

Exploiting Sentiment and Common Sense for Zero-shot Stance Detection

Yun Luo^{1,2}, Zihan Liu², Yuefeng Shi², Stan Z. Li², Yue Zhang^{2,3}

¹ School of Computer Science And Technology, Zhejiang University, Hangzhou, 310024, P.R. China.

² School of Engineering, Westlake University, Hangzhou, 310024, P.R. China.

³ Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou, 310024, P.R. China.

{luoyun, guofang, liuzihan, stan.zq.li, zhangyue}@westlake.edu.cn

Abstract

The stance detection task aims to classify the stance toward given documents and topics. Since the topics can be implicit in documents and unseen in training data for zero-shot settings, we propose to boost the transferability of the stance detection model by using sentiment and commonsense knowledge, which are seldom considered in previous studies. Our model includes a graph autoencoder module to obtain commonsense knowledge and a stance detection module with sentiment and commonsense. Experimental results show that our model outperforms the state-of-the-art methods on the zero-shot and few-shot benchmark dataset–VAST. Meanwhile, ablation studies prove the significance of each module in our model. Analysis of the relations between sentiment, common sense, and stance indicates the effectiveness of sentiment and common sense.

1 Introduction

Stance detection aims to identify the authors' attitudes or positions (*Pro (support)*, *Con (oppose)*, *Neu (neutral)*) towards a specific target such as an entity, a topic. (Mohammad et al., 2017, 2016; Walker et al., 2012; Qiu et al., 2015; Zhang et al., 2017). It is crucial for understanding opinions and analyzing how opinions are presented in texts regarding specific issues, and much work has been done building stance detection models (Wei et al., 2016; Dias and Becker, 2016; Allaway and Mckeown, 2020). There are two salient challenges to the task. First, obtaining rich annotated data in stance detection is time-consuming and labor-intensive. To address this issue, Allaway and Mckeown (2020) propose the dataset VAST containing various topics for few-shot and zero-shot stance detection tasks, requiring the model to classify the stance of topics unseen in the training set. Second, the topic is often not explicitly mentioned in the document, resulting in difficulty. Considering

Example 1 Topic : Olympics	Stance : Pro
Text : The games should proceed. Athletes have made tremendous sacrifices to qualify and be prepared. It would be cruel to deny them their chance. In the future the Games should be held in countries within the top say 15 GDP per capita.	
Example 2 Topic : Nuclear power	Stance : Pro
Text : I totally agree with this premise. As a younger person I was against Nuclear power (I was in college during 3 mile island) but now it seems that nuclear should be in the mix. Fission technology is better , and will continue to get better if we actively promote its development. The prospect of fusion energy also needs to be explored. If it's good enough for the sun and the stars, it's good enough for me.	

Figure 1: Examples for stance detection VAST.

Figure 1 Example 1, the document does not explicitly contain the topic 'Olympics', but 'Games' and 'Athlete' implicitly refer to the topic.

Existing work incorporates external knowledge to solve the challenges (Liu et al., 2021; Jayaram and Allaway, 2021). For example, CKE-Net achieves the state-of-the-art results for zero-shot stance detection, which uses pre-trained model BERT and commonsense knowledge graph on ConceptNet (Liu et al., 2021). However, such a method only considers the knowledge relations between documents and topics (i.e., the commonsense knowledge in two-hop directed paths on the ConceptNet from documents to topics), limiting the generalization of adding other types of related knowledge. In Figure 1 Example 1, the word 'games' can also represent the computer programs in a different document. Such knowledge cannot be used for that document if no relation between 'game' and 'computer program' can be learned from the relations between documents and topics in the dataset.

We consider incorporating two types of general knowledge, including common sense and sentiment. First, we incorporate commonsense

knowledge into the stance detection model using a graph autoencoder module. We take a pre-training method to train the graph autoencoder, separately to the stance detection module. Second, stance detection is significantly influenced by the sentiment information (Li and Caragea, 2019; Sobhani et al., 2016; Hardalov et al., 2022) (case study can be seen in Appendix). In Figure 1 Example 2, the document contains many positive words like ‘good’, and ‘better’ regarding the topic ‘nuclear power’, which implies a *Pro* stance. However, little existing work has considered sentiment knowledge for zero-shot stance detection. We use the sentiment-aware BERT (SentiBERT henceforth) to extract the sentiment information, assisting in classifying the stances of topics.

Existing work on injecting knowledge into NLP models can be broadly classified into two categories. One uses a graph encoder to integrate structural knowledge into a neural encoder (Li et al., 2019; Ghosal et al., 2020; Bai et al., 2021) and the other injects knowledge by using training losses to tune model parameters (Jayaram and Allaway, 2021; Peters et al., 2019; Logan et al., 2019; Liu et al., 2019). In our work, we consider the former for commonsense knowledge and the latter for sentiment due to the sources of information. In the component of knowledge graph encoding, a graph autoencoder consisting of relational graph convolutional network (RGCN) encoders (Schlichtkrull et al., 2018) and a DisMult decoder (Yang et al., 2014) is trained using negative sampling to obtain the relations of concepts on the commonsense knowledge graph. We inject sentiment knowledge encoded by SentiBERT into BERT using a cross attention module and tuning the fusing process by the training loss of the stance detection.

Our model achieves the state-of-the-art performance on the benchmark dataset VAST (Allaway and Mckeown, 2020) in both zero-shot and few-shot stance detection, improving the performance on many challenging linguistic phenomena such as sarcasm and quotations. We analyze the performance of our model with respect to different sentiment and common sense features, finding that the data with the corresponding sentiment and stance pairs (i.e., (*Pos*, *Pro*) and (*Neg*, *Con*)) are the easiest part for models to classify; in addition, increased commonsense knowledge leads to improved performance of the stance detection model. To our knowledge, we are the first to incorporate

both sentiment and common sense into zero-shot stance detection model. The code has been released <https://github.com/LuoXiaoHeics/StanceCS>.

