Enhancing Cross-target Stance Detection with Transferable Semantic-Emotion Knowledge

Bowen Zhang[1], Min Yang[2], Xutao Li[3]*, Yunming Ye[3]*, Xiaofei Xu[1], Kuai Dai[3]
[1]Harbin Institute of Technology, Harbin, China
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[3]Harbin Institute of Technology, Shenzhen, China

Abstract

Stance detection is an important task, which aims to classify the attitude of an opinionated text towards a given target. Remarkable success has been achieved when sufficient labeled training data is available. However, annotating sufficient data is labor-intensive, which establishes significant barriers for generalizing the stance classifier to the data with new targets. In this paper, we proposed a Semantic-Emotion Knowledge Transferring (SEKT) model for cross-target stance detection, which uses the external knowledge (semantic and emotion lexicons) as a bridge to enable knowledge transfer across different targets. Specifically, a semantic-emotion heterogeneous graph is constructed from external semantic and emotion lexicons, which is then fed into a graph convolutional network to learn multi-hop semantic connections between words and emotion tags. Then, the learned semantic-emotion graph representation, which serves as prior knowledge bridging the gap between the source and target domains, is fully integrated into the bidirectional long short-term memory (BiLSTM) stance classifier by adding a novel knowledge-aware memory unit to the BiLSTM cell. Extensive experiments on a large real-world dataset demonstrate the superiority of SEKT against the state-of-the-art baseline methods.

1 Introduction

The goal of stance detection is to automatically predict the attitude (i.e., *favor*, *against*, or *none*) of an opinionated text towards a given target (Du et al., 2017). Recently, deep learning methods, such as convolutional neural network (CNN) and long short-term memory (LSTM) (Augenstein et al., 2016; Du et al., 2017), have dominated the study of stance detection. Impressive stance detection performances have been achieved when a large number of labeled samples are available. However, obtaining rich annotated data is a time-consuming and labor-intensive process. Conventional stance detection methods are struggling to cope well with the data across targets. This motivates the studies of cross-target stance detection (Wei and Mao, 2019), which infers the attitude of the destination target by leveraging a large amount of annotated data from the source target.

So far, several previous studies have been conducted for cross-target stance detection (Augenstein et al., 2016; Xu et al., 2018; Wei and Mao, 2019). These methods leverage either common words or concept-level knowledge shared by different targets to bridge the knowledge gap across the different targets. Such models suffer from two issues when they are applied to cross-target stance detection in practice. First, stance detection often involves analyzing the texts from social media that are short and informal, making it difficult to extract domain-independent common words shared by different targets from the training data. Second, users may express their stance towards a given target in an implicit way. Thus, the existing concept-level based methods may fail to distinguish implicit stance-carrying terms and context information.

To alleviate the aforementioned issues, we propose a semantic-emotion knowledge transferring (SEKT) model for cross-domain stance detection, which leverages external knowledge as a bridge between source and destination targets. The proposed model is motivated by the observation that the data with different targets usually shares certain common external knowledge that can be transferred from the source to destination targets. **First**, we build a semantic-emotion graph (SE-graph) from semantic-related and emotion-related lexicons, which incorporates external knowledge from both word-level and concept-level. In SE-graph, each node is either a word or an emotion tag, and

*corresponding authors: {lixutao, yym}@hit.edu.cn

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3188–3197
July 5 - 10, 2020. ©2020 Association for Computational Linguistics

the edge between each node pair indicates the co-occurrences of the two nodes in the lexicons. **Second**, a graph convolutional network (GCN) (Kipf and Welling, 2016) is employed to learn the graph representation that captures the multi-hop semantic connections between words or emotion tags rather than one-hop connection. **Third**, we extend the standard bidirectional LSTM (BiLSTM) classifier to fully integrate the external knowledge (SE-graph) by adding an additional knowledge-aware memory unit (KAMU) to the LSTM cell. KAMU is capable of controlling the influence of the external knowledge in learning the hidden state of each word.

The main contributions of this paper can be summarized as follows:

- We construct a semantic-emotion heterogeneous graph from external semantic and emotion lexicons, and employ GCN to learn the semantic graph representation. The external knowledge enriches the representation learning of the text and target and can be used as a bridge to enable knowledge transfer across different targets.
- We extend the standard LSTM cell with an additional memory unit, effectively integrating external knowledge into the classifier for stance detection.
- We conduct extensive experiments on a large dataset expanded from SemEval-2016 Task 6 to verify the effectiveness of our model for cross-domain stance detection. The experimental results show that our model consistently outperforms the compared methods.

