SEGP: Stance-Emotion joint Data Augmentation with Gradual Prompt-tuning for Stance detection

Junlin Wang[1,2], Yan Zhou[1(✉)], Yaxin Liu[1,2], Weibo Zhang[1,2], and Songlin Hu[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{wangjunlin,zhouyan,liuyaxin,zhangweibo,husonglin}@iie.ac.cn
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

Abstract. Stance detection is an important task in opinion mining, which aims to determine whether the author of a text is in favor of, against, or neutral towards a specific target. By now, the scarcity of annotations is one of the remaining problems in stance detection. In this paper, we propose a Stance-Emotion joint Data Augmentation with Gradual Prompt-tuning (SEGP) model to address this problem. In order to generate more training samples, we propose an auxiliary sentence based Stance-Emotion joint Data Augmentation (SEDA) method, formulate data augmentation as a conditional masked language modeling task. We leverage different relations between stance and emotion to construct auxiliary sentences. SEDA generates augmented samples by predicting the masked words conditioned on both their context and auxiliary sentences. Furthermore, we propose a Gradual Prompt-tuning method to make better use of the augmented samples, which is a combination of prompt-tuning and curriculum learning. Specifically, the model starts by training on only original samples, then adds augmented samples as training progresses. Experimental results show that SEGP significantly outperforms the state-of-the-art approaches.

Keywords: Stance detection · Data augmentation · Curriculum learning

1 Introduction

The goal of stance detection is to classify a piece of text as either being in support, opposition, or neutrality towards a given target, the target may not be directly contained in the text. With the rapid development of social media, more and more people post online to express their support or opposition towards various targets. Stance detection is known to have several practical application areas such as polling, public health surveillance, fake news detection, and so on. These conditions motivate a large number of studies to focus on inferring

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

the stances of users from their posts. Table 1 shows some examples on target "Wearing a Face Mask", annotated with the stance labels.

Table 1. Examples of stance detection task.

Text	Stance
Wearing a mask is common sense and kind to your fellow human. We all have to do our part to slow the spread of COVID-19.	Favor
Spend the day outside, get some sun and fresh air. Without a face mask. Best way to keep up your immune system.	Against
Any skincare suggestions for breakouts because of face masks?	Neutral

One of the biggest challenges in stance detection task is the scarcity of annotated samples. Data augmentation is commonly used to address data scarcity, which aims to generate augmented samples based on limited annotations. Zhang et al. [37] replace words with WordNet [19] synonyms to get augmented sentences. Wei et al. [33] propose EDA, which is a combination of token-level augmentation approaches. These methods are effective, but the replacement strategies are simple, thus can only generate limited diversified patterns. To enhance the consistency between augmented samples and labels, Wu et al. [35] propose CBERT, the segmentation embeddings of BERT [11] are replaced with the annotated labels during augmentation. However, these methods fail to take targets into consideration. To solve this problem, Li et al. [16] propose ASDA, which uses the conditional masked language modeling (C-MLM) task to generate augmented samples under target and stance conditions.

Although ASDA [16] achieves highly competitive performance, there still exist two limitations. First, they neglect the emotional information during augmentation. It should be noted that emotion can affect the judgment of stance. There exists a number of studies that use emotional information to assist stance detection and achieve good results [6, 14, 20]. Thus, we posit that in addition to stance and target information, the introduction of emotional information through auxiliary sentences can further improve the label consistency of augmented samples. Second, they neglect the linguistic adversity problem [17, 31] during training. This problem is introduced by data augmentation method and therefore can be seen as a form of noising, where noised data is harder to learn from than unmodified original data.

In this paper, we propose a Stance-Emotion joint Data Augmentation with Gradual Prompt-tuning (SEGP) model to address the above limitations. Specifically, we present an auxiliary sentence based Stance-Emotion joint Data Augmentation (SEDA) method that generates target-relevant and stance-emotion-consistent samples based on C-MLM task. We suppose that there are "Consistency", "Discrepancy" and "None" relations between stance and emotion. The auxiliary sentences are constructed on the premise of these relations as well as the target. With the help of C-MLM task, SEDA augment the dataset by predicting

