IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 28, 2020, accepted December 25, 2020, date of publication January 6, 2021,
date of current version January 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048088

Knowledge-Guided Sentiment Analysis Via Learning From Natural Language Explanations

ZUNWANG KE[1], JIABAO SHENG[2], ZHE LI[2], WUSHOUR SILAMU[1], AND QINGLANG GUO[3], (Associate Member, IEEE)

[1]Xinjiang Laboratory of Multi-Language Information Technology, Xinjiang Multilingual Information Technology Research Center, College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
[2]Xinjiang Laboratory of Multi-Language Information Technology, Xinjiang Multilingual Information Technology Research Center, College of Software, Xinjiang University, Urumqi 830046, China
[3]National Engineering Laboratory for Risk Perception and Prevention (NEL-RPP), China Academy of Electronics and Information Technology, Beijing 100041, China

Corresponding author: Zhe Li (lizhe@stu.xju.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant2017YFC0820700, in part by the National Language Commission Research Project under Grant ZDI135-96, in part by the Xinjiang Uygur Autonomous Region Graduate Research and Innovation Project under Grant XJ2020G071, and in part by the China Academy of Electronics and Information Technology, National Engineering Laboratory for Risk Perception and Prevention (NEL-RPP).

ABSTRACT Sentiment analysis is crucial for studying public opinion since it can provide us with valuable information. Existing sentiment analysis methods rely on finding the sentiment element from the content of user-generated. However, the question of why a message produces certain emotions has not been well explored or utilized in previous works. To address this challenge, we propose a natural language explanation framework for sentiment analysis that provides sufficient domain knowledge for generating additional labelled data for each new labelling decision. A rule-based semantic parser transforms these explanations into programmatic labelling functions that generate noisy labels for an arbitrary amount of unlabelled sentiment information to train a sentiment analysis classifier. Experiments on two sentiment analysis datasets demonstrate the superiority it achieves over baseline methods by leveraging explanations as external knowledge to joint training a sentiment analysis model rather than only labels. An ablation study is conducted to clarify the relative contribution of natural language explanations.

INDEX TERMS Sentiment analysis, natural language explanations, domain knowledge, knowledge-aware, semantic parser, classifier.

I. INTRODUCTION

Sentiment analysis is a significant task in natural language processing and is the core for some prevalent downstream tasks including public opinion analysis [3], [19], [21], [24], [33], [44]. This task focuses on predicting the sentiment information of a given input sentence. However, previous works usually require massive labelled data, which limits their applications in situation where data annotation is expensive. The traditional method of providing supervision is through human-generated labels. For example, given a sentence ''Anyway, the food is good, the price is right and they have a decent wine list'', an annotator should label it as ''Positive''. However, the label does not provide information about how the decision is made. A more informative method is to enable the annotators to explain their decisions in natural language, so that the annotation can generalize to other examples. In the above example, an explanation can be ''Positive, because the word 'food' occurs before 'is good' and the word 'price' precedes the word 'right' within 2 words'', which can generalize to instances such as ''Delicious food with a fair price''. Natural language (NL) explanations have shown effectiveness in providing additional supervision, especially in low-resource settings [10], [34]. Additionally, they can be easily collected from human annotators without significantly increasing the annotation effort.

However, exploiting NL explanations as supervision is challenging due to the complex nature of human languages. First, textual data are not well structured, and thus we must parse explanations into logical forms so that machines can better utilize them. Additionally, linguistic variants are ubiquitous, which makes it difficult to generalize an NL explanation to match sentences that are semantically equivalent but have different word usages. When we perform exact matching

The associate editor coordinating the review of this manuscript and approving it for publication was Seyedali Mirjalili.

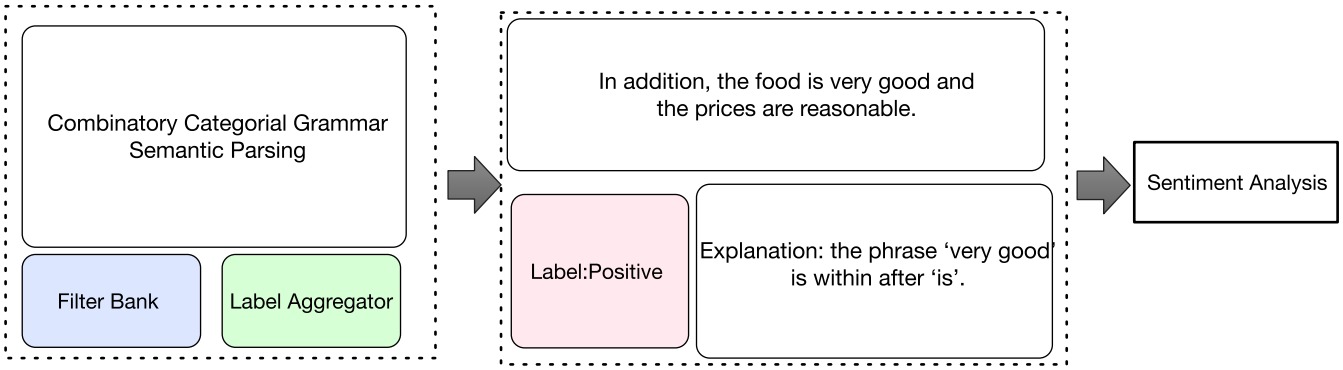

FIGURE 1. High-level illustration of SANLE.

with the previous example explanation, it can fail to annotate sentences with "reasonable prices" or "good bread".

