Department: Affective Computing and Sentiment Analysis
Editor: Erik Cambria, Nanyang Technological University

Commonsense Knowledge Enhanced Memory Network for Stance Classification

Jiachen Du
Harbin Institute of Technology (Shenzhen)

Lin Gui
University of Warwick

Ruifeng Xu
Harbin Institute of Technology (Shenzhen)

Yunqing Xia
Search Technology Center, Microsoft

Xuan Wang
Harbin Institute of Technology (Shenzhen)

***Abstract*—Stance classification aims at identifying in the text the attitude towards the given targets as favorable, negative or unrelated. In existing models for stance classification, only textual representation is leveraged, while commonsense knowledge is ignored. In order to better incorporate commonsense knowledge into stance classification, we propose a novel model named Commonsense Knowledge Enhanced Memory Network which jointly represents textual and commonsense knowledge representation of given target and text. The textual memory module in our model treats the textual representation as memory vectors, and uses attention mechanism to embody the important parts. For commonsense knowledge memory module, we jointly leverage the entity and relation embeddings learned by TransE model to take full advantage of constraints of the knowledge graph. Experimental results on the SemEval dataset show that the combination of the commonsense knowledge memory and textual memory can improve stance classification. Moreover, the visualization of learned representation empirically confirms that the knowledge commonsense extracted by our model can benefit the identification of the stance towards given targets.**

Stance classification is the task of automatically identifying the attitude, such as *favor (positive, pro)*, *against (negative, con)* or *none (unrelated)*, conveyed in the text towards a specific target. Unlike conventional sentiment classification [1], [2], [3], the main challenge of stance classification is that the target relating content may not appear explicitly in the text. Furthermore, it is difficult for models to infer the relationship between the discussed object in the text and in the target. For example, if the target and the discussed entity in the text are from opposite standpoints, the stance polarity can be inconsistent with the sentiment expressed by the text with high probability. As depicted in Figure 1, the text expresses against attitude with respect to the given target "*legalization of abortion*". However, without any commonsense knowledge about the relation between "*abortion*" and "*pro-life*", it is difficult for a pure textual feature based classifier to predict stance polarity.

Previous models for stance classification only mainly leverage the linguistic context to capture the relation between the text and the target in predicting the stance [4], [5]. However, the commonsense knowledge beyond the context has been underutilized, which hinders the performance of current stance classification methods. To address this problem, we present a stance classification model that resembles how we human beings solve the problem: leveraging commonsense knowledge. Such commonsense knowledge is usually not explicitly stated in the text. However, it is very critical for conducting accurate stance prediction in some cases illustrated in Figure 1.

Commonsense knowledge generally refers to the factual knowledge that might not be explicitly available in text [6] but structurally stored in external commonsense knowledge bases (CKBs) such as DBpedia [7], FreeBase [8] ConceptNet [9], and SenticNet [10]. It is vital for a broad range of natural language processing (NLP) tasks such as dialogue modeling [11], short text classification [12], textual reasoning [13] and sentiment analysis [14] to incorporate commonsense knowledge for better performance and interpretability. In previous CKB-based models, the knowledge is usually introduced incorporated by attaching the words in entities as "pseudo words" to the text sequence [12] or appending the entity embeddings to the word vectors on the token level [13]. We argue that entities in CKB and words in text are not always consistent in the same common feature space. Previous methods cannot take advantage of complementary semantics of text and CKB, and they are not suitable for knowledge-sensitive NLP tasks such as stance classification.

In order to make use of the distinct characteristics of text and CKB, we proposed a novel model based on a memory network [15], which separately memorizes text and commonsense knowledge in different components. In addition, we propose the attention mechanism to identify the most possible entities and relations associated with the target from the CKB. To be specific, to memorize CKB representation, our model uses an ordinary end-to-end memory module to store textual representation and a key-value memory module. From CKB representation, the entities and relations mentioned in the text can be used to infer the subject in the text. Then, the stance predicted by the textual representation will affect the weights in key-value memory module which assigns the most possible subject a higher attention. Unlike *dot product* attention used in textual memory module, the attention used in CKB memory module takes advantage of the additional property of knowledge graph embedding to capture the correlation between text and the target in the *knowledge space*, which is complementary to the linguistic feature space.