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

the masked words conditioned on both their context and the auxiliary sentences. Furthermore, to address the linguistic adversity problem in augmented samples, we propose a Gradual Prompt-tuning method, which combines prompt-tuning with curriculum learning to train our model. We design a template that contains target and stance information. After that, we create an artificial curriculum in the training samples according to the disturbance degree in data augmentation. Starting by training on original samples, we feed augmented samples with a higher level of noising into the model as training progresses. The model learns to explicitly capture stance relations between sentence and target by predicting masked words. Our main contributions can be summarized as follows:

- We propose a Stance-Emotion joint Data Augmentation (SEDA) method, which introduces emotional information in the conditional data augmentation of stance detection.
- We further propose a Gradual Prompt-tuning method to overcome the linguistic adversity problem in augmented samples, which combines prompt-tuning with curriculum learning.
- Experimental results show that our methods significantly outperform the state-of-the-art methods.

2 Related Work

2.1 Stance Detection

Stance detection aims to automatically infer the stance of a text towards specific targets [1, 13], which is related to argument mining, fact-checking, and aspect-level sentiment analysis. Early stance detection tasks concentrate on online forums and debates [27, 29]. Later, a series of studies on different types of targets emerge. The targets become political figures [15, 26], controversial topics [7], and so on. At present, the research tasks are mainly divided into three types, in-target stance detection [36], cross-target stance detection [3], and zero-shot stance detection [2]. In this paper, we focus on in-target stance detection, which means the test target can always be seen in the training stage.

2.2 Data Augmentation

Lexical substitution is a commonly used augmentation strategy, which attempts to substitute words without changing the meaning of the entire text.

The first commonly used approach is the thesaurus-based substitution, which means taking a random word from the sentence and replacing it with its synonym using a thesaurus. Zhang et al. [37] apply this and search synonyms in WordNet [19] database. Mueller et al. [21] use this idea to generate additional training samples for their sentence similarity model. This approach is also used by Wei et al. [33] as one of the four random augmentations in EDA.

The second approach is the word-embedding substitution, which replaces some words in a sentence with their nearest neighbor words in the embedding

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

space. Jiao et al. [10] apply this with GloVe embeddings [23] to improve the generalization of their model on downstream tasks, while Wang et al. [30] use it to augment tweets needed to learn a topic model.

The third approach is based on the masked language model, which has to predict the masked words based on their context. Therefore, the model can generate variations of a text using the mask predictions. Compared to previous approaches, the generated text is more grammatically coherent as the model takes context into account when making predictions. Grag et al. [8] use this idea to generate adversarial samples for text classification. Wu et al. [35] formulate the data augmentation as a C-MLM task. Li et al. [16] propose an Auxiliary Sentence based Data Augmentation (ASDA) method that generates samples based on C-MLM task. Inspired by ASDA, we investigate how to introduce more information via auxiliary sentences.

2.3 Curriculum Learning

Curriculum learning is proposed by Bengio et al. [4], which is a training strategy that imitates the meaningful learning order in human curricula. It posits that models train better when training samples are organized in a meaningful order. In the beginning, researchers assume that there exists a range of difficulties in the training samples [28,34]. They leverage various heuristics to sort samples by difficulty and train models on progressively harder samples. Korbar et al. [12] propose instead of discovering a curriculum in existing samples, samples can be intentionally modified to dictate an artificial range of difficulty. Wei et al. [32] combine this idea with data augmentation and propose a curriculum learning strategy, but the performance is still constricted by the gap of objective forms between pre-training and fine-tuning.

2.4 Prompt-tuning

Pre-trained language models like GPT [5] and BERT [11] capture rich knowledge from massive corpora. To make better use of the knowledge, prompt-tuning is proposed. In prompt-tuning, downstream tasks are also formalized as some objectives of language modeling by leveraging language prompts. The results of language modeling can correspond to the solutions of downstream tasks. With specially constructed prompts and tuning objectives [18,24], we can further inject and stimulate the task-related knowledge in pre-trained models, thus boosting the performance. To our knowledge, there is currently a lack of research on applying prompt-tuning to the stance detection task.