2 Related Work

2.1 In-domain Stance Detection

Stance detection aims to infer the attitude of a text towards specific target expression, which is related to argument mining, fact-checking, and aspect-level sentiment analysis. Early stance detection methods were concentrated on debates (Thomas et al., 2006; Somasundaran and Wiebe, 2009; Walker et al., 2012). In recent years, mining users' stance from social media has attracted increasing attention due to its broad applications (Du et al., 2017; Dey et al., 2018; Wei et al., 2018). For example, Du et al. (2017) incorporated target-specific information into stance classification with an attention mechanism. Dey et al. (2018) proposed a two-phase RNN method, where the first phase is to filter the non-neutral text while the second phase is to classify the attitude. Wei et al. (2018) further extended the model to deal with multi-target stance detection and utilized a shared memory network to capture the stance related information towards multiple related targets. Sun et al. (2018) adopted a hierarchical attention method to construct text representation with various linguistic factors.

2.2 Cross-target Stance Detection

There are also several studies being developed for cross-target stance detection problems, which can be divided into two classes. The first one mainly focuses on word-level transfer, which utilizes the common words shared by two targets to bridge the knowledge gap. For example, Augenstein et al. (2016) proposed a bidirectional conditional encoding method by incorporating the target to learn the target-specific words. Xu et al. (2018) further utilized the self-attention mechanism to identify the word importance. The second type of approach attempts to address this transfer learning problem with concept-level knowledge shared by two targets. For example, Wei and Mao (2019) proposed a variational Transfer Network (VTN) method, which complements the commonly used knowledge by inferring the latent topics shared by the two targets.

2.3 Incorporating External Knowledge

There are also plenty of studies that incorporate external resources, such as prior knowledge, grammar rules, domain descriptions, into deep learning framework to address the data sparsity issue (Zhang et al., 2018; Dragoni and Petrucci, 2018; Zhang et al., 2019b; Hu et al., 2016). For example, Lei et al. (2018) integrated the external knowledge in the word embedding layer. Margatina et al. (2019) combined the external knowledge with the hidden layer acquired by RNN. However, these methods ignored the relations between external knowledge and input context. Ma et al. (2018) developed a Sentic LSTM method, which contained an additional affective gate mechanism in the LSTM cell to assist in learning knowledge-aware context representation.

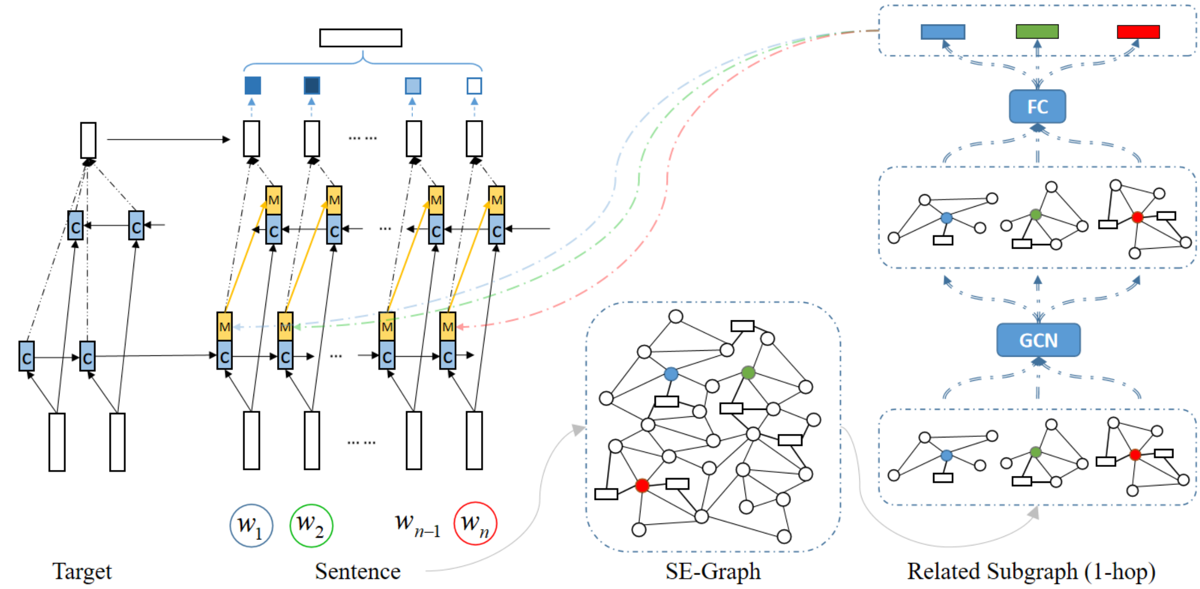

Figure 1: The framework of the proposed SEKT model for cross-target stance detection. It consists of two main components, i.e., SE-graph and knowledge-enhanced BiLSTM.