RELATED WORK

Stance Classification: Recently, there is a growing interest in detecting the stance polarity of text on microblogs. Unlike ordinary sentence-level [16], [17] and aspect-level [18], [19] sentiment classification, stance classification is a more challenging task. SemEval-2016 Task 6 [20] involves two stance detection subtasks in tweets in supervised and weakly supervised settings. Augenstein et al. [4] used two bidirectional recurrent neural network (RNN) to model both target and text for stance detection. However this model requires a very large unlabeled Twitter corpus in order to predict the task-relevant hashtags as an auxiliary task to initialize the word embeddings. Du et al. [5] proposed an RNN based model, which incorporates target-specific infor-

Target: *Legalization of Abortion*
Text: *I just unfollowed someone because they tweeted about being **pro-life**.*
Stance: Against (Negative)

Figure 1. An example of stance classification.

mation into stance classification by using a new attention mechanism. Li et al. [21] introduced deep memory networks for stance detection (they called *attitude identification*), which employs attention mechanism to capture the informative context words by leveraging external memory components. In order to mine the relationship between targets, multi-target stance detection [22], [23] has gained increasing attention and emphasis. Unlike the existing models, which attempt to predict the stance label by using only textual feature, we leverage external CKB to construct a commonsense knowledge memory module to improve the performance of stance classification.

Commonsense Knowledge: With rapid growth of knowledge engineering, several CKBs have been published, such as DBpedia, FreeBase, ConceptNet and SenticNet. In our model, ConceptNet is used to enhance the capacity for modeling the relation between the context and target. ConceptNet is a knowledge representation project, providing a large-scale semantic graph that describes facts and human knowledge. CKBs have been widely used in various NLP tasks, such as open-domain conversation generation, visual question answering and sentiment analysis. For conversation generation, there are several end-to-end conversation models levering CKB [11], [24], which improve the relevance and diversity of generated responses in open-domain conversations. For visual question answering, Su et al. [25] proposed visual knowledge memory networks to leverage self-built CKB for supporting visual question answering. For sentiment analysis, Xu et al. [26] modified ordinary recall gate function in RNN to leverage CKB. For sentiment analysis, Ma et al. [27] integrated integrated external CKB into RNN cell to improve the performance on aspect sentiment classification .

PROPOSED MODEL

We formalize the problem of stance classification as follows. Suppose that the text is a sequence of words $x = \{x_1, x_2, \ldots, x_n\}$, the goal of our model is to predict the stance polarity $y \in \{-1, 0, 1\}$ (corresponding to *Against*, *None* and *Favor* respectively) towards the target $z = \{z_1, \ldots, z_t\}$, where the words in the text and the target are from a global vocabulary. As illustrated in Figure 2, we propose a Commonsense Knowledge Enhanced Memory Network (CKEMN) for stance classification. The CKEMN model consists of two memory modules: (1) the textual memory module, representing the text and the target by computing the attention-weighted sum of word level memory representations; (2) the commonsense knowledge memory module, using key-value neural memories to store the commonsense knowledge representations of the text and target respectively, then applying attention mechanism to extract the related knowledge representation for stance classification. Finally, the textual and commonsense knowledge representations are concatenated and fed into the stance classifier for stance prediction.

Textual Memory Module

The textual memory module, shown in the right part in Figure 2, aims to obtain the textual memory representations of the text and the target. This module is composed of three parts: textual encoder, textual target encoder, and textual output component.

Textual Encoder: Let $E_w \in \mathbb{R}^{d_w \times |V|}$ denote the word embedding lookup matrix initialized by Glove or Word2vec, where $|V|$ is the vocabulary size, d_w is the dimension of word embedding. The one-hot representation of a word $x_i \in \mathbb{R}^{|V|}$ is converted to its embedding vector by $v_i^w = E_w {x_i}^\intercal$. Inspired by recurrent attention on memory framework proposed by Chen et al. [28], we use bidirectional recurrent networks with long short-term memory (LSTM) units as the textual