3 Method

In this section, we first introduce the variables and definitions that appear in this paper. Then provide the overall architecture of SEGP and explain it in detail.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

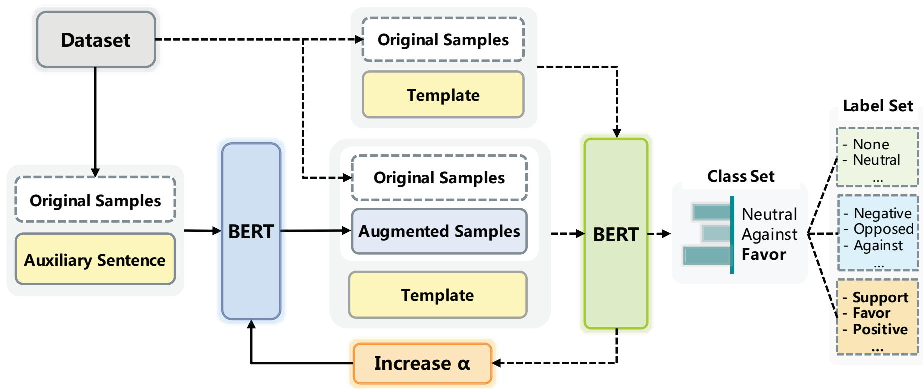

Fig. 1. The overall architecture of SEGP, where α represents the degree of disturbance in the augmentation stage. Solid arrows indicate Stance-Emotion joint Data Augmentation stage and dashed arrows indicate Gradual Prompt-tuning stage.

3.1 Preliminaries

We first give some essential preliminaries. Suppose a given training dataset of size n is $D_{\text{train}} = \{X, S, T, E\}$, where $X = \{x_1, x_2, \ldots, x_n\}$ is the set of input samples. For each $x_i \in X$, it consists of a sequence of l words $x_i = [w_i^1, w_i^2, \ldots, w_i^l]$. We define a stance label set $S = \left\{s_1, s_2, \ldots, s_{|M|}\right\}$, a target set $T = \left\{t_1, t_2, \ldots, t_{|C|}\right\}$ and an emotional label set $E = \left\{e_1, e_2, \ldots, e_{|N|}\right\}$, where the values of $|M|$, $|C|$ and $|N|$ depend on the dataset settings.

3.2 Overall Architecture

In this paper, we propose a Stance-Emotion joint Data Augmentation with Gradual Prompt-tuning (SEGP) model, and the overall architecture is shown in Figure 1. SEGP consists of two stages, as we can see from Figure 1, they are indicated by solid arrows and dashed arrows respectively. The first stage is to get more training samples using the SEDA method. The second is the training stage, which uses the Gradual Prompt-tuning method to overcome the linguistic adversity problem in augmented samples.

3.3 Stance-Emotion joint Data Augmentation

The objective of a data augmentation method is to generate training samples based on the existing limited annotations. In this paper, we propose a novel conditional data augmentation method called SEDA, which is based on C-MLM task. We leverage stance, emotion, and target information to construct auxiliary sentences. SEDA generates target-relevant and stance-emotion-consistent augmented samples by predicting masked words conditioned on context and auxiliary sentences.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

Construction of Auxiliary Sentences. Many approaches achieve better results by taking emotional information as auxiliary information. It should be noted that stance could be inferred independently from the emotional state, the emotions contained in a text may be positive but expresses an opposition stance to a given target. This is due to the complexity of interpreting a stance because it is not always directly consistent with the emotional polarity. We analyze the distribution of stance and emotional labels in COVID-19-Stance dataset. As shown in Figure 2, there is a large gap in the distribution of these two types of labels.

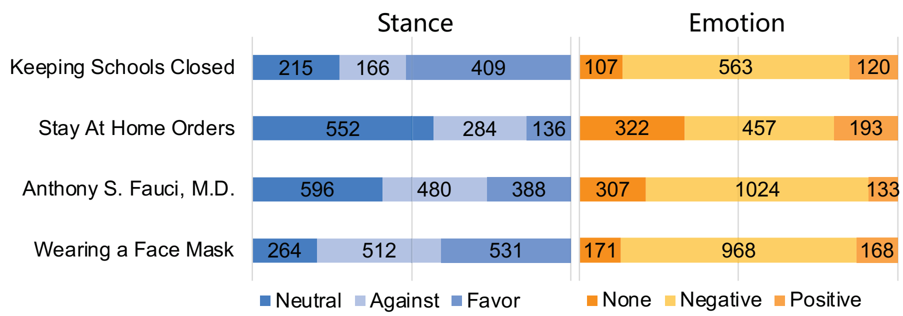

Fig. 2. Stance and emotion distribution in COVID-19-Stance dataset.