3 Our Methodology

3.1 Task Definition and Model Overview

We use $X^s = \{x_i^s, p_i^s\}_{i=1}^{N^s}$ to denote the collection of labeled data in the source domain, where each x denotes the input text and p denotes the corresponding target. N^s represents the number of instances in X^s. Each sentence-target pair $(x^s, p^s) \in X^s$ has a stance label y^s. Given an input sentence x^t and a corresponding target p^t in the target domain, this study aims to predict a stance label for the input sentence x^t towards the given target p^t by using the model learned with the labeled data X^s in source domain.

As illustrated in Figure 1, our model consists of two primary components: a semantic-emotion graph (SE-graph) network and a knowledge-enhanced BiLSTM network. First, we build SE-graph from semantic-related and emotion-related lexicons, where GCN is employed to learn the graph representation that captures the semantic connections between words or emotion tags with the multi-hop connection. Then, we extend the BiLSTM classifier to fully integrate the SE-graph by adding a novel knowledge-aware memory unit (KAMU) to the LSTM cell. Next, we will introduce the main components of our model in detail.

3.2 Semantic-Emotion Knowledge Graph Construction

The data in different domains usually shares certain background knowledge that can possibly be transferred from the source domain to the target domain. Thus, we leverage external knowledge as a bridge between the source and target domains.

To this end, we build a semantic-emotion knowledge graph (SE-graph) to represent the external knowledge that may contribute to cross-target stance detection. The SE-graph utilizes the words or emotion tags in the semantic and emotion lexicons as nodes, and constructs weighted edges between words or emotion tags based on their co-occurrence frequency. First, we utilize the whole words from the semantic lexicon SenticNet (Cambria et al., 2018) as the word-nodes and add edges between the semantic words that capture the word-word semantic connections. Second, we attempt to assign emotion tags to the words in SenticNet by looking for the emotion lexicon EmoLex (Mohammad and Turney, 2013), and add edges between the words and emotion tags that capture the word-tag connection. For example, for a word “*mad*” in SenticNet, its semantic-related words from SenticNet are ”*resent*, *malice*, *rage*, *temper*”, and the corresponding emotion tags from EmoLex are “*#anger*’, *#disgust*”. In this way, we can construct a weighted SE graph G. However, each emotion tag (node) represents a concept-level knowledge, which tends to have many connected nodes. As a result, emotional knowledge may dominate the input text. To alleviate this issue, we re-scale the weights of the word-tag edges by a constant.

The SE-graph can capture the semantic connections between words and emotion tags with multi-hop connections. It can help the stance detector to differentiate the important and appropriate words for knowledge transfer. Intuitively, the nodes with high degrees can be considered as the words that contain common background knowledge, which often act as a bridge between different targets.

3.3 SE-graph Embedding

We learn the embedding of each node in the SE-graph with graph convolutional network (GCN), aiming to fully exploit the multi-hop semantic and emotional connections between the nodes. Due to the semantic locality between the words, we extract a k-hop subgraph from SE-graph for each word. The subgraph is then fed into a GCN to learn the graph representation. Here, we adopt GCN because it has been proved to be effective and efficient to learn graph embedding (Zhang et al., 2019a).

Formally, let $E \in \mathbb{R}^{v \times d}$ be a matrix containing all v nodes in SE-graph with their features, where d is the size of the node embedding. For each node, we extract a k-hop subgraph G_s from the whole graph, which has a degree matrix D and adjacency matrix A. The normalized symmetric adjacency matrix of subgraph G_s can be calculated as: $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. By feeding the subgraph G_s into a two-layer GCN, the corresponding subgraph representation $L \in \mathbb{R}^{n \times c}$ with n nodes can be calculated by:

$$L = \sigma(\tilde{A}\sigma(\tilde{A}EW_0)W_1) \tag{1}$$

where σ represents a non-linear function, $W_0 \in \mathbb{R}^{d*v}$ and $W_1 \in \mathbb{R}^{d*c}$ are trainable parameters. To obtain a more compact graph representation, we further feed L into a fully-connected layer, producing a final graph representation $M \in \mathbb{R}^{d}$.

3.4 Knowledge-enhanced BiLSTM

Preliminary (Vanilla BiLSTM) Generally, two independent BiLSTM networks (denoted as BiLSTM^x and BiLSTM^p) are employed to encode the input sentence x and the target p, respectively. BiLSTM can capture the left and right context of each word in the input. In particular, for the t-th word w_t in the input sequence of the target, BiLSTM^p computes its forward hidden state $\overrightarrow{h}_t^p$ and backward hidden state $\overleftarrow{h}_t^p$. We concatenate both the forward and backward hidden states to form the final hidden state $h_t^p = [\overrightarrow{h}_t^p \oplus \overleftarrow{h}_t^p]$ for word w_t at the t-th position of the input target. After learning the contextual representation of the target, we learn a target-aware sentence representation H^s by initializing BiLSTM^x with the final hidden state of BiLSTM^p.