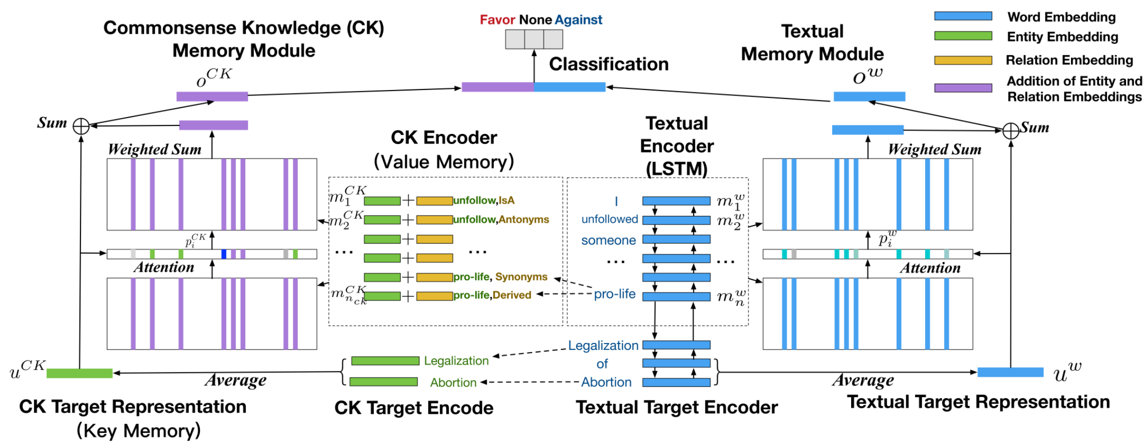

Figure 2. Architecture of Commonsense Knowledge Enhanced Memory Network.

encoder to capture the long-term dependency from both forward and backward directions. The hidden state $\overrightarrow{h_i}$ at step i is used as the representation of word i. The backward LSTM is same as forward LSTM, except that it receives the reversed sequence of words and emits the hidden state $\overleftarrow{h_i}$. The forward and backward hidden states are concatenated to construct the textual memory slot of word i by $\boldsymbol{m}_i^W = \left[\overrightarrow{h_i}, \overleftarrow{h_i}\right] \in \mathbb{R}^{2d_h}$.

Textual Target Encoder: In order to represent the target words in the same vector space of the text, the identical bidirectional LSTM (bi-LSTM) is applied to represent the target. Inspired by bidirectional conditional encoding method proposed by Augenstei et al [4], the bi-LSTM here is initialized by the last hidden states of the bi-LSTM in the textual encoder. The target is represented by taking the average of hidden states of the words in target $u^w = \frac{1}{t}\sum_{j=1}^{t} m_j^t$, where $u^w \in \mathbb{R}^{d_h}$ is the text memory vector of the target, $m_j^t = \left[\overrightarrow{h_j}, \overleftarrow{h_j}\right]$ is the concatenation of the forward and backward hidden states, and t is the target length.

Textual Output Component: As shown in Figure 2, for the feature vector fed into the stance classifier, its right half, i.e., o^w, is computed as the sum of the textual memory vector of the target (i.e., u^w) and the weighted sum of the memory slots of the text $o^w = u^w + \sum_i p_i^w m_i^w$, where the weight p_i^w of m_i^w is the attention score showing the importance of x_i, computed as $p_i^w = \mathrm{Softmax}((u^w)^\intercal m_i^w)$.

Commonsense Knowledge Memory Module

The CKB used in our model is consists of a large number of factual triples $f \in F$. Each fact takes the form of triple like $< e_1, r, e_2 >$, e.g., *<anti-abortion, synonyms, pro-life>*, in which e_1, e_2 are entities or concepts (e_1 is *anti-abortion* and e_2 is *pro-life*) and r is the relationship between the two entities (r is *synonyms*). The goal of commonsense knowledge memory module is to convert the structured commonsense knowledge into key-value based memory representations and extract the most highlighted facts for stance classification. Specifically, the representation of the target is treated as a *key vector* for searching the most related commonsense knowledge memory representation of the text (*i.e., the value vector*) by attention mechanism.

Commonsense Knowledge Encoder: We first retrieve all of entities appeared in text as input to commonsense knowledge memory module by an entity linking tool built on Concept. Extracted entities are denoted as one-hot entity vectors $x^e = \{x_1^e, x_2^e, \ldots, x_{n_e}^e\}, x_i^e \in \mathbb{R}^{|E|}$ for the text , and $z^e = \{z_1^e, z_2^e, \ldots, z_{t_e}^e\}, z_i^e \in \mathbb{R}^{|E|}$ for the target, where n_e and t_e are the number of entities mentioned in text and target. Besides the mentioned entities, the corresponding relations connecting these entities are also retrieved from CKB.