Our research is based on stance label set S = {"Neutral", "Against", "Favor"} and emotional label set E = {"None", "Negative", "Positive"}. In order to integrate these two types of information, we define a cross label set $C = \{S - E\}$, which is generated by stance label s and emotional label e. For example, given s="Favor" and e="Negative", we can obtain the cross label c="Favor-Negative". Before constructing auxiliary sentences, we put forward the following relations between stance and emotion:

- Consistency: When cross label c is in {"Favor-Positive", "Against-Negative", "Neutral-Positive", "Neutral-Negative"}, we suppose that the stance is consistent with emotion, so emotional information can be directly introduced into the auxiliary sentence.
- Discrepancy: When cross label c is in {"Favor-Negative", "Against-Positive"}, we suppose that there is a difference between stance and emotion, so we need to consider this contradiction when constructing auxiliary sentences.
- None: When the emotional label e="None", we suppose that the emotional information is not helpful. In this case, the auxiliary sentence only needs to introduce stance information.

Therefore, we leverage the above mentioned relations to construct three kinds of auxiliary sentences regarding target, stance, and emotion. We also place slots in the auxiliary sentences, $\{a_i\}$ is used to fill target words, $\{s_i\}$ is used to fill stance label, and $\{e_i\}$ is used to fill emotional label. Experiments show that

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

grammar correctness is not important. Table 2 shows how to select the corresponding auxiliary sentence according to a cross label. After obtaining the auxiliary sentence, we prepend both another training sample x_j that has the same target and cross label with x_i. The complete input form for each training sample x_i is: Auxiliary sentence+x_j+"the text is":+x_i.

Table 2. Correspondence between Relations, Cross Labels, and Auxiliary Sentences.

Relations	Cross Labels	Auxiliary Sentences
Consistency	Favor-Positive Against-Negative Neutral-Positive Neutral-Negative	The following texts have $\{s_i\}$ stance and $\{e_i\}$ emotion to $\{a_i\}$.
Discrepancy	Favor-Negative Against-Positive	Although the emotion is $\{e_i\}$, the following texts are both $\{s_i\}$ to $\{a_i\}$.
None	Favor-None Neutral-None Against-None	The following texts have $\{s_i\}$ stance to $\{a_i\}$.

For example, given the input x_i: *I don't need to wear a mask to live a healthy life.* with the stance label s="Against" and emotional label e="Positive". The corresponding target is "Wearing a face mask". First, we get its cross label c="Against-Positive" and choose the discrepancy auxiliary sentence. Second, we find another training sample x_j: *The death rate is falling so fast, we don't need to wear masks at all.* So the complete input is: *Although the emotion is {positive}, the following texts are both {against} to {wearing a face mask}. The death rate is falling so fast, we don't need to wear masks at all. The text is: I don't need to wear a mask to live a healthy life.* The introduction of the auxiliary sentence and x_j not only helps to generate more diversified samples, but also provides a strong guideline to help generate target-relevant and label-compatible samples.

Data Generation. We fine-tune the pre-trained model via C-MLM task. For a training sample x_i from X, we specify that the model can only randomly mask words in the input sample x_i and the mask radio is α. Because we want to preserve all of the target, stance, and emotional information. After prepending the corresponding auxiliary sentence and x_j to obtain the masked sentence, a pre-trained language model like BERT is used to predict the masked words. The prediction of masked words depends not only on the context of x_i, but also on their target, stance, and emotion.

After fine-tuning the model on the training dataset for a few epochs, we use the well-trained model for augmentation. Similar to the fine-tuning procedure, the model randomly masks words of the training sample, then prepend the auxiliary sentence and another training sample. The model is used to predict the

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

masked words, we repeat these steps over training samples to get augmented samples.

3.4 Gradual Prompt-tuning

In this paper, we apply the training strategy of curriculum learning to prompt-tuning. We aim to solve the linguistic adversity problem [17, 31] in augmented samples as well as make better use of the knowledge contained in pre-trained language models.

Prompt-tuning. In order to bridge the gap of objective forms between pre-training and fine-tuning, prompt-tuning is proposed. By tuning a pre-trained language model with the cloze-style task, prompt-tuning can manipulate the model behavior to fit various downstream tasks that more fully utilize task-related knowledge in pre-trained language models. Formally, prompt is consists of a template $P(\cdot)$ and a set of stance labels S. For stance detection task, a pre-trained language model uses input sentences and prompt to predict the stance label for a given target. In order to provide more information, we place two slots into the template, $\{t_i\}$ is used to fill target words, and $[MASK]$ is for the model to fill a label word. We set the template $P(\cdot)$=“The stance to $\{t_i\}$ is $[MASK]$”, and map x to the prompt input x_{prompt}=x+“The stance to $\{t_i\}$ is $[MASK]$”. After that, x_{prompt} is fed into a pre-trained model.