2 Related Work

Stance detection, also known as stance classification (Walker et al., 2012), stance identification (Zhang et al., 2017), stance prediction (Qiu et al., 2015), debate-side classification (Anand et al., 2011), and debate stance classification (Hasan and Ng, 2013), aims to identify the stance of the text author towards a target (an entity, event, idea, opinion, claim, topic, etc.) either explicitly mentioned or implied within the text. For the initial task of stance detection, models are trained an individual classifier for each topic (Lin et al., 2006; Beigman Klebanov et al., 2010; Sridhar et al., 2015; Hasan and Ng, 2013, 2014; Li et al., 2018) or only a small number of topics are both in training and evaluation sets (Faulkner, 2014; Du et al., 2017; Hardalov et al., 2021).

However, given rich and varying topics, data annotation can be time-consuming and labor-intensive. Researchers attempt to solve the task in a cross-target setting (Augenstein et al., 2016; Xu et al., 2018a), training the model in a topic and testing it on another one, and propose several weakly supervised approaches using unlabeled data related to the test topics (Zarrella and Marsh, 2016; Wei et al., 2016; Dias and Becker, 2016). Other studies propose the tasks of zero-shot and few-shot stance detection, which requires training the model in data of several topics and testing it on some unseen topics (Allaway and Mckeown, 2020).

Allaway and Mckeown (2020) propose to solve the task using a topic-grouped attention net, which uses the relation between the training and evaluation topics in an unsupervised way, and they also analyze the relationship between sentiment and stance from the perspective of the model by corrupting sentences with replacing sentiment words. Jayaram and Allaway (2021) use human rationales as attribution priors to provide faithful explanations of models. Liu et al. (2021) propose to incorporate commonsense knowledge to learn the relations between different topics utilizing a CompGCN (a variant of graph convolution networks). However, it limits the content of knowledge (only knowledge from documents to stances in the training data). Our model differs from such a method in that our model adopts the related concepts of both

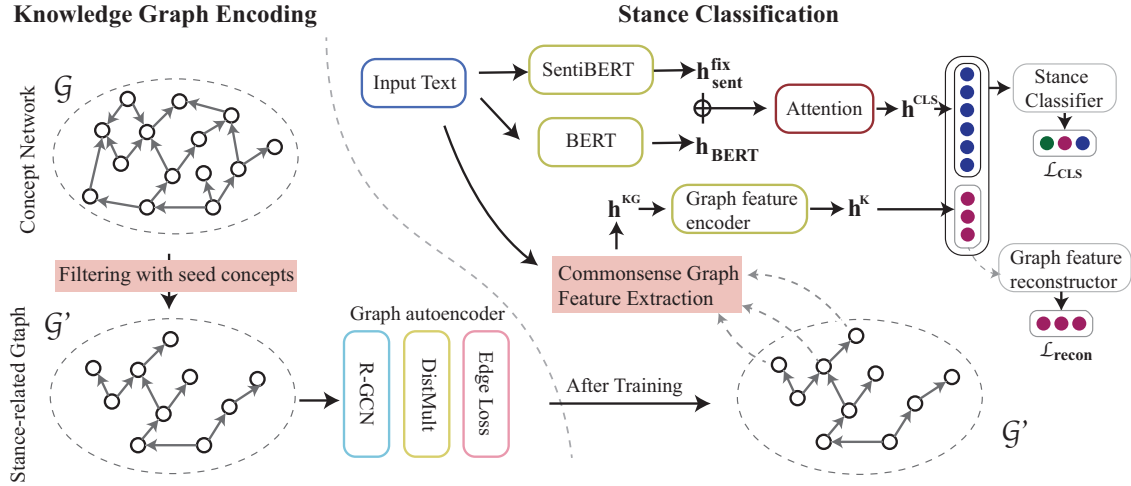


Figure 2: Framework of our proposed model, which contains two components, (1) knowledge graph encoding, (2) stance detection with sentiment and common sense.

documents and topics and uses a pre-trained graph autoencoder to obtain commonsense information. Adversarial learning is also applied to solve the zero-shot task by using unlabeled raw data (Allaway et al., 2021). Unlike the above work, we consider integrating external knowledge for zero-shot stance detection, including sentiment and commonsense information that are rarely considered. To our knowledge, we are the first to systematically incorporate sentiment and commonsense knowledge into the stance detection model and analyze the relationship between them (in Section 4.5 and 4.6).

3 Method

The architecture of our model is illustrated in Figure 2, which contains two components: (1) knowledge graph encoding, which integrates commonsense knowledge from ConceptNet (Section 3.1); (2) stance detection with sentiment and commonsense knowledge (Section 3.2).

3.1 Knowledge Graph Autoencoder

Formally, the ConceptNet is represented as a directed labeled graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{R}\}$, with concepts $v_i \in \mathcal{V}$ and labeled edges $(v_i, r, v_j) \in \mathcal{E}$, where $r \in \mathcal{R}$ is the relation type of edge between v_i and v_j . The concepts in ConceptNet are unigram words or n-gram phrases in the triplet format. For example, one such triplet from ConceptNet is *(teacher, RelatedTo, job)*.

ConceptNet has a large size of approximately 14 million edges. We extract a subset of edges related to the VAST dataset for our task. From the training documents in VAST, we first extract the

set of all unique nouns, adjectives, and adverbs. These words are treated as the seeds that we use to filter the ConceptNet to a sub-graph. We extract all the triplets with a one-edge distance to any of those seed concepts, resulting in a sub-graph $\mathcal{G}' = \{\mathcal{V}', \mathcal{E}', \mathcal{R}'\}$ with 310k concepts and 750k edges. The top 5 relations include ‘RelatedTo’, ‘HasContext’, ‘IsA’, ‘Synonym’ and ‘DerivedFrom’. The sub-graph \mathcal{G}' contains all the concepts related to stance targets in the VAST dataset.

