard	0.288(0.003)	**0.403(0.005)**	0.308(0.024)	0.364(0.019)
FREERL+DP+PMI	0.352(0.001)	0.357(0.001)	**0.352(0.003)**	0.352(0.003)

From Table 3, we can see that some of the prediction results, especially the statistical feature based FREERL models are not as good as in the alignment task. The reasons are as follows:

1. In a zero-shot scenario, there are some unknown entities that suffer from the shortness of the training on the semantic structures in the testing dataset. For these entities, language expression features play a significant role for identification;
2. There are no given object entities to help the fusion framework filter the incorrect pairs. All the pairwise combinations of entities are regarded as candidate pairs, and it is difficult to filter some of the object-object or attribute-attribute pairs by the fusion score, especially the entities that are in the intersection of the object and attribute set. For example, in the sentence "<Doc17088> 用 [NN: 小米] 半年不得不吐槽一下，这款十分注重配置的 [OBJ: 手机]，...，时常无法读取内存卡，更不用说 [ATT: 重启]，[ATT: 发热]...(I have to complain about my [NN:XIAOMI], which used half a year, this very configuration focused [OBJ:mobile phone], ... , often failed to read the memory card, letting alone the [ATT: restart] and [ATT: fever].)". The entity " 手机 (mobile phone)" is an object in the pair < 手机 (mobile phone), 发热 (fever)> and is an attribute in the pair < 小米 (XIAOMI), 手机 (mobile phone)>[5]. Both the pairs not only have strong relevance to the semantic structure but also achieve high scores on language expression feature learning. The fusion framework prefers the latter one because the latter has the highest rank;
3. The way that we use all of the nouns and noun phrases in a sentence as candidates is coarse, it is hardly to identify non-noun aspects correctly.

For all the FREERL combinations except TransE+χ^2, we obtain a significant improvement compared with each original SLM. Considering the statistics-based language expression features, the PMI achieves the best performances combined with TransE, TransH, TransR and STransE in FREERL framework, with the improvement of 7.8%, 10.9%, 11.2%, and 10.1%, respectively. Compared with the alignment task, the PMI and Jaccard, which increase the precision by 10.0% and 5.9% on average, can remarkably benefit all of the fusion models and achieve better performances than χ^2, which

[5] 小米 (XIAOMI) is a famous brand of mobile phone in the Chinese mainland market

improves only by 1.6%. The reason is that many top-ranked pairs calculated by χ^2 are incorrect pairs, such as attribute-attribute pairs. These pairs, especially some rare ones, cannot be filtered without the object knowledge base, and the score of the language expression feature learning will to some extent cover the semantic structure part.

The syntactic dependency based feature gets a remarkable improvement among all the FREERL combinations, attaining a 22.4% improvement on average. It is obvious that the syntactic dependency based features will outperform the statistics-based ones, because the former contains hyponymy information which is insufficient for semantic structure-based learning model in a zero-shot scenario. The features will automatically filter some attribute-attribute or object-object pairs that have high statistical correlation. In terms of the language expression features, the best combination of FREERL are TransE+DP+PMI, TransH+DP+Jaccard, TransR+DP+PMI and STransE+DP+χ^2, achieving 27.6%, 32.1%, 28.3%, and 29.0% improvements, respectively. The FREERL+DP+χ^2 and FREERL+DP+PMI, on the whole, increase the precision with 27.7% and 27.5%, perform better than the FREERL+DP+Jaccard, which improves by 26.2%. In consideration of each structure-based learning model in FREERL framework, the TransH, STransE, and TransR, which make 20.0%, 18.1% and 16.8% improvements on average compared with the baseline, are more reliable and robust than TransE. As mentioned in the last paragraph in Section 5.1, the characteristic of the experiment data can make the multi-relational learning models (TransH, STransE, TransR) fully effective. These types of models can handle more complex relations between entities and perform better than the TransE based fusion model.

