

SGPT: Semantic Graphs based Pre-training for Aspect-based Sentiment Analysis

qian yong

chen chen

zhong qin wang (✉ wangzq870305@gmail.com)

rong xiao

haihong tang

Research Article

Keywords: sentiment analysis, pre-trained language model, aspect sentiment analysis, Semantic Graphs

Posted Date: October 10th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2112315/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

SGPT: Semantic Graphs based Pre-training for Aspect-based Sentiment Analysis

Yong Qian[†], Chen Chen[‡], Zhongqing Wang[‡], Rong Xiao[†], Haihong Tang[†]

sarahyongq@gmail.com
20195227080@suda.edu.cn
wangzq870305@gmail.com
{xiaorong.xr,piaoxue}@taobao.com

Abstract

Previous studies show effective of pre-trained language models for sentiment analysis. However, most of these studies ignore the importance of sentimental information for pre-trained models. Therefore, we fully investigate the sentimental information for pre-trained models and enhance pre-trained language models with semantic graphs for sentiment analysis. In particular, we introduce Semantic Graphs based Pre-training(SGPT) using semantic graphs to obtain synonym knowledge for aspect-sentiment pairs and similar aspect/sentiment terms. We then optimize the pre-trained language model with the semantic graphs. Empirical studies on several downstream tasks show that proposed model outperforms strong pre-trained baselines. The results also show the effectiveness of proposed semantic graphs for pre-trained model.

keywords: sentiment analysis, pre-trained language model, aspect sentiment analysis, Semantic Graphs

1 Introduction

Pre-trained language models learn contextualized word representations on large-scale text corpus through a self-supervised learning method, which are fine-tuned on downstream tasks and can obtain the state-of-the-art (SOTA) performance, and are pervasive and have made a tremendous impact in many NLP fields such as reading comprehension (Lai et al., 2017), question answering (Rajpurkar et al., 2016) and sentiment analysis (Zhang et al., 2018). Leveraging pre-trained language models have achieved promising results in sentiment analysis tasks, including aspect-level sentiment classification (Zeng et al., 2019a), sentence-level sentiment classification (Zhang et al., 2018). These pre-trained models have shown their power in learning general semantic representations from large-scale unlabelled corpora via well-designed pre-trained tasks.

Sentiment analysis involves a wide range of specific tasks (Liu, 2012), such as sentence-level sentiment classification, aspect-level sentiment classification, aspect term extraction, and so on. Sentiment analysis tasks usually depend on different types of sentiment knowledge including sentiment words, word polarity, and aspect-sentiment pairs (Tian et al., 2020). Recently, knowledge has been shown very important for enhancing the language representations, such as SentiLARE (Ke et al., 2020) and SKEP (Tian et al., 2020).

Sentiment analysis especially on large-scale reviews is still very challenging, since it is hard to capture the aspect term, sentiment words from the review text. As shown in the below example, since there are more than one aspect in the example, the traditional pre-trained model cannot capture the sentiment information. In addition, the masking language model of traditional pre-trained model will ignore the continue aspect-opinion phrase (e.g., great color). However, based on the semantic graphs with the correlations between aspect and opinion, it is easy for proposed model to capture sentiment information, and solve the continue phrase masking problem.

The cloth is overall good, with great **color**, but bad **material**.

To address the above challenges, we develop a semantic graph-based pre-training model to employ semantic graphs with sentiment knowledge for pre-trained model. In particular, we explore similar aspect and sentiment words, and build a similar semantic and aspect-sentiment pair graph. We then employ aspect-sentiment pairs to construct the semantic graph. Thirdly, we feed the semantic graph into a the pre-trained language model with sentimental masking. Finally, we jointly optimize the aspect-sentiment pair prediction objective and mask language model. Empirical studies on several downstream tasks show that proposed model out-

performs strong pre-trained baselines. The results also show the effectiveness of proposed semantic graphs for pre-trained model. In summary, our contributions are as follows:

(1) We employ semantic graphs with aspect and sentiment terms to enhance pre-trained language models. The results outperforms state-of-the-art models on several sentiment analysis tasks in both Chinese and English datasets.

(2) Our method significantly outperforms the strong pre-training method RoBERTa (Liu et al., 2019) on three typical sentiment tasks, and achieves much better results on all the datasets.

2 Overview of Proposed Model

In this study, we propose a semantic graph-based pre-trained model to construct the semantic graphs from aspect terms and sentiment words for the pre-trained language model.

As shown in Figure 1, we firstly construct a **semantic graph** with pair-wise aspect and sentiment term relations, and similarity relations among aspect and sentiment terms. We then employ **pre-train a language model** to learn sentiment knowledge from the semantic graphs with three tasks including sentiment masking prediction, aspect-sentiment pair prediction and node similarity. Finally, we fine-tune the pre-training model on three **sentiment analysis tasks**: sentence-level sentiment classification, aspect-level sentiment classification and aspect/sentiment terms extraction.