Attempts have been made to train classifiers with NL explanations. Previous works have relied on identifying the relevant input parts including labelling the features [7], [20], [29], highlighting the rationale phrases in text [1], [41], or marking relevant regions in the images [38]. However, certain types of information cannot be simply attributed to annotating a part of the input, such as missing one word or at least two words. In the above example, a sentence such as "Decent bread at a fantastic enough price" will be rejected because of the "directly preceded" requirement. Therefore, we believe that the generalization ability of NL explanations is under-explored. We emphasize that a good data annotation method should 1) be able to generalize annotations to semantically similar instances (beyond stemming, parts of speech, etc.) and 2) model the uncertainty in annotations.

Towards these aims, as shown in Figure 1, we propose the SANLE framework to learn neural models with explanations, as illustrated in Figure 1. Given a raw corpus and a set of NL explanations, we first parse the NL explanations into machine-actionable logical forms with a combinatory categorical grammar (CCG)-based semantic parser. Unlike previous work, we "soften" the annotation process by generalizing the predicates using a neural module network and changing the labelling course from accurate matching to blurred matching. After the filter removes the incorrect semantic interpretation functions, the correct labelling functions are executed on many unlabelled examples and generate a weakly supervised large training dataset. The annotation generated by natural language explanations is used as external knowledge to jointly train the sentiment analysis classifier. The key idea of these proposals is to learn knowledge embeddings and let knowledge participate in computing the attention weights. Our proposed models can concentrate on different parts of a sentence when different pieces of knowledge are provided, so that they are more competitive for the sentiment analysis classification. We conduct experiments on many sentiment analysis tasks. The experimental results demonstrate the superiority of SANLE over various baseline methods.

We make the following key contributions in this work:

- We address the challenge of limited labelled data and the class imbalance problem for the deep learning-based sentiment analysis on social media. We present a data augmentation method based on NL explanations for the sentiment analysis.
- After utilizing a semantic parser to convert the NL explanations into an executable logical form, the aforementioned method employs a neural module network architecture to generalize various forms of actions to label data instances and accumulates the results with soft logic, which greatly increases the coverage of each NL explanation.
- We perform the above by injecting external knowledge with attention-based BiLSTM for the sentiment classification. The models can attend different parts of a sentence when appending the aspect vector into the input word vectors. The results show that the attention mechanism with knowledge is effective. Our experiments on sentiment analysis tasks show the superiority of this method over baseline methods by leveraging NL explanations as the external knowledge to jointly train the classifier.

II. RELATED WORKS

In this section, we will briefly review related works on the sentiment classification [11], [13], knowledge-aware sentiment analysis [9], [14], and natural language explanation [5], [16], [23] classification.

A. SENTIMENT ANALYSIS

Sentiment analysis and emotion recognition have always attracted attention in multiple fields such as NL processing, psychology, and cognitive science. Reference [6] showed a joint encoding model based on Transformer (TBJE) for the sentiment analysis task. Reference [18] captured the implicit and explicit global structural information that resides in the input space to address the targeted sentiment analysis. To address the aspect-based sentiment analysis in two conceptual tasks (aspect extraction and aspect sentiment classification), Reference [26] fully leveraged the importance of syntactical information to explore the grammatical aspects of the sentence and employed a self-attention mechanism for syntactical learning. Reference [39] also addressed this problem by means of the effective encoding of syntax information.

B. KNOWLEDGE-AWARE SENTIMENT ANALYSIS

Recently, many works have investigated how to incorporate knowledge into the Sentiment Analysis. For the cross-domain sentiment analysis prompted by the requirement to address the domain gap among different applications, Reference [8] took a novel perspective by introducing the external world knowledge to enhance the performance of sentiment analysis. Reference [37] proposed a knowledge-based methodology on social networks for sentiment analysis. This work focused on semantic processing considering the content by handling the opinions of public users as excerpts of knowledge. This approach implements knowledge graphs, similarity measures, graph theory algorithms, and a disambiguation process. Reference [15] considered the linguistic knowledge of texts that could promote language understanding in sentiment analysis tasks and first proposed a context-aware sentiment attention mechanism to acquire the sentiment polarity of each word with its part-of-speech tag by querying SentiWord-Net. Then, they devised a new pre-training task called the label-aware masked language model to construct a knowledge-aware language representation. Reference [36] introduced the Sentiment Knowledge Enhanced Pre-training (SKEP) to learn a unified sentiment representation for multiple sentiment analysis tasks. By using the automatically mined knowledge, SKEP conducts sentiment masking and constructs three sentiment knowledge prediction objectives to embed the sentiment information at the word, polarity and aspect levels into a pre-trained sentiment representation. In particular, the prediction of aspect-sentiment pairs is converted into the multi-label classification to capture the dependency between words in a pair.

C. NATURAL LANGUAGE EXPLANATIONS

In leveraging NL to train the classifiers, supervision in the form of NL has been explored by many works. Reference [34] first demonstrated the effectiveness of NL explanations. A joint concept-learning and semantic-parsing method was proposed for classification problems. However, this method is very limited because it cannot use unlabelled data. To address this issue, Reference [10] proposed parsing NL explanations into labelling functions and using data programming to handle the conflicts and enhancements among different labelling functions. Reference [4] extended the Stanford Natural Language Inference dataset with an extra layer of natural language explanations based on human-annotated implication relations. Furthermore, they built a neural network that could directly provide full-sentence NL justifications. Reference [30] proposed the original elementary knowledge Auto-Generated Explanations (CAGE) architecture, which generated helpful explanations by training a language model when it was fine-tuned on the explanations of human and input problems. Then, the classifier model can use these explanations to make predictions. To augment the classification of sequence with natural language explanations, reference [28] proposed an original neural modular execution tree (NMET) architecture. After transforming natural language explanations into executable logical types via a semantic parser, NMET adopts a neural module network framework to generalize diverse forms of actions to label data examples and cumulates the results with soft logic, which significantly improves the range of each natural language explanation.