We adopt TransE [29] to represent entities and relations in a CKB as low-dimension vectors. For a triple $< e_1, r, e_2 >$ in CKB, the goal of TransE is to minimize the distance between $e_1 + r$ and e_2 in vector space, which assumes $e_1 + r \approx e_2$. Let

$E_{CK} \in \mathbb{R}^{d_{CK} \times (|E|+|R|)}$ denote the entity and relation embedding matrix pre-trained on the whole set of ConceptNet 5.5, where $|E|$ is the number of CKB entities, $|R|$ is the number of relations in CKB, d_{CK} is the dimension of commonsense knowledge graph embedding. The one-hot representation of the extracted entity $x_i^e \in \mathbb{R}^{|E|}$ in the text is converted into its entity vector e_i by $e_i = E_{CK}(x_i^e)^\intercal$. where $e_i \in \mathbb{R}^{d_{CK}}$. Similarly, the relations connecting entities x_i^e are convert to vectors $\{r_1^i, \dots, r_{|r_i|}^i\}$ by looking up the same embedding matrix E_{CK}. For an entity x_i^e , the vectors of relations connecting to entity x_i^e are $r^i = \{r_i^1, r_i^2, \dots, r_i^{|r_i|}\}, r_i^j \in \mathbb{R}^{d_{CK}}$, where $|r_i|$ is the number of connected relations to entity x_i^{CK}. The commonsense knowledge memory slots of entity x_i^e are $\{m_1^{CK}, m_2^{CK}, \dots, m_{|r_i|}^{CK}\}$ (index i omitted) obtained by taking the sum of the entity embedding e_i and the corresponding relation vectors by

$$m_j^{CK} = e_i + r_i^j \tag{1}$$

where $j \in 1, 2, \dots |r_i|$ is the index of relations connecting to entity x_i^c. Note that the total number of commonsense knowledge memory slots of text is $\sum_i^{n_e} |r_i|$. We use the average of entity embeddings $z^e = \{z_1^e, z_2^e, \dots, z_{t_e}^e\}$ as the representation of target $u^{CK} = \frac{1}{t_e}\sum_{j=1}^{t_e} z_j^e$, where z_j^e is the entity embedding of j-th entity mentioned in target.

Commonsense Output Component: The additive property of TransE embedding ($e_1 + r \approx e_2$) allows us to use entity embedding of target as *key* (*query*) to search the most related facts appeared in text (*value*). We use a neural attention function to compute the relatedness score of entities in target and entity-relation tuples in text. The commonsense knowledge output memory representation is obtained by computing the weighted sum of commonsense knowledge memory vectors of text and target representation u^e by

$$\begin{aligned} p_i^{CK} &= \text{Softmax}((u^{CK})^\intercal m_i^{CK}) \\ o^{CK} &= u^{CK} + \sum_i p_i^{CK} m_i^{CK} \end{aligned} \tag{2}$$

where p_i^{CK} is the attention score of commonsense knowledge fact i, and o^{CK} is the output representation of commonsense knowledge memory module.

Stance Classifier

In order to jointly leverage the textual and commonsense knowledge representation of the text and the target, the output of both memory modules are concatenated to obtain the combined representation and then fed to stance classifier:

$$p(\hat{y}|x, z) = \text{softmax}(W_p(o^w \oplus o^{CK}) + b_p) \tag{3}$$

where $o^w \oplus o^{CK} \in \mathbb{R}^{d_{KB}+2d_h}$ is the concatenated representation of text with entities, $o^w \in \mathbb{R}^{2d_h}$ and $o^{KB} \in \mathbb{R}^{d_{CK}}$ are output representation representations of textual and commonsense knowledge memory modules as depicted above, W_p is the weight of stance classifier, b_p is a bias term, and $p(\hat{y}|x, z)$ is the predicted probability of stance polarity. We use the cross-entropy between the predicted and ground-truth labels as the loss function of our model. All components can be trained end-to-end by minimizing the loss function.