3 Semantic Graphs based Pre-training

As shown in Figure 2, we employ semantic graphs to capture sentiment knowledge for pre-trained language model. Our paradigm contains three sentiment pre-training objectives: sentiment words masking prediction L_{sw} , aspect-sentiment pair prediction L_{ap} and aspect-based similarity score L_{ns} .

Given an input sentence, **sentimental masking prediction** attempts to recover the sentiment words masked based on the semantic graph. **Aspect-sentiment pairs similarity** aims to calculate the matching rate of a sentiment description on an aspect along with some related words sampled from the semantic graph. We extend the related words to the **aspect-based similarity score** and further learn the synonym knowledge from another perspective.

Therefore, these three tasks are joint learning to continue pre-train the language model:

$$L = L_{sw} + L_{ap} + L_{ns} \quad (1)$$

3.1 Semantic Graphs Construction

We construct a semantic graph from large-scale unlabeled data. In particular, we extract the aspect words, sentiment words, and aspect-sentiment pairs from the unlabeled data. Our work is mainly based on automation methods along with a slightly manual review.

This is a heterogeneous graph with aspect words and sentiment words as different type nodes. Two nodes are connected if they are semantic or literally similar, aspect-sentiment word pair is also connected. Our method aims to integrate those knowledge into pre-trained language model.

3.2 Sentiment Word Prediction

Inspired by BERT (Devlin et al., 2018) that randomly replaces 15% words with [MASK] and learn to recover them, we attempt recovering masked sentiment words to pay more attention to sentiment descriptions. For sentiment words prediction, each token in a masked sentence \tilde{X} is fed into roBERTa to get a vector representation $\tilde{\mathbf{x}}_i$ and then normalized with softmax layer to produce a probability vector $\hat{\mathbf{y}}_i$ over the entire vocabulary. In this way, the sentiment word prediction objective L_{sw} is to maximize the probability of original sentiment word x_i as follows:

$$\hat{\mathbf{y}}_i = \text{softmax}(\tilde{\mathbf{x}}_i W + b) \quad (2)$$

$$L_{sw} = - \sum_i m_i \times \mathbf{y}_i \log \hat{\mathbf{y}}_i \quad (3)$$

Here, $\mathbf{W} \in R^{d \times v}$ and $\mathbf{b} \in R^{d \times 1}$ are all trainable parameters of the prediction layer. $m_i = 1$ when x_i is masked, otherwise it equals 0. \mathbf{y}_i is the one-hot representation of the original token x_i .

3.3 Aspect-Sentiment Pair Prediction for Pre-Training

We propose a new pre-training task to build these dependency between aspect and sentiment terms. Compared with predicting pairs directly, we calculate a pair prediction over the aspect-sentiment pairs enhanced by their similar words. The is a value between 0 to 1 which means the probability of a pair existing. For an aspect-sentiment pair, we extract a similar aspect words set SA and a similar sentiment words set SS from the semantic graph with algorithm 1. We concatenate SA and SS to construct the input sequence:

$$X_{SA \oplus SS} = [CLS, SA, SEP, SS, SEP] \quad (4)$$

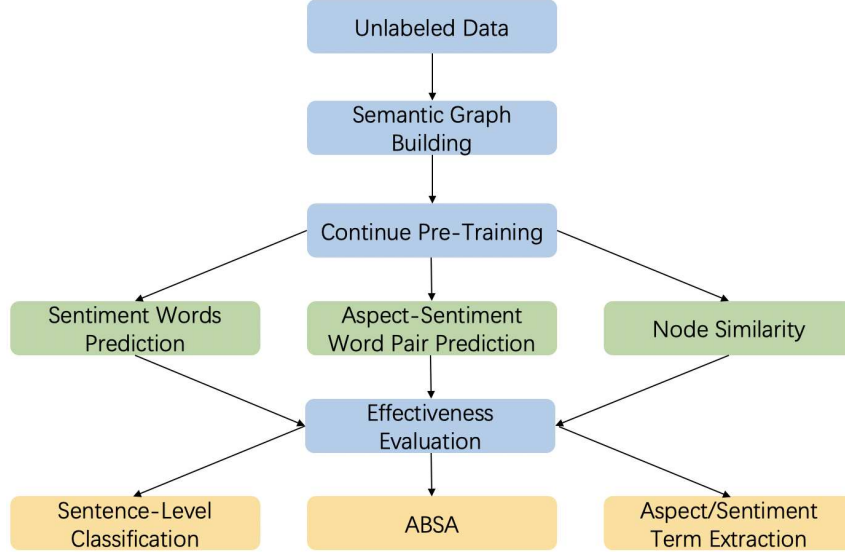


Figure 1: The Overview of proposed models.

where, CLS denotes the entire sequence representation, while SEP is a separator of two sequence. After encoding each element with roBERTa, we use u_{cls} as the embedding of CLS to calculate the pair prediction.

$$\hat{\mathbf{p}}_i = \text{sigmoid}(\mathbf{u}_{cls}W_p + b_p) \quad (5)$$

We expect every aspect-sentiment pair equals 1. Thus, the pair prediction is:

$$L_{ap} = - \sum_{i=1}^2 \mathbf{p}_i \log \hat{\mathbf{p}}_i \quad (6)$$

Where p_i always equals 1 when the input sequence is a pair while p_i equals 0 if the input sequence is not a pair. Then the relation between aspect words and sentiment words will be established in this way.