Compared FREERL based models with original SLMs' combinations, we can see that most of the fusion models achieved significant improvements. The TransH, TransR and STransE based combinations of FREERL gained the average improvements of 1.0%, 2.8% and 3.2%, respectively, and show the effectiveness of our fusion relation embedded learning. TransE+DP, TransH+Jaccard, TransR+DP and STransE+χ^2 in FREERL framework outperform their corresponding original SLMs' combinations with the improvements of 2.8%, 2.7%, 9.5%, and 5.0%. This demonstrates that the multi-relational learning models can more efficiently capture and embed the fused relation of semantic structures and language expression features.

6. Conclusions and Future Work

In this paper, we embedded both semantic structures and language expression features in the representation of opinion object and its corresponding attribute and proposed the FREERL framework. This framework is domain-independent and no manually crafted linguistic rules are needed. The framework can freely fuse any type of language expression features like statistical co-occurrence or dependency syntax into the structure-based embedding of each entity. We designed 7 different types of language expression feature learning models, combined with 4 popular structure-based learning methods into the fusion framework. The trained embeddings of aspect entities and relations were adopted in the task of opinion object and attribute alignment and new pair prediction in a zero-shot scenario. Experiments on COAE2014 and COAE2015 show that the best results in our framework achieve 12.1% and 32.1% improvements over the baselines, respectively.

In the future, the proposed framework will be improved in the following ways:

1. We will employ sentiment information in our framework for fine-grained sentiment analysis.
2. More types of language expression features will be introduced and embedded in the representations of entities.
3. Hierarchical structure will be added to our structure-based learning approach to improve the identification and prediction of aspect entity pairs.

Acknowledgements

The authors would like to thank all anonymous reviewers and Dr.Yanhui Zhai for their valuable comments and suggestions which have significantly improved the quality and presentation of this paper. The works described in this paper are supported by the National Natural Science Foundation of China (NSFC Nos.61632011, 61573231, 61672331, 61432011, 61603229, U1435212), Shanxi Science and Technology Infrastructure (2015091001-0102).

Reference

[1] Appel, O., Chiclana, F., Carter, J., Fujita, H., 2016. A hybrid approach to the sentiment analysis problem at the sentence level. Knowledge-Based Systems 108, 110 – 124.

[2] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O., 2013. Translating embeddings for modeling multi-relational data. In: Advances in Neural Information Processing Systems. pp. 2787–2795.

[3] Cambria, E., 2016. Affective computing and sentiment analysis. IEEE Intelligent Systems 31 (2), 102–107.

[4] Che, W., Li, Z., Liu, T., 2010. Ltp: A chinese language technology platform. In: Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations. Association for Computational Linguistics, pp. 13–16.

[5] Chen, Z., Mukherjee, A., Liu, B., 2014. Aspect extraction with automated prior knowledge learning. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics. Vol. 1. ACL, pp. 347–358.

[6] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., Zhang, W., 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 601–610.

[7] Hu, Y., Boyd-Graber, J., Satinoff, B., Smith, A., 2014. Interactive topic modeling. Machine Learning 95 (3), 423–469.

[8] Jian, L., Yang, L., Suge, W., 2016. The constitution of a fine-grained opinion annotated corpus on weibo. In: China National Conference on Chinese Computational Linguistics. Springer, pp. 227–240.

[9] Lin, Y., Liu, Z., Sun, M., 2016. Knowledge representation learning with entities, attributes and relations. In: Proceedings of the International Joint Conference on Artificial Intelligence. Vol. 1. AAAI Press, pp. 41–52.

[10] Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X., 2015. Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the AAAI Conference on Artificial Intelligence. Citeseer, pp. 2181–2187.

[11] Liu, B., 2015. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

[12] Liu, K., Xu, L., Zhao, J., 2015. Co-extracting opinion targets and opinion words from online reviews based on the word alignment model. IEEE Transactions on Knowledge and Data Engineering 27 (3), 636–650.