III. METHODOLOGY

Our framework generates natural language explanations for datasets through semantic parsers as external knowledge, and the knowledge embeddings are taken as the input with the word embedding to jointly train a sentiment analysis model.

A. GENERATE EXTERNAL KNOWLEDGE

As shown in Figure 2, the SANLE framework varies natural language explanations and unlabeled data into noise labelled training sets. The training set includes three key components: semantic parser, filter bank and label aggregator. The semantic parser exchanges NL explanations into a series of logical forms that represent labelling functions. The filter bank deletes some incorrect labelling functions as probable without requiring the ground truth tags. The remaining labelling functions are adopted for unlabelled instances to generate a matrix of labels. This label matrix is passed into the label aggregator, which integrates these possible conflicting and overlapping labels into one label for every example. The resulting labelling instance is utilized to train any discrimination model.

1) SEMANTIC PARSER

To leverage the unstructured human explanations $\varepsilon = e_{j_{j=1}}^{S'}$, we turn them into logical forms [31], which can be denoted as $F = f_j : X \to 0 : 1_{j=1}^{|S'|}$, where 1 indicates that the logical form matches the input sequence and 0 indicates that it does not. To access the labels, we introduce a function $h : F \to Y$ that maps each logical form f_j to label y_j of its explanation e_j. Examples are provided in Fig. 1. We use the combinatory categorical grammar (CCG)-based semantic parsing approach [2], [43], which couples the syntax with the semantics, to convert each NL explanation e_j to a logical form f_j.

Following [34], we first compile a domain lexicon that maps each word to its syntax and logical predicate. Frequently used predicates are listed in the Appendix. For each explanation, the parser can generate many possible logical forms based on the CCG. To identify the correct logical form from among these possible forms, we use a feature vector $\phi(f) =\in R^d$, where each element counts the number of applications of a particular CCG combinator (similar to [43]). Specifically, given explanation e_i, the semantic parser parameterized by $\theta \in R^d$ outputs a probability distribution over all possible logical forms Z_{e_i}. The probability of a feasible logical form can be calculated as:

$$P_\theta(f|e_i) = \frac{\exp \theta^T \phi(f)}{\sum_{f':f' \in Z_{e_j}} \exp \theta^T \phi(f)}. \quad (1)$$

To learn θ, we maximize the probability of y_i given e_i, which is calculated by marginalizing over all logical forms

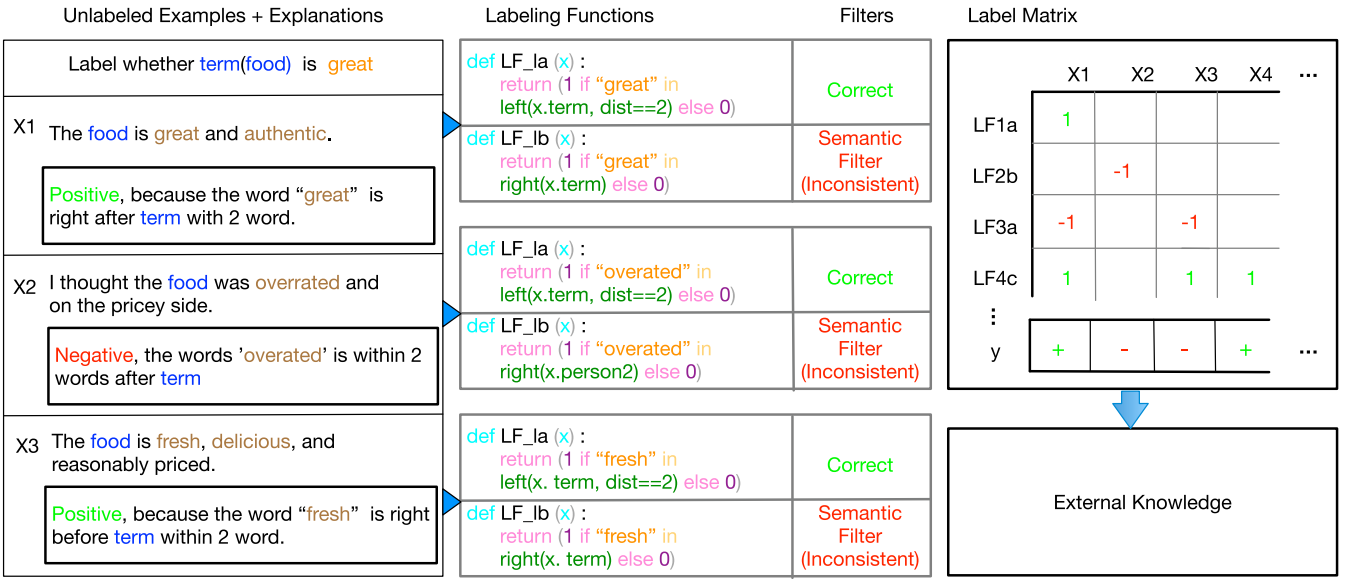

FIGURE 2. Our model combines LFs [10] with GCG [2], and the natural language explanations are parsed into candidate labelling functions (LFs) via semantic parsing GCG. Many incorrect LFs are automatically filtered out by the filter bank. The remaining functions provide heuristic labels over the unlabelled dataset, which are aggregated into one noisy label per example and yield a large, noisily labelled training set for a classifier.

that match x_i (similar to Liang *et al.* (2013)). Formally, the objective function is defined as:

$$L_{parser} = \sum_{i=1}^{|S'|} \log(\sum_{f:f(x_i)=1 \wedge h(f)=y_i} P_{\theta*}(f|e_i)). \quad (2)$$

When the optimal $\theta*$ is derived using the gradient-based method, the parsing result for e_i is defined as $f_i = argmax_f P_{\theta*}(f|e_i)$.