The background knowledge contained in external lexicons is the collection of facts that individuals are expected to know, and plays a crucial role in reading comprehension. We propose a knowledge-enhanced BiLSTM (KE-BiLSTM) model, which incorporates the external background knowledge contained in the semantic-emotion knowledge graph into the BiLSTMs via a novel knowledge-aware memory unit (KAMU). KE-BiLSTM helps to identify discriminative semantic and emotion knowledge from the input text. It is motivated by two considerations:

- The external commonsense knowledge provides rich information of entities and relations between them, and highlights the features that are essential for stance detection. For example, with the external semantic lexicon, we can correctly understand the unusual word "zugzwang" through the semantically related words "chess", "strategy", "forced" contained in the semantic lexicon. Hence, we devise KE-BiLSTM to effectively leverage the graph embedding of SE-graph and fully explore the external knowledge from both word-level and concept-level.

- There exist dynamic interaction patterns and complementarity between the context and the external knowledge within the input sequence for stance detection. Instead of leveraging only the input context in each BiLSTM unit, we take external commonsense knowledge into consideration by adding a novel knowledge-aware memory unit to the BiLSTM, which dynamically controls the amount of external knowledge at each encoding step and thus balances the contextual and knowledge information for stance detection.

As illustrated in Figure 2, KE-BiLSTM consists of two primary parts: a BiLSTM network (depicted in blue) and a knowledge-aware memory unit (depicted in green). Similar to the standard BiLSTM

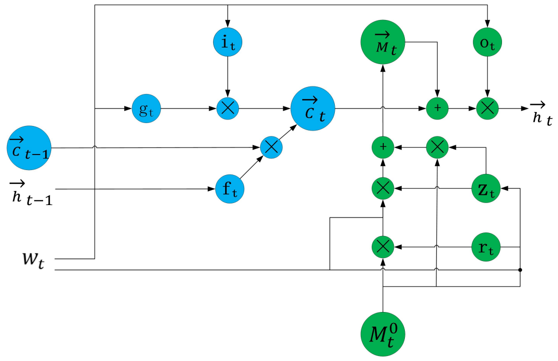

Figure 2: The structure of the knowledge-enhanced BiLSTM unit.

network, KE-BiLSTM also computes forward and backward hidden sequences, which are then combined to form the output representation. Due to limited space, we solely introduce the implementation details of the forward layer. The forward and backward knowledge-enhance LSTMs can be computed in a similar way.

In KE-BiLSTM, the BiLSTM network learns the sequential features of the input text. Formally, in the forward layer of BiLSTM, the input gate i_t, forget gate f_t, output gate g_t, and the memory cell $\overrightarrow{C}_t$ are updated as:

$$i_t = \sigma(W_i w_t + U_i \overrightarrow{h}_{t-1} + V_i \overrightarrow{C}_{t-1}) \qquad (2)$$

$$f_t = \sigma(W_f w_t + U_f \overrightarrow{h}_{t-1} + V_f \overrightarrow{C}_{t-1}) \qquad (3)$$

$$g_t = tanh(W_g w_t + U_g \overrightarrow{h}_{t-1} + V_g \overrightarrow{C}_{t-1}) \qquad (4)$$

$$\overrightarrow{C}_t = f_t \odot \overrightarrow{C}_{t-1} + i_t \odot g_t \qquad (5)$$

where σ represents the sigmoid function. W, U, and V are trainable parameters. w_t is the t-th word of the input text. $\overrightarrow{h}_{t-1}$ is the hidden state for the $t-1$-th word.

We propose a knowledge-aware memory component to incorporate the external knowledge into BiLSTM. For each word w_t, we extract the corresponding entity from SE-graph by performing n-gram matching and acquire a subgraph representation M_t^0. A new knowledge memory $\overrightarrow{M}_t$ at time t is computed with a linear interpolation between the previous M_t^0 and its candidate activation δ_t:

$$\overrightarrow{M}_t = z_t \odot M_t^0 + (1 - z_t) \odot \delta_t \qquad (6)$$

where $z_t \in [0, 1]$ is utilized to balance the importance of M_t^0 and δ_t, which can be computed by:

$$z_t = \sigma(W_z w_t + U_z M_t^0) \qquad (7)$$

where W_z and U_z are parameters to be learned. The candidate activation δ_t is updated as:

$$\delta_t = tanh(W_\delta w_t + U_\delta (r_t \odot M_t^0)) \qquad (8)$$

where W_δ and U_δ are parameters to be learned. r_t is the reset gate which aims to combine the knowledge in M_t^0 and w_t, which is defined as:

$$r_t = \sigma(W_r w_t + U_i M_t^0) \qquad (9)$$

where W_r and U_r are projection parameters.