Following Schlichtkrull et al. (2018), we construct a graph autoencoder to compute the representations of concepts in the sub-graph \mathcal{G}' . The autoencoder takes an incomplete set (randomly sampled with 50% probability in our model) of edges $\hat{\mathcal{E}}'$ from \mathcal{E}' in \mathcal{G}' as input. $\hat{\mathcal{E}}'$ is negative sampled to the overall set of samples denoted \mathcal{T} (details in Training). Then we assign the possible edges $(v_i, r, v_j) \in \mathcal{T}$ with scores to determine the probability these edges are in \mathcal{E}' . Our graph autoencoder consists of a relational concept network (RGCN) (Schlichtkrull et al., 2018) encoder to obtain the latent feature representations of concepts and a DistMult scoring decoder (Yang et al., 2014) to recover the missing facts of triplets.

Encoder. RGCN has a solid ability to accumulate relational evidence in multiple inference steps. In each step, a neighborhood-based convolutional feature transformation process uses the related concepts to induce an enriched stance-aggregated feature vector for each concept. Our model contains two stacked RGCN encoders. We first initialize the parameters of concept feature vectors \mathbf{g}_i . Then the vectors are transformed into stance-aggregated

feature vectors $\mathbf{h}_i \in \mathbb{R}^d$ using the **RGCN encoders**:

$$f(x_i, l) = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in N_i^r} \frac{1}{v_{i,r}} W_r^{(l)} x_j + W_0^{(l)} x_i\right),$$

$$\mathbf{h}_i = \mathbf{h}_i^{(2)} = f(\mathbf{h}_i^{(1)}, 2); \mathbf{h}_i^{(1)} = f(\mathbf{g}_i, 1), \quad (1)$$

where f is the encoder network (requiring inputs of feature vector x_i and the rank of the layer l), N_i^r denotes the neighbouring concepts i with the relation $r \in \mathcal{R}$; $v_{i,r}$ is a normalization constant, which can be set in advance $v_{i,r} = |N_i^r|$ or learned by network learning; σ is the activation function like ReLU and $W_r^{(1/2)}, W_0^{(1/2)}$ are learnable parameters though training.

Training. We use **DistMult factorization as the decoder** to assign scores. For a given triplet (v_i, r, v_j) , the score can be obtain as follows:

$$s(v_i, r, v_j) = \sigma(\mathbf{h}_{v_i}^T R_r \mathbf{h}_{v_j}), \quad (2)$$

where σ is logistic function; $\mathbf{h}_{v_i}, \mathbf{h}_{v_j} \in \mathbb{R}^d$ are the encoding feature vectors through the graph encoder for concept v_i and v_j . Each relation $r \in R$ is also associated with a diagonal matrix $R_r \in \mathbb{R}^{d \times d}$.

Our graph autoencoder module is trained using negative sampling (Schlichtkrull et al., 2018). We randomly corrupt the positive triplets, i.e., triplets in $\hat{\mathcal{E}}'$, to create an equal number of negative samples. The corruption is performed by modifying either of the connected concepts or relations randomly, creating the overall set of samples denoted by \mathcal{T} . The training objective is a binary classification between positive/negative (denoted as u) triplets with a cross entropy loss function:

$$\mathcal{L}_{\mathcal{G}'} = -\frac{1}{2|\hat{\mathcal{E}}'|} \sum_{(v_i, r, v_j, u) \in \mathcal{T}} (u \log s(v_i, r, v_j) + (1-u) \log(1-s(v_i, r, v_j))). \quad (3)$$

3.2 Stance Detection Module

Sentiment Feature Encoding. To learn sentiment knowledge, we follow Zhou et al. (2020) to continually train BERT with sentiment masking. We mask the sentiment-related tokens such as sentiment lexicons, emoticons, and ratings with higher probability than general tokens. The model is trained to reconstruct the masked sentiment tokens and predict the rating of the sentences. The corrupted text

\hat{x} is fed into BERT to obtain each word representation \mathbf{h}_i and the sentence representation \mathbf{h}^{CLS} . Softmax layers are used on \mathbf{h}_i to predict each word’s probability, the sentiment of words, and emoticon probability, respectively. A softmax layer on \mathbf{h}^{CLS} is also used to predict the rating of the text \hat{x} . The tasks are trained using cross-entropy loss. Following Zhou et al. (2020), the SentiBERT are trained on Amazon review dataset (Ni et al., 2019) and Yelp 2020¹ challenge dataset.

After pre-training the SentiBERT, given a document d and a topic t , we concatenate d and t as our model input x in the following format: $x = [CLS] d [SEP] t [SEP]$, SentiBERT to obtain its hidden states:

$$\mathbf{h}_{sent}^{fix} = SentiBERT(x), \quad (4)$$

where the parameters of SentiBERT are fixed in our model to keep sentiment information stabilized.

Commonsense Feature Encoding. After training the graph autoencoder, in order to extract the document-specific commonsense graph feature for the document d and the topic t , the unique nouns, adjectives, and adverbs in the document d and the topic t are extracted at first, which we denote as S . Then we extract a sub-graph \mathcal{G}'_S from \mathcal{G}' , which contains all the triplets either of whose concepts are in S or within the vicinity of radius 1 from any of the concepts in S . Next, we make a forward pass of \mathcal{G}'_S through the encoder of graph autoencoder to obtain the feature vectors \mathbf{h}_j for all unique concepts j in \mathcal{G}'_S . The average of feature vectors \mathbf{h}_j for all unique concepts in \mathcal{G}'_S is regarded as the commonsense graph feature vector \mathbf{h}^{KG} for the document d . The commonsense graph feature vector \mathbf{h}^{KG} is fed into a encoder layer to obtain hidden states \mathbf{h}^K :

$$\mathbf{h}^K = W_k \mathbf{h}^{KG} + b_k \quad (5)$$

where W_k and b_k are the trained parameters of the linear layer.