EXPERIMENTS

Dataset and Commonsense Knowledge: Semeval-2016 Task 6 [20] released a dataset for stance classification on English tweets. IIn total, there are 4,163 tweets in this dataset, and the stance of each tweet is manually annotated as favorable or unfavorable toward on of five targets *Atheism*, *Climate Change*, *Feminist Movement*, *Hillary Clinton*, and *Legalization of Abortion*. This dataset has two subtasks, including subtask-A supervised learning and subtask-B unsupervised learning. In this evaluation, we merely work on the subtask-A, in which the targets provided in the test set can all be found in the training set.

ConceptNet is used as the CKB in our proposed model, which contains 1.5 million entities and 18.1 million relations. The knowledge in ConceptNet is organized as entity-relation triples. The number of retrieved knowledge triples is 17,426 (containing 8,052 entities).

Metrics: The micro average of $F1$-score across targets, which is the official evaluation measure for Semeval-2016 Task 6, is adopted as the evaluation metric for stance classification. First, the $F1$-score is calculated for *Favor* and *Against* categories for all instances in the dataset. Then, the average of F_{Favor} and $F_{Against}$ is calculated as the final metric. Note that the final metric does not consider the *None* class. By

Table 1. Performance comparison on the SemEval Dataset. SVM and TAN train separate model for each target

Models	Semeval English Dataset					
	Atheism	Climate	Feminist	Hillary	Abortion	**Overall**
CNN	52.18	36.70	45.80	56.26	54.30	62.55
LSTM	58.18	40.05	49.06	61.84	51.03	63.21
SVM [30]*	59.48	52.51	41.07	60.79	**64.20**	67.86
BCD [4]	61.47	41.63	48.94	57.67	57.28	67.82
MM [21]	60.55	53.07	**53.58**	62.94	68.05	67.09
TAN [5]*	59.33	**53.59**	55.77	65.38	68.79	68.79
CKEMN (no TMN)	57.78	48.52	58.86	59.81	45.81	65.53
CKEMN (no CMN)	60.33	50.94	57.11	**65.75**	62.19	67.74
CKEMN	**62.69**	53.52	61.25	64.19	64.19	**69.74**

taking the average F-score for only the *Favor* and *Against* classes, we treat *None* as a class that is not of interest.

ANALYSIS

We first analyze the results on the experimental data. Then, the word attention and selected commonsense knowledge learned by our model are visualized. Finally, we demonstrate the learned representations of text and commonsense knowledge.

Main Result

The experimental results of the baselines and our proposed model on SemEval dataset are reported in Table 1. First, it is observed that support vector machine (SVM) [30] performs better than convolutional neural network (CNN) and LSTM, since SVM trains a separate classifier for each target. However, this training strategy is not capable of classifying stance when there is no explicit targets. Bidirectional conditional encoding (BCD) [4], which employs conditional LSTM to learn a representation of the tweet considering the target, also outperforms CNN and LSTM. It is also observed that target-specific attention network (TAN) [5] outperforms other baseline models since it has the capability of capturing the target information to improve the performance of stance detection. Our CKEMN model outperforms all competing baseline methods significantly. This verifies that combining textual representation and CKB can benefit stance classification.

The variants of the proposed CKEMN model, which remove the Textual Memory Network (*TMN*) and the commonsense knowledge memory network (*CKEMN*) respectively, also show satisfactory results. It can be found that when removing the textual or commonsense memory modules, the performance drops dramatically. More importantly, it is empirically found observed that TMN has more effects on performance than CKEMN, since the textual representation is more suitable for capturing the similarity between context and target.It is also shown that the representations of the textual text and commonsense knowledge can be complementary to each other, and combining them can further improve the performance of stance classification. information of the language.