3.4 Node Similarity for Pre-training

Semantic graph node similarity aim to capture model’s sensitivity to aspect/sentiment synonyms. In particular, we sample synonyms from the semantic graph with similar nodes sampling algorithm detailed in Algorithm 1 and get SA and SS . All aspect synonyms are fed into roBERTa to get a representation U_{SA} , so dose sentiment synonyms and gets a representation U_{SS} . As contrastive learning has shown great success on many areas especially unsupervised methods, we apply contrastive learning to capture these synonyms relations. The core idea of contrastive learning is to shorten the distance of positive samples and widen the distance

of negative samples which perfectly meets our requirements.

$$\text{score}(f(x), f(x^+)) \gg \text{score}(f(x), f(x^-)) \quad (7)$$

What we want to do is clustering each similar aspect word together so all synonyms are positive samples and randomly sampled words are negative samples. We use a cosine function to measure the distance between samples and get the loss L_{ns} as following:

$$\text{score}(f(x), f(x^+)) = \cos(f(x), f(x^+)) \quad (8)$$

$$L_{ns} = - \log \frac{e^{\text{score}(f(x), f(x^+))}}{e^{\text{score}(f(x), f(x^+))} + e^{\text{score}(f(x), f(x^-))}} \quad (9)$$

3.5 Sentiment Analysis with Pre-training Models

We verify the effectiveness of our language model on three typical sentiment analysis tasks: sentence-level sentiment classification, aspect-based sentiment classification, and aspect/sentiment terms extraction. We fine-tune some strong models on the same language model as baseline to evaluate the improvement.

Sentence-level Sentiment Classification

This task aims to classify the sentiment polarity of an input sentence. We use the final state vector of classification token [CLS] as the overall representation of an input sentence and a classification layer

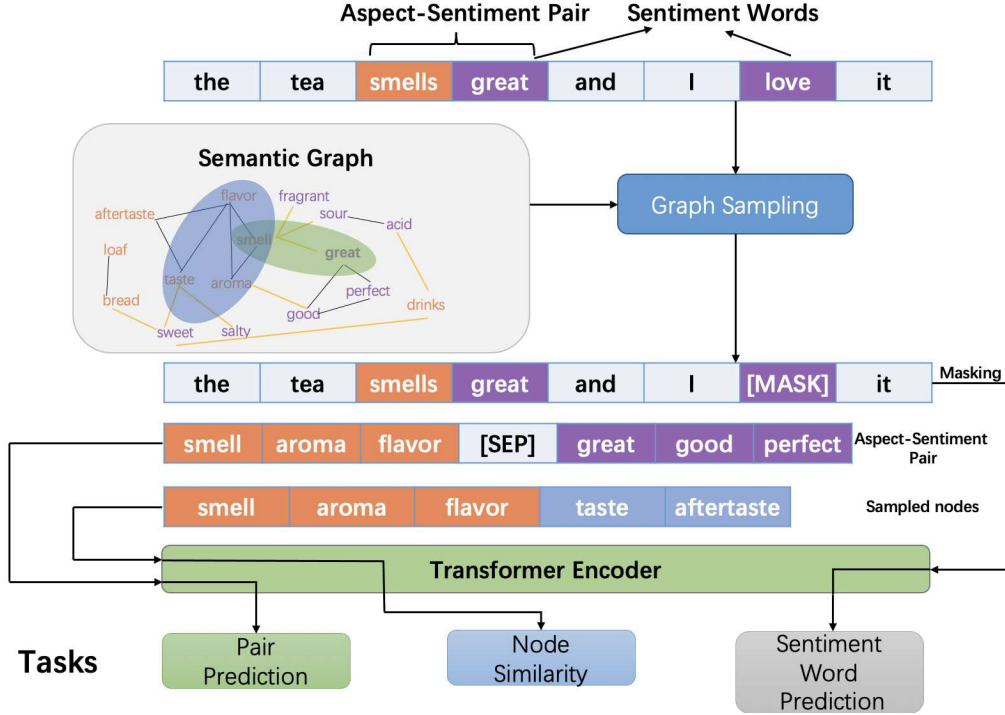


Figure 2: The Framework of our pre-training paradigm.

is added to calculate the sentiment probability on top of the transformer encoder.

Algorithm 1: Similar Nodes Sampling

Input: Graph G , initial node h , max sampling depth K , max sampling words length L , word frequency table T .