2) FILTER BANK

The filter bank input is a series of candidate label functions (LFs) generated by the semantic parser. The goal of the filter bank is to abandon some wrong LFs as probable without claiming extra labels. It includes semantic filters and pragmatic filters.

Each explanation e_i is gathered in the circumstances of a particular labelled instance (x_i, y_i). The semantic filter verifies the LFs that vary with their corresponding instance; Any LF f for which $f(x_i)/ = y_i$ is abandoned. Finally, of all LFs that are similarly interpreted by all other filters, we only retain the most specific (lowest coverage) LF, which prevents multiple related LFs from dominating in a single instance.

3) LABEL AGGREGATOR

The label aggregator integrates diverse suggested labels, which form the LFs, and integrates them into a single label of each probabilistic instance. Specifically, if m LFs through the filter bank are adopted to n instances, the label aggregator executes a function $f : \{-1, 0, 1\}^{mxn} \rightarrow [0, 1]^m$. A simple solution is to adopt a simple majority vote, but this solution neglects to consider the truth that LFs can have a broad range of precision and coverage. Therefore, we utilize data programming [31] to model the relationship of the true labels, and the labelling functions output a factor graph. Furthermore, given the true labels $Y \in \{-1, 1\}^n$ (latent) and label matrix $\Lambda \in \{-1, 0, 1\}$ (observed), where $\Lambda_{i,j} = LF_i(x_j)$, we define the labelling propensity and accuracy as two types of factors:

$$\phi_{i,j}^{Lab}(\Lambda, Y) = 1\Lambda_{i,j} \neq 0 \quad (3)$$

$$\phi_{i,j}^{Acc}(\Lambda, Y) = 1\Lambda_{i,j} = y_i. \quad (4)$$

Hence, for factors concerning to a given point of data x_j as $\phi_j(\Lambda, Y) \in \mathbb{R}^m$,

$$p_w(\Lambda, Y) = Z_w^{-1} \exp(\sum_{j=1}^{n} w \cdot \phi_j(\Lambda, Y)), \quad (5)$$

where $w \in \mathbb{R}^{2m}$ is the weight vector, and Z_w is the normalization constant. To determine this model without identifying the true marks Y, with the perceived labels λ, we minimize the negative log marginal likelihood:

$$\hat{w} = argmin_w - \log \sum_Y p_w(\lambda, Y) \quad (6)$$

By applying stochastic gradient descent (SGD) and Gibbs sampling to infer and utilize the marginals $p_{\hat{w}}(\lambda, Y)$ as probabilistic training labels. We conclude the accuracies of the LFs because they overhang and conflict with one another. Due to high conflict rates in noisier LFs with others, their corresponding accuracy weights in w will be smaller, which decreases their impact on the aggregated labels.

B. SENTIMENT CLASSIFIER

For the results in this paper, our discriminative model is an attention-based BiLSTM Network for sentiment classification. The standard LSTM cannot detect the important part for the aspect-level sentiment classification, and the attention mechanism can capture the key part of a sentence in response