The model first converts the input $x_{prompt} = \left\{w_i^1, w_i^2, \ldots, [MASK], \ldots, w_i^l\right\}$ to sequence $\left\{[CLS], w_i^1, w_i^2, \ldots, [MASK], \ldots, w_i^l, [SEP]\right\}$, then compute the hidden vector $h_{[MASK]}$ of $[MASK]$. Given $s \in S$, the model calculates the probability for s can fill the masked position, where $\mathbf{s}$ is the embedding of s in a pre-trained language model. The probability is calculated as follows:

$$p\left([MASK] = s \mid x_{\text{prompt}}\right) = \frac{\exp\left(s \cdot h_{[MASK]}\right)}{\sum_{s \in S} \exp\left(\mathbf{s} \cdot h_{[MASK]}\right)} \tag{1}$$

There also exists an injective mapping function φ that bridges the set of classes Y and the set of label words S, we define $\varphi = Y \rightarrow S$. With the verbalizer φ, we can formalize the probability distribution over Y with the probability distribution over S at the masked position. i.e., $p(y \mid x) = p([MASK] = \phi(y) \mid x_{prompt})$. We map the supporting stance to “Favor”, the opposing stance to “Against” and other stances to “Neutral”. According to model fills the masked position of x_{prompt} with “Favor”, “Against” or “Neutral”, we can get the stance of x. For prompt-tuning, with a template P$(\cdot)$, a label set S and verbalizer φ, the learning objective is to maximize $\frac{1}{|X|} \sum_{x \in X} \log p\left([MASK] = \phi\left(y_x\right) \mid P(x)\right)$.

Curriculum Learning. The data augmentation method might introduce linguistic adversity and can be seen as a form of noising, where noised data is harder to learn from than unmodified original data. Curriculum learning posits that the model train better when training samples are organized in a meaningful

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

order that gradually shows more concepts and complexity. Therefore, we apply the training strategy of curriculum learning to prompt-tuning. We define the mask radio $0.0 \leq \alpha \leq 0.15$ as disturbance degree for SEDA stage, create an artificial curriculum in training samples according to the disturbance degree of the augmented samples. A larger mask ratio α represents a larger variation in the training samples, thus harder to learn from than unmodified original samples. During training, we begin with a disturbance degree of α=0.0 (equivalent to no augmentation), then linearly increase α by 0.05 every time validation loss plateaus, up to a final of α=0.15.

4 Experiment

In this section, we first present the dataset used for evaluation and several baseline methods. Then introduce experimental details and analyze the results.

4.1 Dataset and Baseline Methods

We carry out experiments on the stance detection dataset COVID-19-Stance [9], which is collected by crawling Twitter, using Twitter Streaming API. It contains the tweets of four targets (i.e., “Stay At Home Orders”, “Wearing a Face Mask”, “Keeping Schools Closed” and “Anthony S. Fauci, M.D”), and the stance label of each tweet is either “Favor” or “Against” or “Neutral”.

We compare SEGP with the following baseline methods:

- BiLSTM [25]: Bi-Directional Long Short Term Memory Network takes tweets as input and is trained to predict the stance towards a target, without explicitly using the target information.
- CT-BERT [22]: A pre-trained language model that predicts the stance by appending a linear classification layer to the hidden representation of $[CLS]$ token, pre-trained on a corpus of messages from Twitter about COVID-19.
- CT-BERT-v2 [22]: It is identical to CT-BERT, but trained on more data, resulting in higher downstream performance.
- EDA [33]: A simple data augmentation method that consists of four operations: synonym replacement, random deletion, random swap, and random insertion.
- ASDA [16]: A data augmentation method that generates target-relevant and label-consistent data samples based on C-MLM task.

4.2 Experimental Results

SEGP is implemented based on CT-BERT-v2 [22], using a batch size of 8. The learning rate of Adam optimizer is 1e-5 and the maximum sequence length is 256. Experimental results are shown in Table 3, the best model configuration is selected according to the highest performance on the development set.