Output: Sampled similar nodes C_h .

- 1 $C_h \leftarrow []$;
 - 2 $S_0 \leftarrow [h], S_1, S_2, \dots, S_K \leftarrow []$;
 - 3 **for** $k = 1, 2, \dots, K$ **do**
 - 4 **for** $t \in S_{k-1}$ **do**
 - 5 Sample all nodes linked with t and append them to S_k ;
 - 6 **end**
 - 7 **end**
 - 8 $\hat{C}_h = S_0 \cap S_1 \cap S_2 \cap \dots \cap S_K$;
 - 9 Sort the nodes in \hat{C}_h by frequency according to table T in incremental order;
 - 10 Choose top-ranked nodes up to the length L and append them to C_h ;
 - 11 **Return** C_h ;
-

Aspect-based Sentiment Classification

The purpose of this task is to analyze fine-grained sentiment polarity for an aspect with a given con-

textual text. Thus, there are two parts in the input: contextual text and aspect description. We combine these two parts with a separator [SEP], and feed them into the language model. The final state of [CLS] also be utilized as the representation for classification.

Aspect and Sentiment Term Extraction

This task is to extracting all aspect description or sentiment statement. The same as other tasks, all tokens are fed into language model to get a representation. Then a CRF layer is added on each tokens to predict if it belongs to an aspect or sentiment term.

| Domain | | Train | Valid | Test |
|--------|-----|-------|-------|-------|
| Furn | POS | 37773 | 5395 | 10791 |
| | NEG | 1795 | 256 | 512 |
| Kith | POS | 25795 | 3712 | 7613 |
| | NEG | 2496 | 358 | 358 |

Table 1: Statistics of Chinese ABSA Evaluation Datasets. POS and NEG refer to positive polarity and negative polarity. There is an obviously unbalanced phenomenon, more than 10:1, between the positive and negative samples.

| Domain | Extraction | | | Classification | | |
|--------|------------|-------|------|----------------|-------|-------|
| | Train | Valid | Test | Train | Valid | Test |
| Furn | 7106 | 1014 | 2029 | 120295 | 17184 | 34369 |
| Kith | 6738 | 962 | 1925 | 92124 | 13160 | 26321 |

Table 2: Statistics of Chinese Evaluation Datasets on Extraction and Sentence-Level Classification. The extraction task includes aspect words and sentiment words which are aggregate.

| Dataset | Train | Dev | Test |
|----------|-------|------|------|
| SST-2 | 67K | 872 | 1821 |
| Amazon-2 | 3.2M | 400K | 400K |
| Sem-R | 3608 | - | 1120 |
| Sem-L | 2328 | - | 638 |
| MPQA2.0 | 287 | 100 | 95 |

Table 3: Statistics of English Evaluation Datasets. Sem-R and Sem-L refer to restaurant and laptop parts of SemEval 2014 Task 4.

| Parameter | Value |
|-----------------|-------|
| Dataset Size CN | 167M |
| Dataset Size EN | 3.2M |
| Masking Rate | 20% |
| Number of Pairs | 2 |
| Sentence Length | 512 |
| Learning Rate | 1e-5 |
| Batch Size | 32 |
| Warm up Ratio | 0.1 |
| Weight Decay | 0 |

Table 4: Pre-Training Parameters.

4 Experimentation

In this section, we introduce our training/evaluation datasets and some experiment setting. Then release the experimental results conducted from different perspectives and analyze the effectiveness compared with different baseline models.

4.1 Data Collections

We mainly develop with Chinese datasets and also pre-train an English version to evaluate the effectiveness on public datasets.

The Chinese data comes from product reviews on TaoBao.com which is one of the top online shopping platforms. The Chinese pre-training dataset includes over 167 million sentences and evaluate them on two domains, i.e., Furniture (Furn) and Kitchen (Kith). And all of the three tasks for Chinese are evaluated as Macro-F1 score. The statistics of Chinese evaluation datasets is shown in Table 1 and Table 2. We split all data as 7:1:2 and get the train/valid/test datasets.

The English dataset for pre-training is amazon-2 (Zhang et al., 2015), 3.2 million of the original training data are reserved for development. We evaluate the performance of the English model on a variety of English sentiment analysis datasets. Table 3 summarizes the statistics of English datasets used in the evaluations. Different tasks are evaluated on different datasets: (1) For sentence-level sentiment classification, Standford Sentiment Treebank (SST-2) (Socher et al., 2013) and Amazon-2

(Zhang et al., 2015) are used. The performance is evaluated in terms of accuracy. (2) Aspect based sentiment classification is evaluated on Semantic Eval 2014 Task4 (Maria et al., 2014). This task contains both restaurant domain and laptop domain, whose accuracy is evaluated separately. (3) For the extraction task, MPQA 2.0 (Wiebe et al., 2005; Wilson, 2008) dataset is used which aims to extract the aspects or the holders of the sentiments. We measured with the method in SRL4ORL (Marasovic and Frank, 2017), which is released and available online.