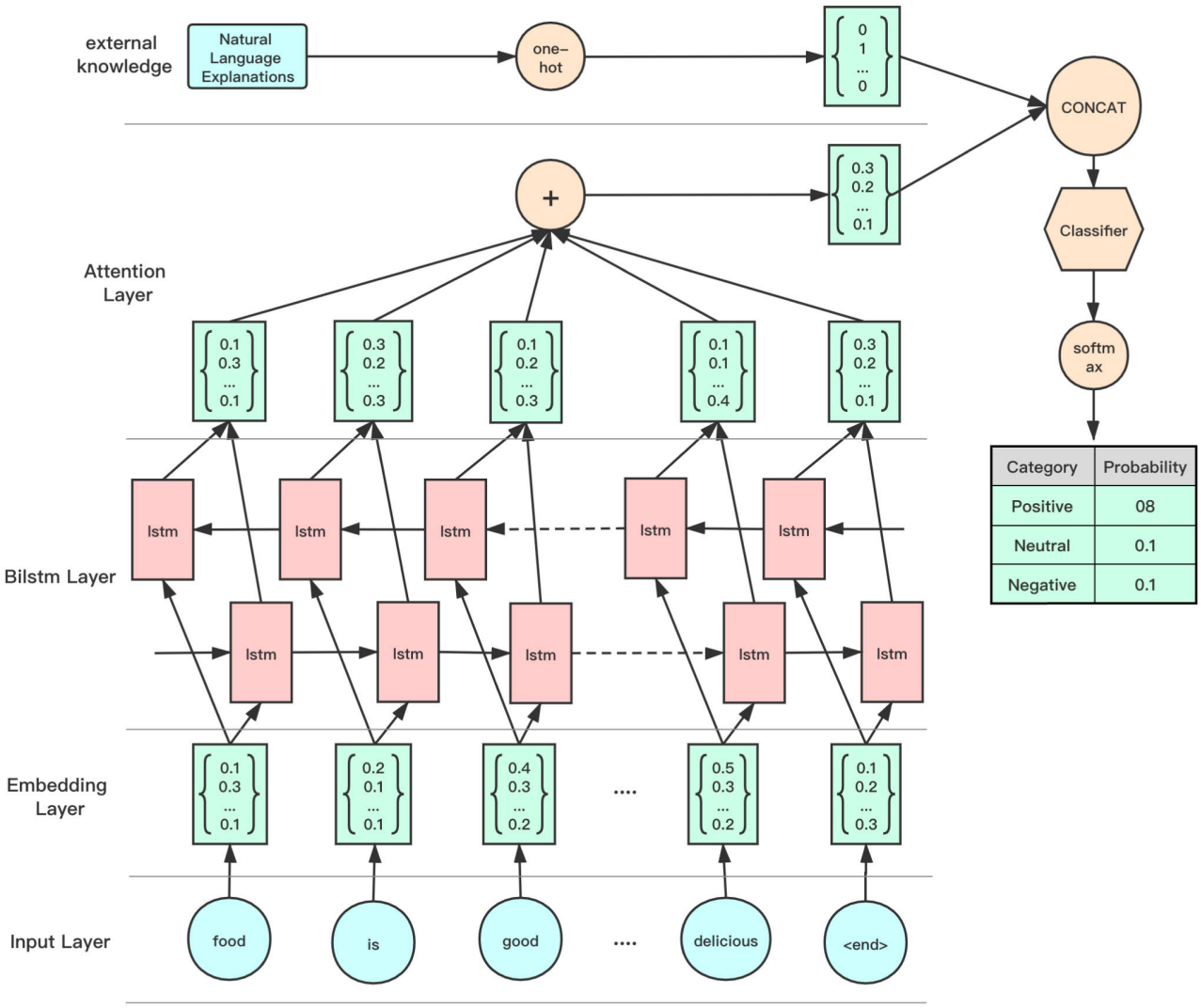

FIGURE 3. Sentiment analysis classifier. Architecture of the attention-based BiLSTM with external knowledge.

to a given aspect. The attention mechanism can concentrate on different parts of a sentence when different aspects are taken as the input. Figure-3 represents the architecture of an Attention-based BiLSTM.

IV. EXPERIMENTS

A. DATASETS

We conduct experiments on the sentiment analysis with sentiment analysis points to define the sentiment regarding a given sentence. For example, in the sentence *The sweet lassi was excellent as was the lamb chettinad and the garlic naan but the rasamalai was forgettable*, the sentiment is positive, and the explanation can be that *The word was is directly succeeded by excellent.* For this task, we use two customer review datasets: Restaurant and Laptop, which are part of SemEval 2014 Task 4 [12], [27]. The dataset consists of customer reviews. Each review contains a list of aspects and corresponding polarities. Our aim is to identify the sentiment of a sentence. The statistics is presented in Table 1.

TABLE 1. Statistics of the SemEval 2014 Task 4. Aspect distribution per sentiment class. Fo., Pr., Se, Am., An. refer to food, price, service, ambience, anecdotes/miscellaneous. "Asp." refers to aspect.

Asp.	Positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
Fo.	867	302	209	69	90	31
Pr.	179	51	115	28	10	1
Se.	324	101	218	63	20	3
Am.	263	76	98	21	23	8
An.	546	127	199	41	357	51
Total	2179	657	839	222	500	94

B. BASELINES

The logical forms are used to partition the unlabelled corpus D into labelled set D_a and unlabelled set D_u. Labelled set D_a can be directly utilized by supervised learning methods.

(1) **CBOW-GloVe** uses a bag-of-words [22] on GloVe embeddings [25] to represent the instance or surface patterns in an NL explanation. Then, it annotates the sentence with the label of its most similar surface pattern (as with cosine similarity).

(2) **PCNN** [42] uses piecewise max-pooling to aggregate CNN-generated features.

(3) **ATAE-LSTM** [40] incorporates the aspect term information into both embedding layer and attention layer to help the model concentrate on different parts of a sentence.

In the semi-supervised baselines, unlabelled data D_u are introduced for training. For methods that require rules as the input, we use surface pattern-based rules transferred from the explanations. The compared semi-supervised methods include the following:

(4) **Data Programming** [10], [31] aggregates the results of strict labelling functions for each instance and uses these pseudo-labels to train a classifier.

(5) **Self-Training** [32] expands the labelled data by selecting the batch of unlabelled data with the highest confidence and generating pseudo-labels for them. The method stops when all unlabelled data are used.

(6) **Pseudo-Labelling** [17] first trains a classifier on a labelled dataset and subsequently generates pseudo-labels for the unlabelled data using the classifier by selecting the label with the maximum predicted probability.

(7) **Mean-Teacher** [35] uses the averaged model weights instead of the label predictions and assumes that similar data points have similar outputs.

Learning from explanations is categorized as a third setting. Both methods generate explanation-guided pseudo-labels for a downstream classifier.

(8) **SANLE** (proposed work) softly applies logical forms to obtain annotations for the unlabelled instances and trains a downstream classifier with these pseudo-labelled instances.

C. HYPER-PARAMETERS

To achieve the ideal classification effect, we repeat the experiments and adjust the model's hyper-parameters. The experimental parameters are selected by minimizing the cross-entropy. The size of an LSTM cell is set to 100. The dropout is set to 0.5 to prevent overfitting. The model uses 128 mini-batches and Adam optimization algorithm. In the experiment, the best accuracy is achieved when the number of iterations in model training is 20.

D. EXPERIMENTAL RESULTS

Table-3 lists the precision, recall, and F1 scores of all sentiment analysis models. The baseline models experimental results are directly cited from previous studies [28]. Our proposed SANLE consistently outperforms all baseline models in a low-resource setting. We also find the following results: (1) Directly applying logical forms to unlabelled data results in poor performance. This method achieves high precision but low recall, as expected. (2) Compared to its downstream classifier baseline, SANLE achieves the best F1 score, which confirms that the expansion of rule coverage by SANLE is effective and provides useful information for classifier training. (3) The semi-supervised methods have unsatisfactory results, which can be explained by the difference between the underlying data distributions of $\mathcal{D}_a$ and $\mathcal{D}_u$.