Finally, the linear transformation of w_t, h_{t-1}, $\overrightarrow{M}_t$ and C_t are combined to calculate the output $\overrightarrow{o}_t$ of the forward KE-BiLSTM layer:

$$\overrightarrow{o}_t = \sigma(W_o w_t + U_o \overrightarrow{h}_{t-1} + V_o \overrightarrow{M}_t + Q_o \overrightarrow{C}_t) \qquad (10)$$

$$\overrightarrow{h}_t = o_t \odot tanh(\overrightarrow{C}_t + \overrightarrow{M}_t) \qquad (11)$$

where $\overrightarrow{o}_t$ and $\overrightarrow{h}_t$ denote the output gate and the hidden state of the forward network of KE-BiLSTM unit at time step t. The hidden state $\overleftarrow{h}_t$ of the backward network at time step t can be computed in a same way. We can get the overall hidden state $h_t = [\overrightarrow{h}_t \oplus \overleftarrow{h}_t]$ for word w_t.

Finally, we can use KE-BiLSTM to learn knowledge-enhanced sentence representation $H^s = \{h_1^s, \dots, h_n^s\}$ and knowledge-enhanced target representation $H^p = \{h_1^p, \dots, h_m^p\}$, where n and m denote the lengths of sentence x and given target p, respectively.

3.5 Stance Detection

We employ an attention mechanism to characterize the effect of the target on enforcing our SEKT model to pay more attention to the important words of the context. In particular, we use the target representation H^p as the attention source to calculate the attention weight α_t for the t-th word:

$$\alpha_t = softmax(\bar{h^p}^{\mathrm{T}} h_t^x) \qquad (12)$$

where $\bar{h^p}$ denote the average vector of target representation H^p. We can learn the attentive sentence representation emb by congregating the embeddings of hidden states H^s with attention vector α:

$$emb = \sum_{t=1}^{n} \alpha_t h_t^x \qquad (13)$$

Target	Favor/Against/None	Avg-length
DT	148/299/260	17.1
HC	163/565/256	17.0
FM	268/ 511/170	18.4
LA	167/544/222	19.0
TP	333/452/460	33.3

Table 1: The statistics of our experimental data extended from SemEval-2016 Task 6.

Finally, the sentence representation emb is fed into a fully-connected layer followed by a softmax layer to compute a stance probability distribution:

$$\hat{y} = softmax(W_y emb + b_y) \tag{14}$$

where W_y is a projection parameter and b_y is a bias term. $\hat{y}$ denotes the predicted stance probability for the input sentence x and target p. Given an annotated training set X^s, we utilize the cross-entropy between the predicted stance $\hat{y}$ and the ground-truth stance y as our loss function for stance detection:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} \log \hat{y}_{ij} \tag{15}$$

where N represents the number of instances in the training set. C denotes the number of possible stance categories. y_i represents the one-hot represented ground-truth label for the i-th instance. $\hat{y}_i$ is the predicted stance probability vector. This model can be optimized with the standard gradient descent algorithm.

4 Experiments

4.1 Experimental Data

We extend the SemEval-2016 Task 6 dataset (denoted as SemEval-2016) to evaluate the performance of our SEKT model for cross-target stance detection. SemEval-2016 is the first stance detection dataset collected from Twitter, which contains 4870 stance-bearing tweets towards different targets. Each tweet is classified as "*favor*", "*against*" or "*none*". Following the previous work (Wei and Mao, 2019), we use the tweets from four targets, including *Donald Trump* (DT), *Hillary Clinton* (HC), *Legalization of Abortion* (LA), and *Feminist Movement* (FM). These targets are commonly utilized to evaluate the cross-target stance classification.

In addition to the four targets in SemEval-2016, we introduce an additional *Trade Policy* (TP) target as the fifth target, which is an incredibly hot topic nowadays. Specifically, 1245 tweets related to TP are collected and manually labeled as "*favor*", "*against*" and "*none*". The statistics of this expanded dataset are reported in Table 1.

Concerning the targets, the expanded dataset can be divided into two groups: *Women's Right* (FM, LA) and *American Politics* (HC, DT, TP). Thus, we constructed 8 cross-target stance detection tasks (*DT→HC, HC→DT, FM→LA, LA→FM, TP→HC, HC→TP, TP→DT, DT→TP*). Here, the left side of the arrow corresponds to the source target and the right side of the arrow denotes the destination target.

4.2 Evaluation Metrics

Two evaluation metrics are adopted to verify our SEKT model. First, following (Wei and Mao, 2019), we leverage the average F1-score as one evaluation metric (denoted as F_{avg}). Second, since the targets in the dataset are imbalanced, we also compute both the micro-averaged F1 (dominating large class) and macro-averaged F1 (dominating small class), and treat their average as another evaluation metric: $F1_m = (F1_{micro} + F1_{macro})/2$.