4.2 Data Pre-Processing

To build the semantic graph, we extract these information as following:

Aspect/Sentiment Term Extraction Aspect and Sentiment descriptions are extracted by BERT-CRF which is trained on labeled datasets up to 85 F1 score.

Aspect-Sentiment Pair Extraction We match the aspect-sentiment pairs with simple constraints. An aspect-sentiment pair refers to the mention of an aspect and corresponding sentiment words. Thus, a sentiment word with its nearest aspect has high probability to be a pair. More specifically, we limit the aspect-sentiment pair must be included in one sentence and only one-to-one pairs are considered.

Similar Words Extraction As similar words

| Model | Aspect-Level | | Sentence-Level | | Extraction | | | |
|-----------------|--------------|-------------|----------------|-------------|-------------|-------------|-------------|-------------|
| | Furn | Kith | Furn | Kith | Furn | | Kith | |
| | M-F1 | M-F1 | M-F1 | M-F1 | Aspect | Sentiment | Aspect | Sentiment |
| BERT | 90.9 | 91.5 | 90.2 | 93.5 | 79.9 | 82.0 | 77.2 | 79.2 |
| RoBERTa | 94.1 | 94.3 | 92.8 | 95.3 | 81.4 | 83.1 | 78.6 | 81.6 |
| SKEP | 94.6 | 94.9 | 92.2 | 95.5 | 82.7 | 84.1 | 80.3 | 83.3 |
| Pair Prediction | 95.2 | 95.8 | 92.6 | 94.1 | 84.2 | 85.2 | 82.8 | 84.8 |
| Node Similarity | 95.4 | 95.9 | 93.2 | 96.4 | 84.8 | 86.9 | 83.0 | 84.7 |
| Ours | 96.2 | 96.7 | 93.2 | 96.8 | 85.0 | 87.3 | 83.8 | 85.0 |

Table 5: Results of Chinese Evaluation on Extraction and Sentence-Level Classification. Model BERT and RoBERTa refer to fine-tune directly on downstream tasks. Pair prediction refer to our pre-training methods with sentiment masking and pair prediction. Node similarity means sentiment masking with node similarity. Ours means the complete method we propose. M-F1 is the abbreviation of macro-F1 score.

| Model | Aspect-Level | | Sentence-Level | | Extraction | |
|-----------------|--------------|--------------|----------------|--------------|---------------------|--------------------|
| | Sem-L | Sem-R | SST-2 | Amazon-2 | MPQA-Holder | MPQA-Target |
| Previous SOTA | 81.35 | 87.89 | 97.10 | 97.37 | 83.67/77.12 | 81.59/73.16 |
| SKEP | 81.32 | 87.92 | 96.70 | 96.94 | 84.25/79.03 | 82.77/74.82 |
| Pair Prediction | 82.46 | 88.79 | 97.18 | 98.02 | 85.31/80.89 | 83.48/76.43 |
| Node Similarity | 84.31 | 89.04 | 98.15 | 98.25 | 86.42 /81.78 | 84.69/77.01 |
| Ours | 84.45 | 89.03 | 98.27 | 98.33 | 86.16/ 82.05 | 84.79/77.56 |

Table 6: Results of English Evaluation

can be seen as words with a same category, we employ DBSCAN clustering algorithm to get coarse-grained synonyms represented by the average pooling of all piece words’ Word2Vec embeddings. A recycling mechanism is applied to further clustering big clusters, which contain lots of irrelevant words, into small clusters by grid searching different parameters. Finally, we review all similar words manually to get more accurate synonyms.

4.3 Experiment Setting

We use RoBERTa as the base language model and continue pre-training it with our paradigm. We concatenate different sentences until the sentence length up to 512 and train them with batch size 32. An adam optimizer is applied with learning rate 1e-5 and warmup ratio 0.1 without weight decay. For sentiment masking, we mask sentiment words as far as possible up to 20% and two aspect-sentiment pairs at most needed to be extracted. At the stage of fine-tune for downstream tests, we take the same parameters as SKEP (Tian et al., 2020) did.

4.4 Comparison with Baselines

We compare SGPT with two strong pre-training baselines: RoBERTa and a similar continue pre-

training method SKEP.

The results on Chinese datasets are shown in table 5. For classification tasks, there is no obvious improvement from RoBERTa to SKEP while SGPT outperforms SKEP 1.6 and 1.8 on aspect-level, along with 1.0 and 1.3 on sentence-level. Meanwhile both pair prediction and node similarity performs better than SKEP. More details, SGPT overcomes the unbalanced problem with average 3 point improvements on negative samples. For extraction task, sentiment continuing pre-training’s effectiveness is also significantly. SKEP has a 1-2 point improvement compared with RoBERTa and SGPT has a further 2-3 point improvement based on SKEP.