TABLE 2. Experiment results on sentiment analysis. The best and second-best results in each metric are bold and underlined, respectively.

	Restaurant			Laptop		
Metric	Precision	Recall	F1	Precision	Recall	F1
LF(ε)	**86.5**	4.0	7.7	**90.0**	7.1	13.1
CBOW-GloVe ($\mathcal{R}+\mathcal{D}$)	62.8	75.3	68.5	53.4	**72.6**	61.5
PCNN ($\mathcal{D}_a$)	67.1	79.0	72.6	53.1	71.4	60.9
ATAE-LSTM ($\mathcal{D}_a$)	65.1	78.4	71.1	49.0	66.0	56.2
ATAE-LSTM ($\mathcal{D}_l$)	65.3	78.9	71.4	48.9	55.6	52.0
Data Programming ($\varepsilon+\mathcal{S}$)	65.0	78.8	71.2	53.4	72.5	61.5
Self Training ($\mathcal{D}_a$+$\mathcal{D}_u$)	65.3	78.4	71.2	50.1	67.7	57.6
Pseudo Labeling ($\mathcal{D}_a$+$\mathcal{D}_u$)	64.9	78.0	70.9	50.4	68.4	58.0
Mean Teacher ($\mathcal{D}_a$+$\mathcal{D}_u$)	68.8	75.7	72.0	54.4	72.3	<u>62.1</u>
Mean Teacher ($\mathcal{D}_l$+$\mathcal{D}_{lu}$)	68.3	<u>81.0</u>	<u>74.1</u>	55.0	70.3	61.7
SANLE ($\varepsilon+\mathcal{D}$)	<u>69.1</u>	**81.5**	**74.7**	<u>54.7</u>	**72.6**	**62.3**

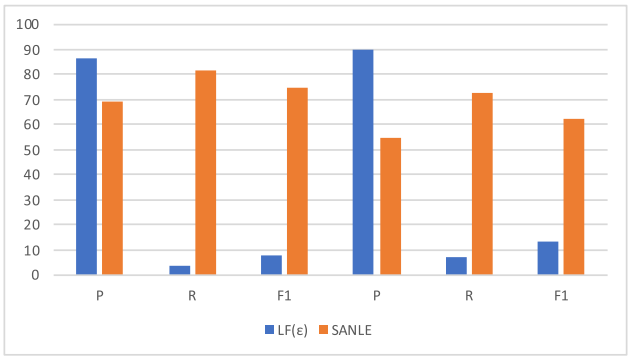

FIGURE 4. Ablation study on semantic parsing. The precision, recall, and F1 score on the test set are reported. We compared two semantic parsers.

E. ABLATION STUDY

To explore the effect of key factors on our proposed model, we assess the achievement of SANLE with respect to its rate of advancement by semantic parsing, its reliance on correctly parsed relevant information, and the logical form mechanism that it utilizes.

1) EFFECTIVENESS OF SEMANTIC PARSING

As shown in Figure 4, we conduct ablation studies on Laptop and Restaurant. We changed the semantic parser from CGC to LF [10] to see how much rule softening helped in our framework. We can easily conclude that the CGC semantic parser module plays a vital role. Modifying it results in a significant performance drop, which demonstrates the effectiveness of generalization when applying logical forms and indicates the importance of semantic learning when performing fuzzy matching.

2) EFFECT OF INCORRECT PARSING

In Figure 5, we describe the semantic parser case before and after filtering. The semantic parser accuracy is based on an accurate match with a manually produced parse per explanation. The simple filter-bank-based heuristic strongly eliminates the most inaccurate semantic parsers in the sentiment analysis tasks. Intuitively, the filters are powerful, since it is difficult for a semantic parser to be parsed from an instance, perfectly label its own example and not label all cases in the training dataset with the same description or identically to extra LF. While users provide explanations, the signs that they express present good origin points, but they are unlikely to be optimal. This result explains that the filter bank is required

TABLE 3. Examples for SANLE explanations.

Sentence	In addition, the food is very good and the prices are reasonable.
Lable	positive
Explanation	the phrase 'very good' is within after 'is'.
Sentence	Their calzones are horrific, bad, vomit-inducing, YUCK.
Lable	negative
Explanation	the words 'horrific, bad, vomit-inducing, YUCK' is after 'are'.
Sentence	Great laptop that offers many great features!
Lable	positive
Explanation	the word 'great' occurs after 'offers' by no more than 2 words.
Sentence	One night I turned the freaking thing off after using it. The next day I turn it on, no GUI, screen all dark, power light steady, hard drive light steady and not flashing as it usually does.
Lable	positive
Explanation	The word 'freaking' occurs before 'thing'.

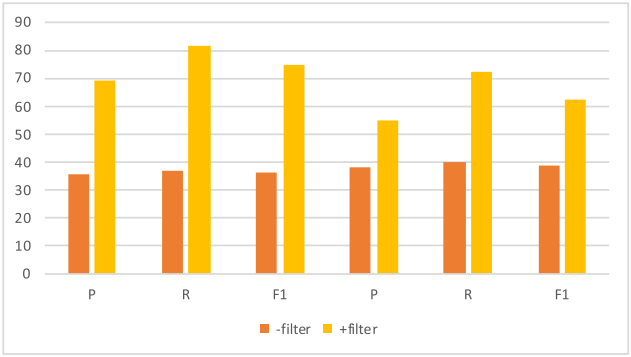

FIGURE 5. Precision, recall, and F1 scores obtained using SANLE with no filter bank and as normal.

to liquidate certainly unrelated semantic analysis labels, but with this issue, the naive semantic parser based on a rule and a perfect parser have almost equal average F1 scores.