[13] Liu, K., Xu, L., Zhao, J., et al., 2014. Extracting opinion targets and opinion words from online reviews with graph co-ranking. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics. ACL, pp. 314–324.

[14] Liu, Q., Gao, Z., Liu, B., Zhang, Y., 2015. Automated rule selection for aspect extraction in opinion mining. In: Proceedings of the International Joint Conference on Artificial Intelligence. AAAI Press, pp. 1291–1297.

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

[16] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems. pp. 3111–3119.

[17] Nguyen, D. Q., Sirts, K., Qu, L., Johnson, M., June 2016. Stranse: a novel embedding model of entities and relationships in knowledge bases. In: Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. ACL, San Diego, California, pp. 460–466.

[18] Oneto, L., Bisio, F., Cambria, E., Anguita, D., 2016. Statistical learning theory and elm for big social data analysis. ieee CompUTATionAl inTelliGenCe mAGAzine 11 (3), 45–55.

[19] Palatucci, M., Pomerleau, D., Hinton, G. E., Mitchell, T. M., 2009. Zero-shot learning with semantic output codes. In: Advances in neural information processing systems. pp. 1410–1418.

[20] Palmer, M., Gildea, D., Xue, N., 2010. Semantic role labeling. Synthesis Lectures on Human Language Technologies 3 (1), 1–103.

[21] Poria, S., Cambria, E., Gelbukh, A., 2016. Aspect extraction for opinion miningwith a deep convolutional neural network. Knowledge-Based Systems.

[22] Poria, S., Cambria, E., Gelbukh, A., Bisio, F., Hussain, A., 2015. Sentiment data flow analysis by means of dynamic linguistic patterns. IEEE Computational Intelligence Magazine 10 (4), 26–36.

[23] Poria, S., Chaturvedi, I., Cambria, E., Bisio, F., 2016. Sentic lda: Improving on lda with semantic similarity for aspect-based sentiment analysis. In: Proceedings of the International Joint Conference on Neural Networks.

[24] Poria, S., Chaturvedi, I., Cambria, E., Hussain, A., Dec 2016. Convolutional mkl based multimodal emotion recognition and sentiment analysis. In: 2016 IEEE 16th International Conference on Data Mining (ICDM). pp. 439–448.

[25] Qiu, G., Liu, B., Bu, J., Chen, C., 2011. Opinion word expansion and target extraction through double propagation. Computational Linguistics 37 (1), 9–27.

[26] Toh, Z., Wang, W., 2014. Dlirec: Aspect term extraction and term polarity classification system. In: Proceedings of the International Workshop on Semantic Evaluation (SemEval 2014). pp. 235–240.

[27] Wang, L., Liu, K., Cao, Z., Zhao, J., de Melo, G., 2015. Sentiment-aspect extraction based on restricted boltzmann machines. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 616–625.

[28] Wang, Z., Zhang, J., Feng, J., Chen, Z., 2014. Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the AAAI Conference on Artificial Intelligence. Citeseer, pp. 1112–1119.

[29] Xie, R., Liu, Z., Jia, J., Luan, H., Sun, M., 2016. Representation learning of knowledge graphs with entity descriptions. In: Thirtieth AAAI Conference on Artificial Intelligence.

[30] Yang, B., Cardie, C., 2012. Extracting opinion expressions with semi-markov conditional random fields. In: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. ACL, pp. 1335–1345.

[31] Yang, B., Cardie, C., 2014. Joint modeling of opinion expression extraction and attribute classification. Transactions of the Association for Computational Linguistics 2, 505–516.

[32] Zafarani, R., Abbasi, M. A., Liu, H., 2014. Social media mining: an introduction. Cambridge University Press.

[33] Zhang, L., Liu, B., 2014. Aspect and entity extraction for opinion mining. In: Data Mining and Knowledge Discovery for Big Data. Springer, pp. 1–40.

[34] Zhao, Y., Qin, B., Liu, T., Yang, W., 2015. Aspect-object alignment with integer linear programming in opinion mining. PloS one 10 (5), e0125084.