Meanwhile, English experiments are shown in table 3. SGPT outperforms other methods over three task as well as in Chinese datasets and indicates that SGPT is universal to cover different language and different tasks.

4.5 Influence of Different Factors

We then analyze the influence of different factors with both pair prediction and node similarity separately to verify the influence factors on Table 3.

Effective of Aspect-Sentiment Pair Prediction

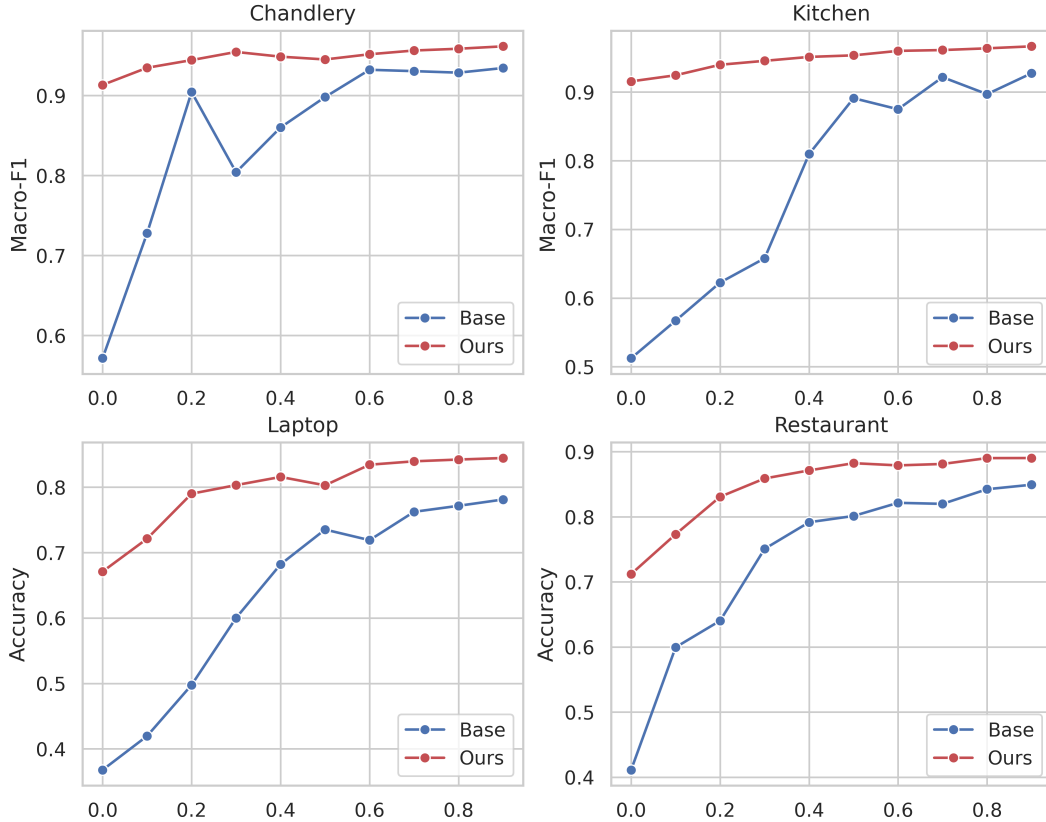


Figure 3: The Results under Different Data Scales.

For aspect based sentiment analysis, we propose the aspect-sentiment pair prediction to build the dependence between aspect words and sentiment words. The results reflect that continuing pre-training language model with pair prediction can absorb pairs information successfully and adapt great on downstream task related to aspect or sentiment.

Effective of Node Similarity The node similarity aims to learn the synonym knowledge from semantic graphs. Compared with aspect-sentiment pair prediction, the experiment results shows that it almost outperforms far more than all pair prediction’s results except aspect term extraction. We think node similarity is an even more powerful mechanism for gaining sentiment information and contrastive learning also plays a great role in this.

Finally, we combine the two factors into one model. The results shows that our composite method benefits from the two with different advantages and get a better performance than both of them separately.

4.6 Influence of Training Size

To verify how capable SGPT solves the unbalance label problem and few-shoot problems, we design an experiment which increases the data scale from 10% to 100%, then compare the effectiveness between SGPT and fine-tuning directly. As shown in figure 3, SGPT has great performance even on 10% of all training data and gets more than 30 point improvement compared with the base model. The base model is close to SGPT until 60% of data has been used but still has a big gap with SGPT.

4.7 Case Study

We give two aspect based sentiment analysis examples in table 7 and illustrate the situation where RoBERT or SKEP can’t solve but can be overcome by SGPT.

The first example is about unconventional expression. For most of reviews, "cost" usually means needing more and SKEP makes the decision from this inertia thinking. While SGPT takes more words into consideration that linked in the semantic graph and recognizes "costs less" is a positive expression.