F. CASE STUDY

In this section, we show the results of applying our proposed model to the NL interpretation of the dataset. As seen from the examples given in the table, using our correct semantic parser can provide good NL interpretations of sentences. our semantic parser, can accurately identify words expressing emotional polarity and make a reasonable interpretation of the sentence, as external knowledge to jointly train the sentiment classifier in order to improve the accuracy of the emotion dichotomy.

G. DISCUSSION

We obtain the appropriate logical forms from user-given explanations, and we have various choices for how to use them. Reference [34] proposes adopting those logical forms as characteristics in a linear classifier, essentially using a traditional supervision approach with user-specified features. We prefer to apply them as functions to weakly supervise a larger training dataset over data programming [31]. In the above experiment, we find that the use of NL explanations can effectively improve the performance of the sentiment analysis tasks. More commonly, from a machine learning viewpoint, labels are the principal asset, but they are a low-bandwidth sign among the annotators and the learning algorithm. NL as external knowledge presents a much higher-bandwidth connection pipe. We have displayed encouraging results in the sentiment analysis task, and it will be fascinating to enlarge our framework to other tasks and more interactive perspectives.

V. CONCLUSION AND FUTURE WORK

In this paper, we have presented SANLE, a framework that augments sequence classification by exploiting NL explanations as the external knowledge supervision in a low-resource setting. We addressed the challenges of modelling the compositionality of NL explanations and handling the linguistic variants. Semantic parsers, semantic filter banks, and label aggregators were introduced to generalize different types of actions through logical forms, which substantially increased the coverage of the NL explanations. Our current study has demonstrated the potential efficiency and effectiveness of semantically augmented data in combating the labelled data scarcity and class imbalance problems of publicly available sentiment analysis datasets. We conducted extensive experiments on two datasets and proved the effectiveness of our model.

In future work, we plan to augment data via NL explanations in low-resource language sentiment analysis datasets to build comprehensive datasets for the sentiment analysis and conduct experiments on sentiment analysis via deep learning. We will evaluate the effectiveness of the augmented data in alleviating overfitting and its usefulness in facilitating deeper neural networks for the sentiment analysis. Further experiments will be conducted to examine the generalization of sentiment analysis models to unseen causes of emotions.

ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable feedback. *(Zunwang Ke, Jiabao Sheng, and Zhe Li contributed equally to this work.)*

REFERENCES

[1] S. Arora and E. Nyberg, "Interactive annotation learning with indirect feature voting," in *Proc. Hum. Lang. Technol., Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Companion Volume, Student Res. Workshop Doctoral Consortium (NAACL)*, 2009, pp. 55–60.

[2] Y. Artzi, K. Lee, and L. Zettlemoyer, "Broad-coverage CCG semantic parsing with AMR," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1699–1710.

[3] R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in *Proc. 3rd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2016, pp. 452–455.

[4] O.-M. Camburu, T. Rocktäschel, T. Lukasiewicz, and P. Blunsom, "E-SNLI: Natural language inference with natural language explanations," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 9539–9549, 2018.

[5] O.-M. Camburu, B. Shillingford, P. Minervini, T. Lukasiewicz, and P. Blunsom, "Make up your mind! Adversarial generation of inconsistent natural language explanations," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association Computational Linguistics, 2020, pp. 4157–4165. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.382

[6] J.-B. Delbrouck, N. Tits, M. Brousmiche, and S. Dupont, "A transformer-based joint-encoding for emotion recognition and sentiment analysis," in *Proc. 2nd Grand-Challenge Workshop Multimodal Lang. (Challenge-HML)*, 2020, pp. 1–7.

[7] G. Druck, B. Settles, and A. McCallum, "Active learning by labeling features," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, vol. 1, 2009, pp. 81–90.

[8] D. Ghosal, D. Hazarika, A. Roy, N. Majumder, R. Mihalcea, and S. Poria, "KinGDOM: Knowledge-guided DOMain adaptation for sentiment analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association Computational Linguistics, 2020, pp. 3198–3210. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.292

[9] C. Gong, J. Yu, and R. Xia, "Unified feature and instance based domain adaptation for aspect-based sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*. Stroudsburg, PA, USA: Association Computational Linguistics, 2020, pp. 7035–7045. [Online]. Available: https://www.aclweb.org/anthology/2020.emnlp-main.572

[10] B. Hancock, P. Varma, S. Wang, M. Bringmann, P. Liang, and C. Ré, "Training classifiers with natural language explanations," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, vol. 1, 2018, p. 1884.

[11] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An interactive multi-task learning network for end-to-end aspect-based sentiment analysis," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*. Florence, Italy: Association Computational Linguistics, 2019, pp. 504–515. [Online]. Available: https://www.aclweb.org/anthology/P19-1048

[12] I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó. Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz, "SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals," in *Proc. Workshop Semantic Eval., Recent Achievements Future Directions (DEW)*, 2009, pp. 33–38.

[13] M. Hu, Y. Peng, Z. Huang, D. Li, and Y. Lv, "Open-domain targeted sentiment analysis via span-based extraction and classification," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*. Florence, Italy: Association Computational Linguistics, 2019, pp. 537–546. [Online]. Available: https://www.aclweb.org/anthology/P19-1051

[14] J. Huang, Y. Meng, F. Guo, H. Ji, and J. Han, "Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*. Stroudsburg, PA, USA: Association Computational Linguistics, 2020, pp. 6989–6999. [Online]. Available: https://www.aclweb.org/anthology/2020.emnlp-main.568

[15] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, "Sentilare: Linguistic knowledge enhanced language representation for sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 6975–6988.