Stance Classification. The input x is first fed into BERT to obtain its hidden states:

$$\mathbf{h}_{BERT} = BERT(x) \quad (6)$$

Then the hidden states of $\mathbf{h}_{BERT}, \mathbf{h}_{sent}^{fix}$ are concatenated and fed into a **cross attention** module to fuse the information of BERT and SentiBERT:

$$\mathbf{h}^{CLS} = CrossAttention([\mathbf{h}_{BERT}, \mathbf{h}_{sent}^{fix}])[CLS], \quad (7)$$

¹<https://www.yelp.com/dataset>

	#Exp	#Doc	#Zero-shot	#Few-shot
Train	13477	638	1481	4003
Dev	2062	114	682	383
Test	3066	159	786	600

Table 1: Statistics on the VAST dataset.

where \mathbf{h}^{CLS} is the hidden states of $[CLS]$ token in BERT. The hidden states vectors of \mathbf{h}^K and \mathbf{h}^{CLS} are concatenated to for classification:

$$p = \text{Softmax}(W[\mathbf{h}^{CLS}, \mathbf{h}^K] + b), \quad (8)$$

where W and b are the parameters and p is the probability distribution on the three stance labels.

Training. Given the input and its golden label (x_i, y_i) , the loss function \mathcal{L}_{cls} for classifying stance is cross entropy:

$$\mathcal{L}_{cls} = -\frac{1}{|N|} \sum_{(x_i, y_i)} y_i \log p(y_i), \quad (9)$$

where $|N|$ is the number of data samples. To further ensure stronger topic invariance constraints of \mathbf{h}_{KG} , we add a shared decoder layer D_{recon} with a reconstruction loss:

$$\mathcal{L}_{recon} = -E_{\mathbf{h}^{KG}}(\|D_{recon}(\mathbf{h}^K) - \mathbf{h}^{KG}\|_2^2). \quad (10)$$

The overall loss function is:

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{recon}. \quad (11)$$

4 Experiments

We verify the effectiveness of sentiment and common sense influence for zero-shot and few-shot stance detection. We also prove the significance of each module in our model in Section 4.4 and analyze the relationship between sentiment (common sense) and stance in Section 4.5 (4.6).

4.1 Settings

Dataset: We adopt the dataset for zero-shot and few-shot stance detection task—VARied Stance Topics (VAST) (Allaway and Mckeown, 2020), which is practical and useful for real-world applications. The dataset consists of thousands of topics, and the statistics are summarized in Table 1. The zero-shot topics only appear in the test set, and the few-shot topics only contain a few training examples.

Training Details We perform experiments using the official pre-trained BERT model provided by

Huggingface². For the pre-trained model with sentiment information, we adopt the model provided by Zhou et al. (2020), which is a continually trained BERT on sentiment datasets. We train our model on 1 GPU (Nvidia GTX2080Ti) using the Adam optimizer (Kingma and Ba, 2014). For training the graph autoencoder, the initial learning rate is 1e-2. For the stance detection training process, the initial learning rate is 1.5e-5, the max sequence length for BERT and SentiBERT is 256, the batch size for training is 4, and the model is trained for three epochs.

Baselines We compare our model with several state-of-the-art baselines: (1) **BiCond** (Augenstein et al., 2016), a model for cross-domain target stance detection task which uses one BiLSTM to encoding the topic and another BiLSTM to encoded the text; (2) **CrossNet** (Xu et al., 2018b), a model based on the BiCond adding an aspect-specific attention layer for cross-target setting; (3) **SENT** (Zhang et al., 2020), a model using the semantic-emotion heterogeneous graph to enhance BiLSTM for cross-traget stance detection; (4) **BERT-sep**, a model that encodes the text and topic separately, using BERT, and then classification with a two-layer feed-forward neural network; (5) **BERT-joint** (Allaway and Mckeown, 2020), a model with contextual conditional encoding followed by a two-layer feed-forward neural network; (6) **TGA-Net** (Allaway and Mckeown, 2020), a model using contextual conditional encoding and topic-grouped attention. In addition, we also consider the models BERT-joint-ft and TGA-Net-ft where the BERT module is fine-tuned; (7) **Prior-Bin:gold** (Jayaram and Allaway, 2021), a model applying human rationales as attributions to assist the stance detection; (8) **BERT-GCN** (Liu et al., 2021), a model applying the conventional GCN (Kipf and Welling, 2016), which considers node information aggregation; (9) **CKE-Net** (Liu et al., 2021), a model based on BERT, using the CompGCN (Vashishth et al., 2019) to obtain the commonsense information.

4.2 Results

The results are shown in Table 2. Compared with previous models, our model achieves the state-of-the-art performance in zero-shot, few-shot, and all the topics of VAST. In particular, the macro F1 scores are 72.6%, 70.2%, and 71.3%, which are 2.4%, 0.1%, and 1.2% higher than CKE-Net

²<https://huggingface.co/>

Model	F1 Zero-shot				F1 Few-Shot				F1 All			
	pro	con	neu	all	pro	con	neu	all	pro	con	neu	all
BiCond	.459	.475	.349	.427	.454	.463	.259	.392	.457	.468	.306	.410
Cross-Net	.462	.434	.404	.434	.508	.505	.410	.474	.486	.471	.408	.455
SEKT	.504	.442	.308	.418	.510	.479	.215	.474	.507	.462	.263	.411
BERT-sep	.414	.506	.454	.458	.524	.539	.544	.536	.473	.522	.501	.499
BERT-joint	.546	.584	.853	.660	.543	.597	.796	.646	.545	.591	.823	.653
TGA-Net	.554	.585	.858	.666	.589	.595	.805	.663	.573	.590	.831	.665
BERT-joint-ft	.579	.603	.875	.685	.595	.621	.831	.684	.588	.614	.853	.684
TGA-Net-ft	.568	.598	.885	.684	.628	.601	.834	.687	.599	.599	.859	.686
Prior-Bin:gold	.643	.581	.852	.692	.632	.563	.881	.692	.652	.597	.824	.691
BERT-GCN	.583	.606	.869	.686	.628	.634	.830	.697	.606	.620	.849	.692
CKE-Net	.612	.612	.880	.702	.644	.622	.835	.701	.629	.617	.857	.701
Our Model												
BS	.625	.667	.870	.717	.601	.667	.828	.699	.591	.669	.858	.706
S-RGCN	.582	.669	.838	.699	.561	.623	.809	.665	.607	.657	.842	.702
B-RGCN	.594	.657	.885	.712	.568	.678	.851	.699	.591	.663	.865	.706
BS-RGCN(<i>proposed</i>)	.608	.674	.895	.726	.600	.665	.839	.702	.604	.669	.866	.713