4.3 Implementation Details

In the experiments, we use the 300-dimensional word2vec pre-trained on English Google News corpus to initialize the word embeddings. Follow (Augenstein et al., 2016), the node features is pre-trained on unlabelled corpora. The hidden size of LSTM is set to 100. Dropout (dropout rate = 0.2) is used to avoid overfitting. The Adam optimizer is applied to train the model, with the mini-batch size of 8 and the learning rate of 0.001.

4.4 Baseline Methods

We evaluate and compare our model with several strong baselines, which are described as follows:

- **BiLSTM**: This method uses BiLSTM to encode the sentence and target separately. The hidden states from both directions are combined to infer the stance label.

- **BiCond** (Augenstein et al., 2016): This method is similar to BiLSTM but uses a conditional encoding method that learns a target-dependent sentence representation for stance detection.

- **CrossNet** (Xu et al., 2018): This model is a variant of BiCond, which leverages a self-

Source-Target:	FM→LA	LA→FM	HC→DT	DT→HC	HC→TP	TP→HC	DT→TP	TP→DT
BiLSTM	0.448	0.412	0.298	0.358	0.291	0.395	0.311	0.341
BiCond	0.450	0.416	0.297	0.358	0.292	0.402	0.317	0.347
CrossNet	0.454	0.433	0.431	0.362	0.298	0.417	0.314	0.374
VTN	0.473	0.478	**0.479**	0.364	-	-	-	-
BERT	0.479	0.339	0.436	0.365	0.261	0.231	0.241	**0.456**
CrossNet-C	0.449	0.439	0.442	0.369	0.297	0.413	0.324	0.355
CrossNet-CF	0.467	0.457	0.457	0.396	0.307	0.411	0.377	0.398
CrossNet-CA	0.473	0.475	0.455	0.407	0.301	0.442	0.409	0.396
TextCNN-E	0.469	0.458	0.380	0.404	0.309	0.450	0.356	0.396
SEKT (Ours)	**0.536**	**0.513**	0.477	**0.420**	**0.335**	**0.460**	**0.444**	0.395

Table 2: Performance comparison of cross-target stance detection in terms of $F1_{avg}$ on 8 tasks.

Source-Target:	FM→LA	LA→FM	HC→DT	DT→HC	HC→TP	TP→HC	DT→TP	TP→DT
BiLSTM	0.401	0.379	0.433	0.401	0.236	0.418	0.207	0.389
BiCond	0.403	0.392	0.442	0.408	0.239	0.424	0.207	0.396
CrossNet	0.442	0.431	0.461	0.418	0.244	0.425	0.211	0.407
BERT	0.499	0.395	0.412	0.399	0.353	0.295	**0.391**	**0.478**
CrossNet-C	0.473	0.399	0.439	0.403	0.251	0.428	0.221	0.414
CrossNet-CF	0.497	0.438	0.434	0.404	0.280	0.437	0.302	0.428
CrossNet-CA	0.507	0.434	0.452	0.401	0.283	0.453	0.375	0.440
TextCNN-E	0.513	0.466	0.360	0.385	0.283	0.472	0.191	0.433
SEKT (Ours)	**0.523**	**0.510**	**0.463**	**0.432**	0.300	**0.489**	**0.391**	0.435

Table 3: Performance comparison of different models for cross-target stance detection.

attention layer to capture important words in the input text.

- **VTN** (Wei and Mao, 2019): The model utilizes the latent topics shared between the two targets as transferable knowledge for cross-target adaptation.

- **BERT** (Devlin et al., 2019): The method fine-tunes a pre-trained BERT model to perform cross-target detection. Specifically, we convert the given context and target to "[CLS] + target + [SEP] + context" structure for source and target domain, respectively.

We also extend CrossNet and TextCNN to incorporate external knowledge (SE-graph), resulting in stronger competitors.

- **CrossNet-C**: Similar to (Margatina et al., 2019), we extend the original CrossNet model by incorporating external knowledge. Here, three variants are considered, where CrossNet-C adopts the attentional concatenation, CrossNet-CF uses the feature-based gating mechanism, and CrossNet-CA adopts an attentional affine transformation.

- **TextCNN-E**: TextCNN (Kim, 2014) is an important baseline for text classification. Here, we extend TextCNN to the cross-target setting, denoted as TextCNN-E. Specifically, each word is represented as a 3D tensor by concatenating the embeddings of k semantically/emotionally-related words.