[16] S. Kumar and P. Talukdar, "NILE : Natural language inference with faithful natural language explanations," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association Computational Linguistics, 2020, pp. 8730–8742. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.771

[17] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks,".

[18] H. Li and W. Lu, "Learning explicit and implicit structures for targeted sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 5481–5491.

[19] N. Li and D. D. Wu, "Using text mining and sentiment analysis for online forums hotspot detection and forecast," *Decis. Support Syst.*, vol. 48, no. 2, pp. 354–368, Jan. 2010.

[20] P. Liang, M. I. Jordan, and D. Klein, "Learning from measurements in exponential families," in *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, 2009, pp. 641–648.

[21] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing*, vol. 2. 2010, pp. 627–666.

[22] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[23] S. Murty, P. W. Koh, and P. Liang, "ExpBERT: Representation engineering with natural language explanations," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association Computational Linguistics, 2020, pp. 2106–2113. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.190

[24] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Comput. Hum. Behav.*, vol. 31, pp. 527–541, Feb. 2014.

[25] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.

[26] M. H. Phan and P. O. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 3211–3220.

[27] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*. Dublin, Ireland: Association Computational Linguistics, 2014, pp. 27–35. [Online]. Available: https://www.aclweb.org/anthology/S14-2004

[28] Y. Qin, Z. Wang, W. Zhou, J. Yan, Q. Ye, X. Ren, L. Neves, and Z. Liu, "Learning from explanations with neural module execution tree," in *Proc. Int. Conf. Learn. Represent.*, 2020.

[29] H. Raghavan, O. Madani, and R. Jones, "Interactive feature selection," in *Proc. IJCAI*, vol. 5, 2005, pp. 841–846.

[30] N. F. Rajani, B. McCann, C. Xiong, and R. Socher, "Explain yourself! Leveraging language models for commonsense reasoning," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 4932–4942.

[31] A. J. Ratner, C. M. D. Sa, S. Wu, D. Selsam, and C. Ré, "Data programming: Creating large training sets, quickly," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3567–3575.

[32] C. Rosenberg, M. Hebert, and H. Schneiderman, "Semi-supervised self-training of object detection models," in *Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION)*, vol. 1, Jan. 2005, pp. 29–36.

[33] J. Smailović, M. Grčar, N. Lavrač, and M. Žnidaršič, "Predictive sentiment analysis of tweets: A stock market application," in *Proc. Int. Workshop Hum.-Comput. Interact. Knowl. Discovery Complex, Unstructured, Big Data*. Berlin, Germany: Springer, Jul. 2013, pp. 77–88.

[34] S. Srivastava, I. Labutov, and T. Mitchell, "Joint concept learning and semantic parsing from natural language explanations," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1527–1536.

[35] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 1195–1204.

[36] H. Tian, C. Gao, X. Xiao, H. Liu, B. He, H. Wu, H. Wang, and F. Wu, "SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis," 2020, *arXiv:2005.05635*. [Online]. Available: http://arxiv.org/abs/2005.05635

[37] J. Vizcarra, K. Kozaki, M. Torres Ruiz, and R. Quintero, "Knowledge-based sentiment analysis and visualization on social networks," *New Gener. Comput.*, pp. 1–31, Aug. 2020.

[38] L. von Ahn, R. Liu, and M. Blum, "Peekaboom: A game for locating objects in images," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI)*, 2006, pp. 55–64.

[39] K. Wang, W. Shen, Y. Yang, X. Quan, and R. Wang, "Relational graph attention network for aspect-based sentiment analysis," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*. Stroudsburg, PA, USA: Association Computational Linguistics, 2020, pp. 3229–3238. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-main.295

[40] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.

[41] O. Zaidan and J. Eisner, "Modeling annotators: A generative approach to learning from annotator rationales," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2008, pp. 31–40.

[42] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant supervision for relation extraction via piecewise convolutional neural networks," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1753–1762.

[43] L. S. Zettlemoyer and M. Collins, "Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars," in *Proc. 21st Conf. Uncertainty Artif. Intell.*, 2005, pp. 658–666.

[44] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 8, no. 4, p. e1253, 2018.

ZUNWANG KE is currently pursuing the Ph.D. degree with Xinjiang University. He is also a Lecturer with Xinjiang University. He has presided over one key project. His research interest includes natural language processing.

JIABAO SHENG is currently pursuing the master's degree in software engineering with Xinjiang University. Her research interest includes knowledge graph. She is a member of the Chinese Information Processing Society of China.

ZHE LI is currently pursuing the master's degree in software engineering with Xinjiang University. His research interests include text generation and social computing. He is a member of the Chinese Information Processing Society of China and the Chinese Association for Artificial Intelligence.

WUSHOUR SILAMU is currently a Professor with Xinjiang University, where he is also a Ph.D. Supervisor. He is also an Academician of the Chinese Academy of Engineering. He is also an Executive Director of the Chinese Association for Artificial Intelligence. He has published more than 200 articles and presided over 65 key projects, including seven national 863 projects and one national 973 project. He has presided over the formulation of five international standards and more than 14 national standards. His research interest includes multilingual natural language processing. He received three National Science and Technology Progress Award.

QINGLANG GUO (Associate Member, IEEE) is currently pursuing the Ph.D. degree in engineering with the University of Science and Technology of China. He also works as an Engineer with the National Engineering Laboratory for Risk Perception and Prevention (NEL-RPP), China Academy of Electronics and Information Technology. His research interests include social networks and graph convolution networks.