Table 2: Overall results. The suffix "ft" means BERT is fine-tuned. BS – the combination of BERT and SentiBERT; S-RGCN – the combination of SentiBERT and the graph autoencoder; B-RGCN – the combination of BERT and the graph autoencoder; BS-RGCN – our proposed model.

Model	Imp	mIT	mIS	Qte	Sarc
BERT-joint	.571	.590	.524	.634	.601
TGA-Net	.594	.605	.532	.661	.637
BERT-joint-ft	.617	.621	.547	.668	.673
BERT-GCN	.619	.627	.547	.668	.673
CKE-Net	.625	.634	.553	.695	.682
BS-RGCN	.621	.647	.556	.701	.717

Table 3: Accuracies on five challenges on the test set.

model, respectively. The results of B-RGCN (our model without SentiBERT module) are 71.2% and 69.9%, with a higher macro F1 score on zero-shot topics but a similar result on few-shot topics compared with CKE-Net. The performances of both our model and B-RGCN increase largely on the zero-shot topics but less on few-shot topics, which implies that our graph autoencoder module can achieve a similar effect compared with the GCN module of CKE-Net in the few-shot topics but can improve the effectiveness in extracting relation information in zero-shot topics. This verifies the intuition that only considering the relations between documents and topics limits the transferability of CKE-Net for the zero-shot task. Compared with Prior-Bin:gold, the macro F1 scores of our model are 3.4%, 1.0%, and 2.2% higher on zero-shot, few-shot, and all the topics sets, respectively. It implies that commonsense knowledge and sentiment information are more effective than the set of specific human rationales by Prior-Bin:gold as attributions.

Our model achieves better performance on *Con* labels (67.4%, 66.5%, 66.9%) compared with *Pro*

labels (60.8%, 60.0%, 60.4%), which is similar to most of the previous models (BERT-GCN, TGA-Net, and so on). The phenomenon also appears in B-RGCN and BS, which are our models without SentiBERT and without BERT, respectively (the analysis of the ablation study is explained in Section 4.4 in detail). The results suggest that the use of SentiBERT does not cause the imbalanced performance on different stances and the detection difficulty is mainly on *Pro* labels. In addition, the results of *Neu* stance labels are the highest (89.5%, 83.9%, 86.6%) than those of other labels. It indicates that it is easier for models to classify the *Neu*, where the topics are mostly unrelated to documents.

4.3 Breakdown Evaluation

We also test our model on five special phenomena of the test set on VAST following Allaway and Mckeown (2020): (1) **Imp**: non-neutral stance examples where the topics are not explicit in the documents; (2) **mIT**: documents having multiple stance topics with different topics; (3) **mIS**: documents having multiple stance topics with different and non-neutral labels; (4) **Qte**: documents with quotations; (5) **Sarc**: documents with sarcasm.

The results are shown in Table 3. Our model achieves the state-of-the-art performance on mIT, mIS, Qte, and Sarc with 64.7%, 55.6%, 70.1%, and 71.7%, respectively. In particular, the improvement of our model on mIS implies that different types of knowledge features help models extract stance topics-related information. The most challenging

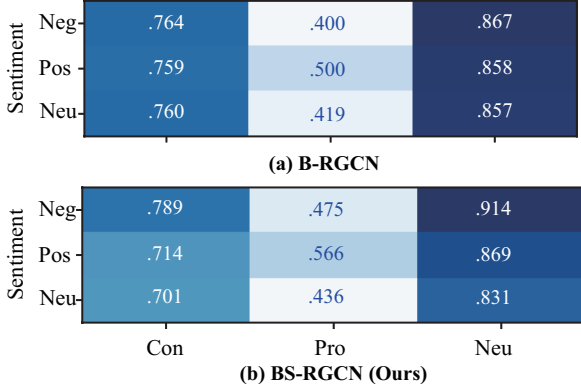


Figure 3: Accuracies of B-RGCN and BS-RGCN on all test data w.r.t different sentiment and stance pairs.

task is mIS, with a macro F1 score of 55.6% by our model. The results demonstrate that it is highly challenging to classify the topics with different stances since the stance information extracted in the model is more related to the whole sentence but more minor to the topics. The macro F1 score of Sarc increases the most, 3.5% higher than that of CKE-NET, implying that the sentiment information helps boost the model performance in understanding sarcasm, which is a sentiment-related linguistic phenomenon. The accuracy of our model on Imp is the second-highest (slightly lower than that of CKE-Net), which indicates that introducing commonsense graph knowledge can help improve the model performance on the zero-shot task.

4.4 Ablation Study

We conduct ablation studies of BS, S-RGCN, and B-RGCN to understand the significance of the graph autoencoder, BERT, and SentiBERT modules, respectively. The results are shown in Table 2. First, BS fuses BERT and SentiBERT feature vectors using Eq(5-6) and classifies the stance using \mathbf{h}^{CLS} with a linear layer. It achieves macro F1 scores of 71.7%, 69.9%, and 70.6% on the zero-shot, few-shot, and all the topics, which are 3.2%, 1.5%, and 2.2% higher than those of BERT-joint-ft, respectively, which proves that sentiment information can help boost the performance of stance detection task.