The second example is fine-grain sentiment po-

| ID | Review | Sentiment Polarity | RoBERTa | SKEP | Ours |
|----|--|--------------------|---------|------|------|
| 1 | This air conditioner costs <i>less electricity</i> and it's refrigeration is very good | POS | NEG | NEG | POS |
| 2 | Perfect! Although the <i>price</i> is high , the quality is very good. | NEG | POS | POS | NEG |

Table 7: Case Study. Italics words are aspect terms and bold words are sentiment descriptions.

larity contrary to the sentence-level polarity. The pair prediction task in SKEP can't identify every aspect-sentiment pair when the sentiment description is intensive so that it assesses "price" from a holistic perspective. While SGPT benefits from pair prediction without predicting all pairs and recognizes the real sentiment polarity.

5 Related Works

5.1 Sentiment Analysis with Knowledge

Various types of sentiment knowledge, including sentiment words, aspect-sentiment pairs and prior sentiment polarity from Senti WordNet (Ke et al., 2020), have been proved to be useful for a wide range of sentiment analysis tasks. Sentiment words with their polarity are widely used for sentiment analysis, including sentence level sentiment classification (Taboada et al., 2011b; Shin et al., 2017; Zhang et al., 2018), aspect-level sentiment classification (Vo and Zhang, 2015; Zeng et al., 2019a), sentiment extraction (Li and Lam, 2017), emotion analysis (Gui et al., 2017; Fan et al., 2019) and so on. Lexicon-based method (Turney, 2002; Taboada et al., 2011a) directly utilizes polarity of sentiment words for classification. Traditional feature-based approaches encode sentiment word information in manually-designed features to improve the supervised models (Bakshi et al., 2016; Agarwal et al., 2011). In contrast, deep learning approaches enhance the embedding representation with the help of sentiment words (Shin et al., 2016), or absorb the sentiment knowledge through linguistic regularization (Qian et al., 2016). Aspect-sentiment pair knowledge is also useful for aspect-level classification and sentiment extraction. Previous works often provide weak supervision by this type of knowledge, either for aspect level classification (Zeng et al., 2019b) or for sentiment extraction (Yang et al., 2017; Ding et al., 2017). One related study is SKEP (Tian et al., 2020), which utilize sentiment knowledge to embed sentiment information at the word, polarity and aspect level into pre-trained sentiment representation. But it's hard to re-

cover aspect-sentiment pair due to aspect-sentiment pair appear continuously(85% probability) in product review(product reviews in tabao.com).

5.2 Knowledge Enhanced PLMs

Recently, to enable PLMs with world knowledge, several attempts (Wang et al., 2019; Peters et al., 2019; Zhang et al., 2019; Liu et al., 2020; Wang et al., 2020) have been made to inject knowledge into BERT leveraging Knowledge Graphs (KGs). Most of these work adopts the "BERT+entity linking" paradigm, whereas, it is not suitable for E-commerce product reviews due to the lack of quality entity linkers as well as KGs in this domain. SKEP conducts aspect-sentiment pair masking, sentiment word masking, common token masking and utilize three sentiment knowledge prediction objectives, with sentiment word prediction, word polarity prediction and aspect-sentiment pair prediction and aspect-sentiment pairs is converted into multi-label classification. However, aspect-sentiment word appear continuously in product reviews, which make it difficult to predict when masking the pair. Our work also differs from the work skep, we develop a novel pre-training paradigm that leverage semantic graphs and incorporate sentiment knowledge into pretraining, a detailed comparison between our model and skep pre-trained language models can be found in §3.

6 Conclusion

In this study, we design three pre-training tasks to continue pre-training language model for downstream sentiment analysis tasks. We propose a semi automatic method to build a semantic graph and employee language model to adopt the graph knowledge with these tasks. Sentiment words masking for paying more attention to sentiment term, aspect-sentiment pair prediction for building the dependence between aspect and sentiment, node similarity for learning synonym knowledge, which all get great improvement over different downstream tasks and different languages.

In the future, we will try to apply SGPT on more sentiment analysis tasks, to further see the generalization of SGPT, and we are also interested in exploiting more efficiency method of building semantic knowledge.

7 Declarations

- Ethical Approval: not applicable
- Competing interests: not applicable
- Authors' contributions: Yong qian, chen chen and Zhongqing Wang wrote the main manuscript text and Yong Qian did all experiments. Chen chen prepared figures 1-3. All authors reviewed the manuscript.
- Funding: not applicable
- Availability of data and materials: All the data can be found online.
SST-2:<http://nlp.stanford.edu/sentiment>
MPQA2:http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/mpqa_corpus_2_0/
Amazon-2:<https://snap.stanford.edu/data/web-Amazon.html>
SRL4ORL:<http://alt.qcri.org/semeval2014/task4/data/uploads/>
Taobao dataset cannot be provided at present because Alibaba has not yet approved, once approved we will release dataset.