4.5 Overall Performance

We report the experimental results in terms of $\mathrm{F1}_{avg}$ and $\mathrm{F1}_{m}$ in Table 2 and Table 3, respectively. From the results, we can observe that BiLSTM has the worst performance because BiLSTM neither exploits the target information nor considers knowledge transfer for the cross-target stance detection. BiCond performs slightly better than BiLSTM, since it explicitly encodes the target information. As an extension to BiCond by introducing the attention mechanism, CrossNet shows a marginal improvement (e.g., 13.4% on HC→DT for $\mathrm{F1}_{avg}$, 3.9% on LA→FM for $\mathrm{F1}_{m}$). This may be because that the attention mechanism can learn the informative stance-aware sentence representation. However, this knowledge transfer scheme is based on word-level information, which often suffers from the data scarcity problem. VTN, which is a concept-level knowledge transfer model, achieves the best performance among all the baseline methods. It is noteworthy that the performance of BERT is not stable. Promising results are achieved on FM→LA and HC→DT, but it performs unsatisfactorily on other tasks. The reason may be that BERT does not explicitly employ any knowledge transfer

	SEKT	w/o SE	w/o KAMU
FM→LA	0.536 (0.523)	0.461 (0.492)	0.471 (0.499)
LA→FM	0.513 (0.510)	0.443 (0.455)	0.475 (0.469)
HC→DT	0.477 (0.463)	0.449 (0.439)	0.449 (0.450)
DT→HC	0.420 (0.432)	0.400 (0.404)	0.411 (0.407)
HC→TP	0.335 (0.279)	0.314 (0.278)	0.321 (0.280)
TP→HC	0.460 (0.489)	0.448 (0.466)	0.453 (0.471)
DT→TP	0.444 (0.391)	0.407 (0.371)	0.411 (0.376)
TP→DT	0.395 (0.435)	0.394 (0.420)	0.395 (0.431)

Table 4: Ablation test results in terms of $F1_{avg}$ and $F1_m$ (in the parentheses) by discarding SE graph (w/o SE) and knowledge-aware memory unit (w/o KAMU).

strategy. The proposed SEKT method yields better performance than all the baselines in most of the tasks. For example, our method improves 5.7% on FM→LA, 3.5% on LA→FM, 5.5% on DT→HC over the best competitors in terms of F1$_{avg}$. The advantage of SEKT comes from its two characteristics: (i) we develop a GCN based model to fully exploit the external knowledge from both semantic and emotion lexicons; (ii) a knowledge-aware memory unit is proposed to better fuse the external knowledge.

We also compare our SEKT model with the competitors that also integrate the semantic-emotion knowledge graph with GCN, e.g., CrossNet-C, CrossNet-CF, CrossNet-CA and TextCNN-E. The results are demonstrated in Table 2 and Table 3. CrossNet-C produces the worst performance in general. The reason is that concatenating the external knowledge and context representation could make the external knowledge lost in the sentence encoding process. CrossNet-CF and CrossNet-CA perform better than CrossNet-C since they incorporate the external knowledge into the hidden layers of BiLSTM. As expected, SEKT achieves the best performance, which verifies the effectiveness of the KAMU model.

4.6 Ablation Study

To investigate the impact of each part on our SEKT model, we perform the ablation test by discarding SE graph knowledge (denoted as w/o SE) and knowledge-aware memory unit (denoted as w/o KAMU), respectively. Specifically, for the w/o SE model, the external knowledge is expressed by a weighted sum of the embeddings of four semantically/emotionally-related words. For the w/o KAMU model, we replace the KE-BiLSTM structure by the standard BiLSTM layer, and the external knowledge is combined in the hidden layer.

hop No.	**DT→HC**	**LA→FM**	**DT→TP**
1	0.401	0.489	0.431
2	0.417	**0.513**	**0.444**
3	**0.420**	0.479	0.424
4	0.374	0.369	0.408

Table 5: The experimental results with respect to varying number of hops in GCN.

The ablation results are summarized in Table 4. From the results, we observe that both the SE graph and KAMU make great improvements to our SEKT method. The external semantic and emotional knowledge can help SEKT to capture multi-hop semantic correlations between words or emotion tags. On the one hand, KAMU helps to fully incorporate the external knowledge into the BiLSTM network, which makes the representation learning model more general to new targets.

Number of Hops Based on our empirical observation, capturing the multi-hop semantic correlation is one of the most important parts for the overall performance of SEKT. Thus, we also investigate the impact of the number of hops used in GCN. In particular, we evaluate the performance of SEKT by varying the number of hops from 1 to 4 with a step size of 1. From Table 5, we can observe that the best results are achieved when the number of hops is 2 or 3. This is because GCN with a mediate hop number can capture semantic correlations between words while preventing from introducing unnecessary noises.