Second, B-RGCN and S-RGCN are models without fusing the BERT and SentiBERT feature vectors. The feature vectors of $[CLS]$ tokens from BERT or SentiBERT (the parameters of SentiBERT are not fixed) are directly concatenated with knowledge graph feature vectors to classify the stance. The macro F1 scores of S-RGCN are 69.9% and

66.5% on the zero-shot topics and the few-shot topics, 1.3%, and 3.4% lower than those of B-RGCN, respectively. It indicates that it is not sufficient to use a sentiment-specific model to do stance classification. The macro F1 score of B-RGCN on the zero-shot set is 71.2%, 1.0% higher than that of CKE-Net, which shows that our graph autoencoder module can achieve better performance for zero-shot stance detection than CompGCN. However, BS, B-RGCN, and S-RGCN do not outperform BS-RGCN in the zero-shot topics and all the topics set, which shows that the graph autoencoder, BERT, and SentiBERT are all useful for the stance detection task.

4.5 Sentiment and Stance

Allaway and Mckeown (2020) indicate that models of BERT-Joint are reliant on sentiment cues, and the models learn the strong association between the *Neg* (negative) sentiment and the *Con* stance, yet weak association between *Pos* (positive) sentiment and *Pro* stance. Their analysis is based on experiments where the documents are corrupted by replacing the text’s sentiment words. Here we take a different perspective and carry out experiments with respect to different stances and sentiment pairs on both B-RGCN and BS-RGCN. We use opinion lexicon (Hu and Liu, 2004) to classify the sentiment of document, (i.e, if a document contains more positive/negative words, we treat it as a document with the *Pos* (positive)/*Neg* (negative) sentiment; otherwise, we treat it as a document with the *Neu* (neutral) sentiment).

The results are shown in Figure 3 (the model trained on all the topics is tested in this experiment). For BS-RGCN, the accuracy on the corresponding stance and sentiment (*Neg, Con*) is 78.9%, higher than 71.4% of (*Pos, Con*) and 70.1% of (*Neu, Con*). Similarly, the accuracy on (*Pos, Pro*) is 56.6%, higher than 47.5% of (*Neg, Pro*) and 43.6% of (*Neu, Pro*). This suggests that data samples with corresponding sentiment and stance pairs (*(Pos, Pro)*, *(Neg, Con)*) are easier to classify by our model. The performance of B-RGCN is similar to BS-RGCN, with an accuracy of 76.4% for (*Neg, Pro*), a little higher than those of (*Pos, Con*) (75.9%) and (*Neu, Con*) (76.0%). The same model achieves an accuracy of 50% of (*Pos, Pro*), 10% higher than that of (*Neg, Pro*), and 8.1% higher than that of (*Neu, Pro*). The model without the sentiment module can also predict corresponding sentiment and

Context	Topic	Gold Label	Output
I have lived in brazil for the last five years (and off and on over the last 27 years). I know of no one here who is even remotely excited about the Olympics. It would seem that people don't care. The economy is tanking and government is at a complete standstill . We have more important things on our mind right now.	Olympics	Con	Con
I can't even believe that this is a debate. Cutting the most basic foreign language programs ? How does one appreciate that there is a world outside of America? Google translate? Suny, everyone is laughing at you and you're too smug to notice.	College	Con	Con
Good idea. I have always had a cat or two. While being inhumane , declawing places a cat in danger. Should my charming indoor kitty somehow escape outside, he would have no way to defend himself.	nail removal	Con	Con

Table 4: Case Study for our trained stance detection model. Case I shows the effectiveness of using sentiment information; Case II shows the importance of commonsense knowledge; Case III shows both the sentiment and commonsense knowledge help the stance detection model.

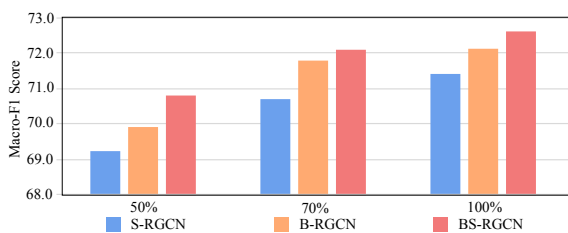


Figure 4: Macro-F1 scores of S-RGCN, B-RGCN, and BS-RGCN on zero-shot test data w.r.t different percents of commonsense knowledge for pre-training.

stance pairs with higher accuracy, demonstrating that sentiment information can help stance detection models. The accuracies for B-RGCN and BS-RGCN are both significantly higher on data with *Con* stances than those with *Pro* stances. The phenomenon indicates that it is difficult for models to predict *Pro* stance in the VAST dataset, and the difference in performance is not caused by the difference of associations between data of (*Pos*, *Pro*) and (*Neg*, *Con*). For the data of *Neu* stance, the performance is less related to sentiments. The models can achieve much better results on *Neu* stance data, where the topics may be not related to the documents, 91.4% on (*Neg*, *Neu*), 86.9% on (*Pos*, *Neu*), 83.1% on (*Neu*, *Neu*) for BS-RGCN, and 86.7% on (*Neg*, *Neu*), 85.8% on (*Pos*, *Neu*), 85.7% on (*Neu*, *Neu*) for B-RGCN. The phenomenon demonstrates that it is easy for the model to judge whether the topic is related to the documents.

4.6 Common Sense and Stance

We show the relationship between common sense and stance by pre-training the graph autoencoder w.r.t different percentages of extracted concepts (Section 3.1). Using the commonsense feature with

the pre-trained autoencoder, we show the results of the stance detection models B-GCN, S-GCN, and BS-RGCN on the zero-shot task. The results are given in Figure 4. As observed, the performance of the three models increases with increasing coverage of commonsense knowledge. It indicates that commonsense knowledge is directly useful for stance detection models.