References

- Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow, and Rebecca J Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)*, pages 30–38.
- Rushlene Kaur Bakshi, Navneet Kaur, Ravneet Kaur, and Gurpreet Kaur. 2016. Opinion mining and sentiment analysis. In *2016 3rd international conference on computing for sustainable global development (INDIACom)*, pages 452–455. IEEE.
- Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Ying Ding, Jianfei Yu, and Jing Jiang. 2017. Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.
- Chuang Fan, Hongyu Yan, Jiachen Du, Lin Gui, Li-dong Bing, Min Yang, Ruifeng Xu, and Ruibin Mao. 2019. A knowledge regularized hierarchical approach for emotion cause analysis. 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 5618–5628.
- Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Qin Lu, and Jiachen Du. 2017. A question answering approach to emotion cause extraction. *arXiv preprint arXiv:1708.05482*.
- Pei Ke, Haozhe Ji, Siyang Liu, Xiaoyan Zhu, and Minlie Huang. 2020. Sentilare: Linguistic knowledge enhanced language representation for sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6975–6988.
- Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. 2017. Race: Large-scale reading comprehension dataset from examinations.
- Xin Li and Wai Lam. 2017. Deep multi-task learning for aspect term extraction with memory interaction. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2886–2892.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.
- Weijie Liu, Peng Zhou, Zhe Zhao, Zhiruo Wang, Qi Ju, Haotang Deng, and Ping Wang. 2020. K-bert: Enabling language representation with knowledge graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2901–2908.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Ana Marasovic and Anette Frank. 2017. SRL4ORL: improving opinion role labelling using multi-task learning with semantic role labeling. *CoRR*, abs/1711.00768.
- Pontiki Maria, Galanis Dimitris, Pavlopoulos John, Papanageorgiou Harris, Androutsopoulos Ion, and Manandhar Suresh. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 27–35.
- Matthew E Peters, Mark Neumann, Robert L Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A Smith. 2019. Knowledge enhanced contextual word representations. *arXiv preprint arXiv:1909.04164*.

- Qiao Qian, Minlie Huang, Jinhao Lei, and Xiaoyan Zhu. 2016. Linguistically regularized lstms for sentiment classification. *arXiv preprint arXiv:1611.03949*.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. [Squad: 100,000+ questions for machine comprehension of text](#).
- Bonggun Shin, Timothy Lee, and Jinho D Choi. 2016. Lexicon integrated cnn models with attention for sentiment analysis. *arXiv preprint arXiv:1610.06272*.
- Bonggun Shin, Timothy Lee, and Jinho D. Choi. 2017. Lexicon integrated cnn models with attention for sentiment analysis. *Alexandra Balahur Dobrescu*, pages 149–158.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. 2011a. [Lexicon-based methods for sentiment analysis](#). *Computational Linguistics*, 37(2):267–307.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly D. Voll, and Manfred Stede. 2011b. Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.
- Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. Skep: Sentiment knowledge enhanced pre-training for sentiment analysis. *arXiv preprint arXiv:2005.05635*.
- Peter D Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.
- Duy-Tin Vo and Yue Zhang. 2015. Target-dependent twitter sentiment classification with rich automatic features. In *Twenty-fourth international joint conference on artificial intelligence*.
- Ruize Wang, Duyu Tang, Nan Duan, Zhongyu Wei, Xuanjing Huang, Cuihong Cao, Daxin Jiang, Ming Zhou, et al. 2020. K-adapter: Infusing knowledge into pre-trained models with adapters. *arXiv preprint arXiv:2002.01808*.
- Xiaozhi Wang, Tianyu Gao, Zhaocheng Zhu, Zhiyuan Liu, Juanzi Li, and Jian Tang. 2019. Kepler: A unified model for knowledge embedding and pre-trained language representation. *arXiv preprint arXiv:1911.06136*.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2):165–210.
- Theresa Ann Wilson. 2008. *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. Ph.D. thesis, University of Pittsburgh.
- Zhilin Yang, Ruslan Salakhutdinov, and William W Cohen. 2017. Transfer learning for sequence tagging with hierarchical recurrent networks. *arXiv preprint arXiv:1703.06345*.
- Ziqian Zeng, Wenxuan Zhou, Xin Liu, and Yangqiu Song. 2019a. A variational approach to weakly supervised document-level multi-aspect sentiment classification. In *Conference of the North*.
- Ziqian Zeng, Wenxuan Zhou, Xin Liu, and Yangqiu Song. 2019b. A variational approach to weakly supervised document-level multi-aspect sentiment classification. *arXiv preprint arXiv:1904.05055*.
- Lei Zhang, Shuai Wang, and Bing Liu. 2018. Deep learning for sentiment analysis : A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, page e1253.
- Xiang Zhang, Junbo Jake Zhao, and Yann LeCun. 2015. [Character-level convolutional networks for text classification](#). *CoRR*, abs/1509.01626.
- Zhengyan Zhang, Xu Han, Zhiyuan Liu, Xin Jiang, Maosong Sun, and Qun Liu. 2019. Ernie: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.