Case Study

To make it more intuitive, we randomly select an example from the test set, and show the word and knowledge graph (KG) attention scores obtained by textual memory and KG memory modules respectively in Figure 3. For KG attention, the word-level attention score is obtained by taking the average of all entity-relation pairs corresponding to a word. It is observed that the word attention focuses on "*unfollowed*" that is a word expressing emotion of the author, but it ignores the relation between the word "*pro-life*" in text and the target. Unlike the word attention, KG attention focuses on word "*pro-life*", which is more contextually related to the target in the external KG. This exemplary case indicates that the proposed KG memory module improves the performance of stance classification mainly by capturing the complementary knowledge rather than the superficial

CONCLUSION

In this paper, we proposed a novel memory network-based model which combines the textual representation and the corresponding commonsense knowledge representation for stance classification. The main contribution of this model lies in that it represents the

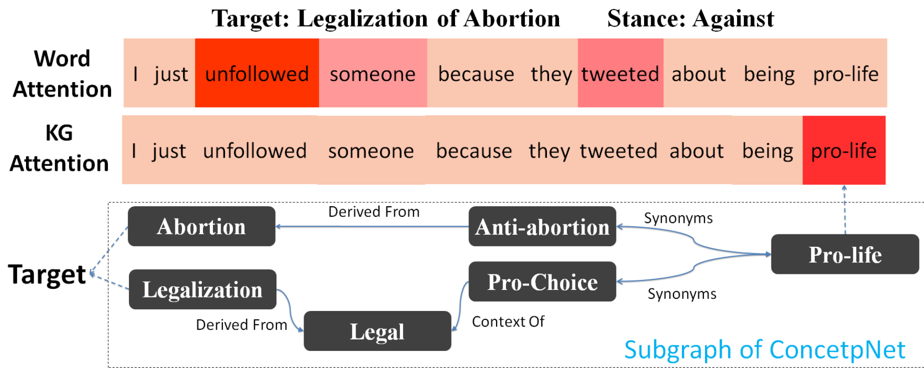

Figure 3. Example of attention visualization.

commonsense knowledge as memory vectors for stance classification. Specifically, a neural memory module is proposed to make use of the additional property of knowledge graph embeddings to better represent the structural knowledge. Experimental results show that our model outperforms several strong baselines.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (No. 61632011 and 61876053), Key Technologies Research and Development Program of Shenzhen (No. JSGG20170817140856618), Shenzhen Foundational Research Funding (JCYJ20180507183527919 and JCYJ20180507183608379) and EU-H2020 (grant no. 794196).

REFERENCES

1. I. Chaturvedi et al., “Distinguishing between facts and opinions for sentiment analysis: Survey and challenges,” *Information Fusion*, vol. 44, 2018, pp. 65–77.
2. Y. Xia et al., “Word Polarity Disambiguation Using Bayesian Model and Opinion-Level Features,” *Cognitive Computation*, vol. 7, no. 3, 2015, pp. 369–380.
3. E. Cambria et al., “Sentic Medoids: Organizing Affective Common Sense Knowledge in a Multi-Dimensional Vector Space,” LNCS, vol. 6677Springer, 2011, pp. 601–610.
4. I. Augenstein et al., “Stance detection with bidirectional conditional encoding,” *Proceedings of EMNLP*, 2016, pp. 876–885.
5. J. Du et al., “Stance Classification with Target-Specific Neural Attention Networks,” *Proceedings of IJCAI*, 2017, pp. 3988–3994.
6. E. Cambria et al., “Common Sense Computing: From the Society of Mind to Digital Intuition and Beyond,” LNCS, vol. 5707Springer, 2009, pp. 252–259.
7. S. Auer et al., “Dbpedia: A nucleus for a web of open data,” The semantic webSpringer, 2007, pp. 722–735.
8. K. Bollacker et al., “Freebase: a collaboratively created graph database for structuring human knowledge,” *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, 2008, pp. 1247–1250.
9. R. Speer, J. Chin, and C. Havasi, “ConceptNet 5.5: An Open Multilingual Graph of General Knowledge.” *Proceedings of AAAI*, 2017, pp. 4444–4451.
10. E. Cambria et al., “SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings,” *AAAI*, 2018, pp. 1795–1802.
11. T. Young et al., “Augmenting End-to-End Dialog Systems with Commonsense Knowledge,” *Proceedings of AAAI*, 2018, pp. 4970–4977.
12. J. Wang et al., “Combining knowledge with deep convolutional neural networks for short text classification,” *Proceedings of IJCAI*, 2017, pp. 2915–2921.
13. T. Botschen, D. Sorokin, and I. Gurevych, “Frame-and Entity-Based Knowledge for Common-Sense Argumentative Reasoning,” *Proceedings of the 5th Workshop on Argument Mining*, 2018, pp. 90–96.
14. N. Ofek et al., “Unsupervised Commonsense Knowledge Enrichment for Domain-Specific Sentiment Analysis,” *Cognitive Computation*, vol. 8, no. 3, 2016, pp. 467–477.
15. S. Sukhbaatar et al., “End-to-end memory networks,” *Proceedings of NIPS*, 2015, pp. 2440–2448.
16. Y. Li et al., “Learning word representations for sentiment

analysis," *Cognitive Computation*, vol. 9, no. 6, 2017, pp. 843–851.