4.7 Case Study

We also show some cases from the test data using the model trained on all the topics. In the first case, sentiment words such as ‘tanking’ or ‘standstill’ imply the negative sentiments towards the influence of the Olympics on the economy of Brazil, which further expresses an opposing stance towards ‘Olympics’. Our model outputs the correct label towards the target thanks to the sentiment information. In the second case, no explicit expression of the target ‘College’ is contained in the document. Only some implications, including ‘foreign language programs’, have relation to the ‘College’, and with the commonsense knowledge encoding, our model outputs the correct stance. The third case proves that both common sense and sentiment information can benefit the stance detection model, that ‘inhumane’ expresses a negative sentiment, and the topic ‘nail removal’ is implicitly involved by the word ‘declawing’. Our model can also give the correct stance for case III.

5 Conclusion

We proposed a stance detection model incorporating commonsense knowledge and sentiment information, achieving state-of-the-art zero-shot and few-shot stance detection results on the standard

dataset. The ablation study showed the significance of each module, such as knowledge graph autoencoder, SentiBERT, and BERT. We also analyzed the relation between sentiment/common sense and stance, which indicate the effectiveness of this external knowledge.

Acknowledgements

Yue Zhang is the corresponding author. We would also like to thank the anonymous reviewers for the detailed and thoughtful reviews. The work is funded by the Zhejiang Province Key Project 2022SDXHDX0003.

References

- Emily Allaway and Kathleen Mckeown. 2020. Zero-shot stance detection: A dataset and model using generalized topic representations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8913–8931.
- Emily Allaway, Malavika Srikanth, and Kathleen Mckeown. 2021. Adversarial learning for zero-shot stance detection on social media. *arXiv preprint arXiv:2105.06603*.
- Pranav Anand, Marilyn Walker, Rob Abbott, Jean E Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 1–9.
- Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. [Stance detection with bidirectional conditional encoding](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.
- Xuefeng Bai, Yulong Chen, Linfeng Song, and Yue Zhang. 2021. [Semantic representation for dialogue modeling](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4430–4445, Online. Association for Computational Linguistics.
- Beata Beigman Klebanov, Eyal Beigman, and Daniel Diermeier. 2010. [Vocabulary choice as an indicator of perspective](#). In *Proceedings of the ACL 2010 Conference Short Papers*, pages 253–257, Uppsala, Sweden. Association for Computational Linguistics.
- Marcelo Dias and Karin Becker. 2016. Inf-ufrgs-opinion-mining at semeval-2016 task 6: Automatic generation of a training corpus for unsupervised identification of stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 378–383.
- Jiachen Du, Ruifeng Xu, Yulan He, and Lin Gui. 2017. Stance classification with target-specific neural attention networks. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 3988–3994.
- Adam Faulkner. 2014. Automated classification of stance in student essays: An approach using stance target information and the wikipedia link-based measure. In *FLAIRS Conference*.
- Deepanway Ghosal, Devamanyu Hazarika, Abhinaba Roy, Navonil Majumder, Rada Mihalcea, and Soujanya Poria. 2020. [KinGDOM: Knowledge-Guided DOMain Adaptation for Sentiment Analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3198–3210, Online. Association for Computational Linguistics.
- Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2021. [Cross-domain label-adaptive stance detection](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9011–9028, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Momchil Hardalov, Arnav Arora, Preslav Nakov, and Isabelle Augenstein. 2022. Few-shot cross-lingual stance detection with sentiment-based pre-training. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10729–10737.
- Kazi Saidul Hasan and Vincent Ng. 2013. [Stance classification of ideological debates: Data, models, features, and constraints](#). In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 1348–1356, Nagoya, Japan. Asian Federation of Natural Language Processing.
- Kazi Saidul Hasan and Vincent Ng. 2014. [Why are you taking this stance? identifying and classifying reasons in ideological debates](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 751–762, Doha, Qatar. Association for Computational Linguistics.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Sahil Jayaram and Emily Allaway. 2021. [Human rationales as attribution priors for explainable stance detection](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5540–5554, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Chang Li, Aldo Porco, and Dan Goldwasser. 2018. Structured representation learning for online debate stance prediction. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3728–3739, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Pengfei Li, Kezhi Mao, Xuefeng Yang, and Qi Li. 2019. Improving relation extraction with knowledge-attention. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 229–239, Hong Kong, China. Association for Computational Linguistics.
- Yingjie Li and Cornelia Caragea. 2019. Multi-task stance detection with sentiment and stance lexicons. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6299–6305, Hong Kong, China. Association for Computational Linguistics.
- Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on? identifying perspectives at the document and sentence levels. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)*, pages 109–116, New York City. Association for Computational Linguistics.
- Rui Liu, Zheng Lin, Yutong Tan, and Weiping Wang. 2021. Enhancing zero-shot and few-shot stance detection with commonsense knowledge graph. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3152–3157, Online. Association for Computational Linguistics.
- Zhibin Liu, Zheng-Yu Niu, Hua Wu, and Haifeng Wang. 2019. Knowledge aware conversation generation with explainable reasoning over augmented graphs. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1782–1792, Hong Kong, China. Association for Computational Linguistics.
- Robert Logan, Nelson F. Liu, Matthew E. Peters, Matt Gardner, and Sameer Singh. 2019. Barack’s wife hillary: Using knowledge graphs for fact-aware language modeling. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5962–5971, Florence, Italy. Association for Computational Linguistics.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 31–41, San Diego, California. Association for Computational Linguistics.
- Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. Stance and sentiment in tweets. *ACM Trans. Internet Technol.*, 17(3).
- Jianmo Ni, Jiacheng Li, and Julian McAuley. 2019. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 188–197.
- Matthew E. Peters, Mark Neumann, Robert Logan, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. 2019. Knowledge enhanced contextual word representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 43–54, Hong Kong, China. Association for Computational Linguistics.
- Minghui Qiu, Yanchuan Sim, Noah A Smith, and Jing Jiang. 2015. Modeling user arguments, interactions, and attributes for stance prediction in online debate forums. In *Proceedings of the 2015 SIAM international conference on data mining*, pages 855–863. SIAM.
- M. Schlichtkrull, Thomas Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. *ArXiv*, abs/1703.06103.
- Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko. 2016. Detecting stance in tweets and analyzing its interaction with sentiment. In *Proceedings of the fifth joint conference on lexical and computational semantics*, pages 159–169.
- Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 116–125, Beijing, China. Association for Computational Linguistics.
- Shikhar Vashishth, Soumya Sanyal, Vikram Nitin, and Partha Talukdar. 2019. Composition-based multi-relational graph convolutional networks. *arXiv preprint arXiv:1911.03082*.
- 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, NAACL HLT '12, page 592–596, USA. Association for Computational Linguistics.