We first compare SEGP with BiLSTM [25], CT-BERT [22] and CT-BERT-v2 [22]. It can be seen that SEGP is superior to all baselines in accuracy and F1

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

score, which demonstrates the validity of our model in stance detection tasks. Besides, we compare SEGP with different data augmentation methods, i.e., EDA and ASDA. We can observe that SEGP performs the best, while EDA and ASDA methods have limited improvement in performance. Furthermore, when target="Anthony S. Fauci, M.D.", the result is even worse than CT-BERT that only trained on original samples.

SEGP has better performance on all targets, which proves it can not only generate more diversified samples but also have the ability to overcome the linguistic adversity problem and better utilize task-related knowledge in pre-trained language models.

Table 3. Performance of SEGP and different baseline methods for stance detection on four targets in the COVID-19-Stance dataset. The performance is reported in terms of accuracy(Acc), precision(P), recall(R), and F1 score(F1). We highlight the best results in bold.

Model	**Wearing a Face Mask**				**Stay At Home Orders**			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
BiLSTM	57.80	56.90	58.00	56.70	73.50	67.90	64.00	64.50
CT-BERT	81.00	81.80	80.30	80.30	84.30	81.60	78.80	80.00
CT-BERT-v2	81.25	80.49	81.99	80.13	86.00	82.56	88.00	84.78
EDA	81.50	79.77	78.61	79.07	85.50	81.96	84.50	83.09
ASDA	82.50	80.96	80.24	80.53	87.00	83.04	85.09	83.99
SEGP	**84.50**	**83.20**	**83.49**	**83.34**	**89.00**	**86.33**	**89.37**	**87.71**
Model	**Anthony S. Fauci, M.D.**				**Keeping Schools Closed**			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
BiLSTM	63.80	63.90	63.10	63.00	62.70	57.00	54.50	54.80
CT-BERT	81.70	81.60	83.00	81.80	77.20	76.50	76.10	75.50
CT-BERT-v2	80.25	80.16	81.36	80.42	81.00	78.81	79.14	78.85
EDA	80.50	80.82	81.01	80.55	83.00	80.92	81.66	80.98
ASDA	81.00	81.49	81.04	81.06	83.50	81.29	80.95	81.01
SEGP	**82.50**	**82.60**	**82.57**	**82.57**	**86.00**	**84.04**	**84.45**	**84.23**

4.3 Analysis of Stance-Emotion joint Data Augmentation

We conduct experiments to prove the following two points: (1) the effectiveness of introducing emotional information into data augmentation; (2) the effectiveness of introducing emotional information through different types of auxiliary sentences.

In order to prove the first point, we compare the results of Stance-Emotion joint Data Augmentation (SEDA) with ASDA, which does not take emotional information into account. We present several augmented samples generated by these two methods in Table 4. It can be observed that the generated words of SEDA are more consistent with the label information. Furthermore, according

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

to the experimental results in Table 5, SEDA outperforms ASDA on all targets, which further demonstrates the validity of emotional information.

Table 4. Examples generated by ASDA and SEDA. Italicized texts represent generated words.

Target	Wearing a Face Mask
Source	In the USA, Walmart will now serve mask-less customers. Hopefully the same will happen in the UK .
ASDA	In the USA, Walmart will *today* serve mask-less customers. Hopefully the *fight* will *spread* in the UK.
SEDA	In the USA, Walmart will now serve mask-less customers. Hopefully the same will happen *sooner to* the *globe*.

In order to prove the second point, we compare the results of using different auxiliary sentences. The auxiliary sentences are constructed based on the relations between stance and emotion. "Consistency only" means we only use the "Consistency" relation between stance and emotion to introduce emotional information, thus SEDA(Consistency only) only contains the auxiliary sentence: The following texts have $\{s_i\}$ stance and $\{e_i\}$ emotion to $\{a_i\}$. "Discrepancy only" means we only use the "Discrepancy" relation, thus SEDA(Discrepancy only) only contains: Although the emotion is $\{e_i\}$, the following texts are both $\{s_i\}$ to $\{a_i\}$. SEDA is what we propose in this paper, which introduces emotional information based on "Consistency", "Discrepancy" and "None" relations. Therefore, as shown in Table 2, SEDA contains three types of auxiliary sentences. The experimental results in Table 5 show the performance impact of different auxiliary sentences, we can see that SEDA performs the best, indicating the effectiveness of the way we introduce emotional information.