5 Error Analysis

To better understand the limitations of SEKT, we additionally carry out an analysis of the errors made by SEKT. Specifically, we randomly select 100 instances that are incorrectly predicted by SEKT from the expanded SemEval-2016 dataset. We revealed several reasons for the classification errors, which can be divided into the following categories. First, SEKT fails to classify some sentences that contain latent opinions or require deep comprehension. For example, for the sentence "I guess NBC does not like to hear the truth.$_{[favor]}$" with a target "Donald Trump", SEKT tends to predict an incorrect *against* stance. This is because the SEKT model cannot learn the implicit relationship between NBC[*] and TRUMP, which is not acquirable from the semantic-emotion lexicons. The

[*] National Broadcasting Company

second error category is caused by special hashtags with implicit meanings. For example, SEKT cannot correctly predict the stance for the sentence "The gift that keeps on giving. #makeitstop #SemST"$_{[against]}$. This may be because the information in the sentence is not sufficient enough such that SEKT cannot capture the sequential patterns of the stance-related words. It suggests that certain data augmentation strategy needs to be devised in the future so as to capture the sequential patterns between stance-related words from short texts.

6 Conclusion

In this paper, we proposed a semantic-emotion knowledge transferring (SEKT) model for cross-target stance classification, which used the external knowledge from semantic and emotion lexicons as commonsense knowledge to bridge the gap across different targets. Specifically, we first built a SE-graph from semantic and emotion lexicons, which leveraged external knowledge from both word-level and concept-level. Second, the GCN was employed to learn the graph representation that captured multi-hop semantic connections between words or emotion tags. Third, we extend the standard BiLSTM classifier to fully integrate the external knowledge by adding a novel knowledge-aware memory unit to the BiLSTM cell. The experimental results demonstrated that the SEKT model significantly outperformed the state-of-the-art methods for cross-target stance detection.

7 Acknowledgements

This research was supported in part by the National Key R&D Program of China, 2018YFB2101100, 2018YFB2101101 and NSFC under Grant Nos. U1836107, 61972111, 61572158 and 61602132. Min Yang was partially supported by National Natural Science Foundation of China (No. 61906185), Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515011705 and No. 2018A030313943), and the project AWS13C008.

References

I Augenstein, T Rocktaeschel, A Vlachos, and K Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Sheffield.

Erik Cambria, Soujanya Poria, Devamanyu Hazarika, and Kenneth Kwok. 2018. Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Kuntal Dey, Ritvik Shrivastava, and Saroj Kaushik. 2018. Topical stance detection for twitter: A two-phase lstm model using attention. In *European Conference on Information Retrieval*, pages 529–536. Springer.

Mauro Dragoni and Giulio Petrucci. 2018. A fuzzy-based strategy for multi-domain sentiment analysis. *International Journal of Approximate Reasoning*, 93:59–73.

Jiachen Du, Ruifeng Xu, Yulan He, and Lin Gui. 2017. Stance classification with target-specific neural attention networks. International Joint Conferences on Artificial Intelligence.

Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, and Eric Xing. 2016. Harnessing deep neural networks with logic rules. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2410–2420.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751.

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Zeyang Lei, Yujiu Yang, and Min Yang. 2018. SAAN: A sentiment-aware attention network for sentiment analysis. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 1197–1200. ACM.

Yukun Ma, Haiyun Peng, and Erik Cambria. 2018. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Katerina Margatina, Christos Baziotis, and Alexandros Potamianos. 2019. Attention-based conditioning methods for external knowledge integration. *arXiv preprint arXiv:1906.03674*.

Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3):436–465.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*, pages 226–234. Association for Computational Linguistics.

Qingying Sun, Zhongqing Wang, Qiaoming Zhu, and Guodong Zhou. 2018. Stance detection with hierarchical attention network. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2399–2409.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 conference on empirical methods in natural language processing*, pages 327–335. Association for Computational Linguistics.

Marilyn A Walker, Pranav Anand, Robert Abbott, and Ricky Grant. 2012. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 592–596. Association for Computational Linguistics.

Penghui Wei, Junjie Lin, and Wenji Mao. 2018. Multi-target stance detection via a dynamic memory-augmented network. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1229–1232. ACM.

Penghui Wei and Wenji Mao. 2019. Modeling transferable topics for cross-target stance detection. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1173–1176. ACM.

Chang Xu, Cecile Paris, Surya Nepal, and Ross Sparks. 2018. Cross-target stance classification with self-attention networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 778–783.

Bowen Zhang, Xiaofei Xu, Min Yang, Xiaojun Chen, and Yunming Ye. 2018. Cross-domain sentiment classification by capsule network with semantic rules. *IEEE Access*, 6:58284–58294.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019a. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv preprint arXiv:1909.03477*.

Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. 2019b. Integrating semantic knowledge to tackle zero-shot text classification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1031–1040.