17. I. Chaturvedi et al., "Bayesian network based extreme learning machine for subjectivity detection," *Journal of The Franklin Institute*, vol. 355, no. 4, 2018, pp. 1780–1797.
18. N. Majumder et al., "IARM: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis," *Proceedings of EMNLP*, 2018, pp. 3402–3411.
19. H. Peng et al., "Learning multi-grained aspect target sequence for Chinese sentiment analysis," *Knowledge-Based Systems*, vol. 148, 2018, pp. 167–176.
20. S. M. Mohammad et al., "Semeval-2016 task 6: Detecting stance in tweets," *Proceedings of the 9th Semeval*, 2016, pp. 31–41.
21. C. Li, X. Guo, and Q. Mei, "Deep Memory Networks for Attitude Identification," *Proceedings of the 2017 WSDM*, 2017, pp. 671–680.
22. P. Sobhani, D. Inkpen, and X. Zhu, "A dataset for multi-target stance detection," *Proceedings of EACL, vol. 2*, 2017, pp. 551–557.
23. C. Xu et al., "Cross-Target Stance Classification with Self-Attention Networks," *Proceedings of ACL*, 2018, pp. 778 – 783.
24. H. Zhou et al., "Commonsense Knowledge Aware Conversation Generation with Graph Attention," *Proceedings of IJCAI*, 2018, pp. 4623–4629.
25. Z. Su et al., "Learning Visual Knowledge Memory Networks for Visual Question Answering," *Proceedings of CVPR*, 2018, pp. 7736–7745.
26. Z. Xu et al., "Incorporating loose-structured knowledge into conversation modeling via recall-gate LSTM," *Proceedings of IJCNN*, 2017, pp. 3506–3513.
27. Y. Ma, H. Peng, and E. Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM," *Proceedings of AAAI*, 2018, pp. 5876–5882.
28. P. Chen et al., "Recurrent attention network on memory for aspect sentiment analysis," *Proceedings of EMNLP*, 2017, pp. 452–461.
29. A. Bordes et al., "Translating embeddings for modeling multi-relational data," *Proceedings of NIPS*, 2013, pp. 2787–2795.
30. S. M. Mohammad, P. Sobhani, and S. Kiritchenko, "Stance and sentiment in tweets," *ACM Transactions on Internet Technology*, vol. 17, no. 3, 2017.

Jiachen Du is a PhD student at Harbin Institute of Technology. His research interests include natural language processing and sentiment analysis. Contact him at dujiachen@stmail.hitsz.edu.cn.

Lin Gui is a Marie Sklodowska-Curie Individual Fellow with the University of Warwick, U.K. He received the Ph.D. degree from the Harbin Institute of Technology, Shenzhen, China. His research interests include natural language processing, and sentiment analysis. Contact him at lin.gui@warwick.ac.uk.

Ruifeng Xu is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. He received the Ph.D. degree in computer science from The Hong Kong Polytechnic University. He has published more than 100 papers in natural language processing, sentiment analysis, and social media analysis. Ruifeng Xu is the corresponding author of this work. Contact him at xuruifeng@hit.edu.cn.

Yunqing Xia is principal applied scientist with Search Technology Center, Microsoft, Beijing, China. He received the Ph.D. degree in computer science from Institute of Computing Technology, CAS, China. His research interests include sentiment analysis, text classification and clustering. Yunqing Xia is the corresponding author of this work. Contact him at yxia@microsoft.com

Xuan Wang is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China. He received the Ph.D. degree in computer science from the Harbin Institute of Technology, China. His research interests include cybernetic security and artificial intelligence. Contact him at wangxuan@cs.hitsz.edu.cn