Table 5. Performance comparison of introducing emotional information in different ways. We highlight the best results in bold.

Model	Wearing a Face Mask				Stay At Home Orders			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
ASDA	82.50	80.96	80.24	80.53	87.00	83.04	85.09	83.99
SEDA(Consistency only)	81.50	79.57	80.18	79.83	86.50	83.05	86.40	84.51
SEDA(Discrepancy only)	80.50	78.98	77.74	78.25	86.00	82.33	86.68	84.17
SEDA	**83.50**	**82.50**	**82.26**	**82.36**	**87.50**	**84.51**	**86.32**	**85.36**
Model	**Anthony S. Fauci, M.D.**				**Keeping Schools Closed**			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
ASDA	81.00	81.49	81.04	81.06	83.50	81.29	80.95	81.01
SEDA(Consistency only)	80.00	79.93	81.17	80.32	83.50	80.89	81.51	81.14
SEDA(Discrepancy only)	80.50	80.31	81.66	80.79	82.00	80.38	81.77	80.85
SEDA	**82.00**	**82.11**	**82.20**	**82.09**	**85.50**	**83.79**	**83.09**	**83.40**

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

4.4 Analysis of Gradual Prompt-tuning

We further explore the effectiveness of curriculum learning by comparing SEGP with SEP, which does not use the training strategy of curriculum learning. Curriculum learning requires a series of training samples with different disturbance degrees. In our method, the disturbance degree is determined by the mask ratio α in augmentation stage. Therefore, the artificial curriculums in the training samples are created according to α. Experimental results are shown in Table 6, which indicates that we can further improve performance by combining prompt-tuning with curriculum learning.

Table 6. Performance comparison of applying different training strategies. We highlight the best results in bold.

Model	Wearing a Face Mask				Stay At Home Orders			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
SEP	83.50	82.50	82.26	82.36	87.50	84.51	86.32	85.36
SEGP	**84.50**	**83.20**	**83.49**	**83.34**	**89.00**	**86.33**	**89.37**	**87.71**

Model	Anthony S. Fauci, M.D.				Keeping Schools Closed			
	Acc	**P**	**R**	**F1**	**Acc**	**P**	**R**	**F1**
SEP	82.00	82.11	82.20	82.09	85.50	83.79	83.09	83.40
SEGP	**82.50**	**82.60**	**82.57**	**82.57**	**86.00**	**84.04**	**84.45**	**84.23**

5 Conclusion

In this paper, we propose SEGP to address the scarcity of annotations problem in stance detection. SEGP is mainly composed of two stages, i.e., Stance-Emotion joint Data Augmentation (SEDA) and Gradual Prompt-tuning. With the help of C-MLM task, SEDA generates target-relevant and label-compatible samples by predicting the masked word conditioned on both their context and the auxiliary sentences. Gradual Prompt-tuning can make better use of the augmented samples as well as the knowledge contained in pre-trained models. The experimental results show that SEGP obtains superior performance over all baseline methods. Since our methods are not designed for a certain model, we will investigate how to extend them to other tasks in the future.

References

1. AlDayel, A., Magdy, W.: Stance detection on social media: State of the art and trends. Information Processing & Management **58**(4), 102597 (2021)
2. Allaway, E., McKeown, K.: Zero-shot stance detection: A dataset and model using generalized topic representations. arXiv preprint arXiv:2010.03640 (2020)
3. Augenstein, I., Rocktäschel, T., Vlachos, A., Bontcheva, K.: Stance detection with bidirectional conditional encoding. arXiv preprint arXiv:1606.05464 (2016)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

4. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th annual international conference on machine learning. pp. 41–48 (2009)
5. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. arXiv preprint arXiv:2005.14165 (2020)
6. Chauhan, D.S., Kumar, R., Ekbal, A.: Attention based shared representation for multi-task stance detection and sentiment analysis. In: International Conference on Neural Information Processing. pp. 661–669. Springer (2019)
7. Du, J., Xu, R., He, Y., Gui, L.: Stance classification with target-specific neural attention networks. International Joint Conferences on Artificial Intelligence (2017)
8. Garg, S., Ramakrishnan, G.: Bae: Bert-based adversarial examples for text classification. arXiv preprint arXiv:2004.01970 (2020)
9. Glandt, K., Khanal, S., Li, Y., Caragea, D., Caragea, C.: Stance detection in covid-19 tweets. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. vol. 1 (2021)
10. Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F., Liu, Q.: Tinybert: Distilling bert for natural language understanding. arXiv preprint arXiv:1909.10351 (2019)
11. Kenton, J.D.M.W.C., Toutanova, L.K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT. pp. 4171–4186 (2019)
12. Korbar, B., Tran, D., Torresani, L.: Cooperative learning of audio and video models from self-supervised synchronization. Advances in Neural Information Processing Systems **31** (2018)
13. Küçük, D., Can, F.: Stance detection: A survey. ACM Computing Surveys (CSUR) **53**(1), 1–37 (2020)
14. Li, Y., Caragea, C.: 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). pp. 6299–6305 (2019)
15. Li, Y., Caragea, C.: A multi-task learning framework for multi-target stance detection. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. pp. 2320–2326 (2021)
16. Li, Y., Caragea, C.: Target-aware data augmentation for stance detection. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 1850–1860 (2021)
17. Li, Y., Cohn, T., Baldwin, T.: Robust training under linguistic adversity. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. pp. 21–27 (2017)
18. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. arXiv preprint arXiv:2107.13586 (2021)
19. Miller, G.A.: Wordnet: a lexical database for english. Communications of the ACM **38**(11), 39–41 (1995)
20. Mohammad, S.M., Sobhani, P., Kiritchenko, S.: Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT) **17**(3), 1–23 (2017)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48

21. Mueller, J., Thyagarajan, A.: Siamese recurrent architectures for learning sentence similarity. In: Proceedings of the AAAI conference on artificial intelligence. vol. 30 (2016)
22. Müller, M., Salathé, M., Kummervold, P.E.: Covid-twitter-bert: A natural language processing model to analyse covid-19 content on twitter. arXiv preprint arXiv:2005.07503 (2020)
23. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). pp. 1532–1543 (2014)
24. Reynolds, L., McDonell, K.: Prompt programming for large language models: Beyond the few-shot paradigm. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. pp. 1–7 (2021)
25. Schuster, M., Paliwal, K.K.: Bidirectional recurrent neural networks. IEEE transactions on Signal Processing **45**(11), 2673–2681 (1997)
26. Sobhani, P., Inkpen, D., Zhu, X.: A dataset for multi-target stance detection. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. pp. 551–557 (2017)
27. Somasundaran, S., Wiebe, J.: Recognizing stances in ideological on-line debates. In: Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text. pp. 116–124 (2010)
28. Tsvetkov, Y., Faruqui, M., Ling, W., MacWhinney, B., Dyer, C.: Learning the curriculum with bayesian optimization for task-specific word representation learning. arXiv preprint arXiv:1605.03852 (2016)
29. Walker, M., Anand, P., Abbott, R., Grant, R.: 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. pp. 592–596 (2012)
30. Wang, W.Y., Yang, D.: That's so annoying!!!: A lexical and frame-semantic embedding based data augmentation approach to automatic categorization of annoying behaviors using# petpeeve tweets. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 2557–2563 (2015)
31. Wang, X., Pham, H., Dai, Z., Neubig, G.: Switchout: an efficient data augmentation algorithm for neural machine translation. arXiv preprint arXiv:1808.07512 (2018)
32. Wei, J., Huang, C., Vosoughi, S., Cheng, Y., Xu, S.: Few-shot text classification with triplet networks, data augmentation, and curriculum learning. arXiv preprint arXiv:2103.07552 (2021)
33. Wei, J., Zou, K.: Eda: Easy data augmentation techniques for boosting performance on text classification tasks. arXiv preprint arXiv:1901.11196 (2019)
34. Weinshall, D., Cohen, G., Amir, D.: Curriculum learning by transfer learning: Theory and experiments with deep networks. In: International Conference on Machine Learning. pp. 5238–5246. PMLR (2018)
35. Wu, X., Lv, S., Zang, L., Han, J., Hu, S.: Conditional bert contextual augmentation. In: International Conference on Computational Science. pp. 84–95. Springer (2019)
36. Zhang, B., Yang, M., Li, X., Ye, Y., Xu, X., Dai, K.: Enhancing cross-target stance detection with transferable semantic-emotion knowledge. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 3188–3197 (2020)
37. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. Advances in neural information processing systems **28**, 649–657 (2015)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_48