Wan Wei, Xiao Zhang, Xuqin Liu, Wei Chen, and Tengjiao Wang. 2016. pkudblab at semeval-2016 task 6: A specific convolutional neural network system for effective stance detection. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 384–388.

Chang Xu, Cecile Paris, Surya Nepal, and Ross Sparks. 2018a. Cross-target stance classification with self-attention networks. *arXiv preprint arXiv:1805.06593*.

Chang Xu, Cécile Paris, Surya Nepal, and Ross Sparks. 2018b. [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, Melbourne, Australia. Association for Computational Linguistics.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.

Guido Zarrella and Amy Marsh. 2016. Mitre at semeval-2016 task 6: Transfer learning for stance detection. *Proceedings of SemEval*, pages 470–475.

Bowen Zhang, Min Yang, Xutao Li, Yunming Ye, Xiaofei Xu, and Kuai Dai. 2020. [Enhancing cross-target stance detection with transferable semantic-emotion knowledge](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3188–3197, Online. Association for Computational Linguistics.

Shaodian Zhang, Lin Qiu, Frank Chen, Weinan Zhang, Yong Yu, and Noémie Elhadad. 2017. We make choices we think are going to save us: Debate and stance identification for online breast cancer cam discussions. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1073–1081.

Jie Zhou, Junfeng Tian, Rui Wang, Yuanbin Wu, Wenming Xiao, and Liang He. 2020. Sentix: A sentiment-aware pre-trained model for cross-domain sentiment analysis. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 568–579.

the relation between sentiments and stances. Opinion lexicon (Hu and Liu, 2004) is adopted as the sentiment vocabulary. As shown in Table 5, there are many samples (7 in 10) that sentiment knowledge plays a significant role for stance detection, and few samples have a conflicting relation.

A Appendix

Human Labeling for Sentiment and Stance Detection

In this part, we manually label some samples (randomly selected) from VAST dataset to prove

Context	Topic	Stance	Relation
The reason that Deep Mind winning is so impressive is that Google managed to accomplish this with virtually no warning. It was less than a year ago where the best computer program was not in the top 20,000 in the world. It was less than 6 months ago when the program beat a player in the top 1,000. Yesterday the program beat the the best player in the world. Am I wrong to be shocked at how fast complicate AI has advanced?	Artificial Intelligence	Pro	+
I totally agree with this premise. As a younger person I was against Nuclear power (I was in college during 3 mile island) but now it seems that nuclear should be in the mix. Fission technology is better , and will continue to get better if we actively promote its development. The prospect of fusion energy also needs to be explored. If it's good enough for the sun and the stars, it's good enough for me.	Nuclear Power	Pro	+
This is a horrible idea. Anyone who has worked on the border, or in Mexico (as I do), knows there are plenty of middle and upper-class Mexicans who come to the U.S. for an education. I think Dr. Lee is really perpetuating stereotypes here. In my opinion, affirmative action should be based on economic class, no matter what the race.	Mexico	Pro	0
Good idea. I have always had a cat or two. While being inhumane , declawing places a cat in danger. Should my charming indoor kitty somehow escape outside, he would have no way to defend himself. Why don't humans have their finger-and tonails removed to save on manicures? Answer: they are important to the functioning and protection of our bodies.	nail removal	Con	+
The mandate of private corporations is to make a profit. And if the profit is made at the EXPENSE of the society that allow the corporation to exist, well, too bad . Oil companies foul the environment. Financial companies drive the economy into the Great Recession . Airlines have no regard for the people they transport. As long as they make a profit, they are allowed to abuse the public until they are stopped. That is the way it has been since Swift and Armour canned and sold rotten meat and Carnegie sent Pinkertons to shoot striking miners.	private corporation profit	Con	+
One's own, and learning another language is important and a great work out for the brain! Back in the day, I learned Spanish! In retrospect Latin would have been the better way to go, since mastery of that makes learning the languages like French, Italian, Portuguese, Romanian, and Spanish, very much easier to learn!	Latin helpful language	Pro	+
Without government to ensure their behavior, companies will attempt to make a profit even to the DETRIMENT of the society that supports the business. We have seen this in the environment, in finances, in their treatment of workers and customers. Enough.	company	Pro	0
The "you have a short live, so enjoy" attitude alone did not lead to the Renaissance , the age of Enlightenment, or the Industrial Revolution. It did not le ()ad to the invention of the light bulb, or the telephone, or the internet, or the NYT electronic discussion board. Just "enjoying" life alone means you are enjoying the fruit of someone else's hard work.	Renaissance	Pro	0
Of course their salaries should be raised. But this should be separated from the discussion about legality. Salaries should be raised and only legal workers should be employed. Its really a no brainer. And any discussion about only Mexicans being prepared to do this work so it has to be illegal is completely disingenuous .	illegal labor	Con	+
Also, and usually not acknowledged, is that we are slowly eroding the fertility of the soil. There is no more usable soil , we are farming everything that can be farmed. Current methods depend on petrochemical fertilizers. Even with their use, fertility is slowly declining . As human population continues to grow, the result is obvious.	soil	Con	+

Table 5: Manually labeling samples for the relation between sentiments and stances. The positive/negative words related to the topics are labeled with red/teal colors. The topic related words in the documents are bold. '+' for sentiment words supporting the stance, '0' for no relation.