MedAI at SemEval-2021 Task 10: Negation-aware Pre-training for Source-free Negation Detection Domain Adaptation

Jinquan Sun, Qi Zhang, Yu Wang, Lei Zhang
Alibaba Group Inc.
{jinquan.sjq, mickey.zq, tonggou.wangyu}@alibaba-inc.com
lei.zhang.lz@alibaba-inc.com

Abstract

Due to the increasing concerns for data privacy, source-free unsupervised domain adaptation attracts more and more research attention, where only a trained source model is assumed to be available, while the labeled source data remains private. To get promising adaptation results, we need to find effective ways to transfer knowledge learned in source domain and leverage useful domain specific information from target domain at the same time. This paper describes our winning contribution to SemEval 2021 Task 10: *Source-Free Domain Adaptation for Semantic Processing*. Our key idea is to leverage the model trained on source domain data to generate pseudo labels for target domain samples. Besides, we propose Negation-aware Pre-training (NAP) to incorporate negation knowledge into model. Our method wins the 1st place with F1-score of 0.822 on the official blind test set of *Negation Detection Track*.

1 Introduction

The *Negation Detection Track* of SemEval 2021 Task 10: *Source-Free Domain Adaptation for Semantic Processing* provides a new setting for unsupervised domain adaptation task which ask participants to conduct negation detection in target domain only with model trained on source domain (namely source model) and unlabeled target domain data. Negation detection is a span-in-context classification problem, where the model will jointly consider both the target mention to be classified and its surrounding context. For example, in sentence *Has no <e> diarrhea <e\> and no new lumps or masses*, the target span *diarrhea* is negated by its surrounding context *no*. This task is important for physicians to extract key information from clinical text. The test dataset used is based on MIMIC-III version 1.4 (Johnson et al., 2016), which is a large, freely-available english database comprising de-identified health-related data.

We approach this task as a problem of learning with pseudo labels. Our main interests include 1) negation knowledge infusion through pre-training on target domain and 2) high-quality pseudo label generation. We divide the task into two stages: Negation-aware Pre-training (NAP) stage and Pseudo label Training stage. In the NAP stage, token-level and sentence-level negation semantic are embedded into model. In the pseudo label training stage, confidence threshold search and mean self-entropy are used to select target domain samples with highly confident pseudo labels.

2 Related Work

Traditional negation detection method are mostly rule-based. These methods (Chapman et al., 2001; Sanchez-Graillet and Poesio, 2007; Huang and Lowe, 2007; Sohn et al., 2012) used regular expression algorithm, dependency parsing and grammatical parsing to perform negation cue detection and scope resolution. Recent years, deep learning has been applied to negation detection task. In (Qian et al., 2016), Convolutional Neural Network was used to recognize negation scope in the sentence. (Lazib et al., 2019) and (Gautam et al., 2018) leveraged recurrent neural network variants to perform negation scope resolution and achieved better performance with BiLSTM, which further indicates the potential in deep learning-based methods. Joint model to detect negation cues and targets simultaneously had been studied by Bhatia et al. (2019). More recently, popular transformer-based model (Khandelwal and Sawant, 2019) had also been used to perform negation detection.

Due to data privacy and data transmission problem, several source-free unsupervised domain adaptation methods (Liang et al., 2020; Kim et al., 2020; Yang et al., 2020) have been proposed for image classification task. These methods mostly focus

Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021), pages 1283–1288
Bangkok, Thailand (online), August 5–6, 2021. ©2021 Association for Computational Linguistics

on generating high-quality pseudo labels for target domain samples before or during training phase and do not involve self-supervised pre-training.

In the natural language processing filed, pre-training is popular. We train language models on huge corpora and fine-tune the pre-trained architectures (Devlin et al., 2018; Liu et al., 2019; Zhang et al., 2019; Yang et al., 2019) in downstream tasks, achieving state-of-the-art results on most NLP tasks. Prior studies (Radford et al., 2018; Chronopoulou et al., 2019; Gururangan et al., 2020; Lee et al., 2020) has further shown the potential of domain-adaptive and task-adaptive pre-training.

3 Method

We approach the task of source-free domain adaptation for negation detection as a problem of learning with pseudo labels. To generate high-quality pseudo labels, we use mean self-entropy as metric to search appropriate probability threshold, which is inspired by (Li et al., 2020). Besides, in order to learn more negation semantic knowledge from target domain, we propose negation-aware pre-training to incorporate negation knowledge by self-supervised training.

3.1 Negation-aware Pre-training

In prior studies, negation cues are important for rule-based and machine learning-based methods. We propose Negation-Aware Pre-training NAP to embed the knowledge of negation cues into representation. As shown in Figure 1, common token masking, negation cue prediction and pseudo negation detection are included in NAP. Common token masking is conducted to capture target domain language knowledge. Negation cue prediction could help the model recognize token-level negation information based on collected negation cue lexicon. Pseudo negation detection is a sequence classification task and designed for complex sentence-level negation knowledge. It is the same as our final negation detection task, however, the target mention and corresponding label is generated by simple heuristic rules. Although these data is somewhat noisy, with the help of pseudo negation detection pre-training, more negation information could be embedded into model.

Negation Cue Lexicon Negation cues are key to the Negation-Aware Pre-training. Based on the lexicon created by (Weng et al., 2020), we divide negation cues into 2 categories: Pre-negation (PREN), Post-negation (POSN). Pre-negation and Post-negation mean the negation cues locate before or after the target mentions respectively. Table 1 shows several examples of each type of negation cues.

PREN	POSN
not	unlikely
none	be ruled out
nor	be excluded
without	be resolved
deny	be absent
no evidence of	be negative

Table 1: Examples of negation cues.

Pseudo Pre-training Data Generation To perform pseudo negation detection task, we need a large number of labeled data. We design simple yet effective rules to generate training data. To simplify the process, we assume all clinical event mentions are single noun. For sentence *without evidence of diarrhea, vomit*, we first locate the negation cue *without*. Since *without* is a pre-negation cue, we take the 3 tokens behind it (*evidence of diarrhea*) as target context. In the target context, the furthest noun *diarrhea* from *without* is selected as target mention. Finally, we generate a pseudo training data: *without evidence of <e> diarrhea <e\>, vomit* with negated label. For samples with post-negation cue, the process is similar, but the target context is before the negation cue. Generally, the selected target mentions are negated by surrounding cues, but there is a special case: double negative. For instance, in sentence *The report can not rule out diarrhea*, *diarrhea* is not negated, since *rule out* is negated by *not*. We assign non-negated label for these double-negative cases. We also generate more non-negated samples by randomly select sentences including no negation cues.

Objectives of Pre-training As illustrated in Figure 1, there are 3 tasks included in the pre-training stage. Common token masking is inherited from RoBERTa (Liu et al., 2019). It is used to capture low level language information in test domain. Specifically, during pre-training, except for the tokens in negation cue lexicon, about 15% of input tokens are random sampled and masked. The model is trained to recover original tokens in the corrupted input sequence. We denote the objective of common token masking task as L_{mlm}. Negation cue prediction as a token classification task is import

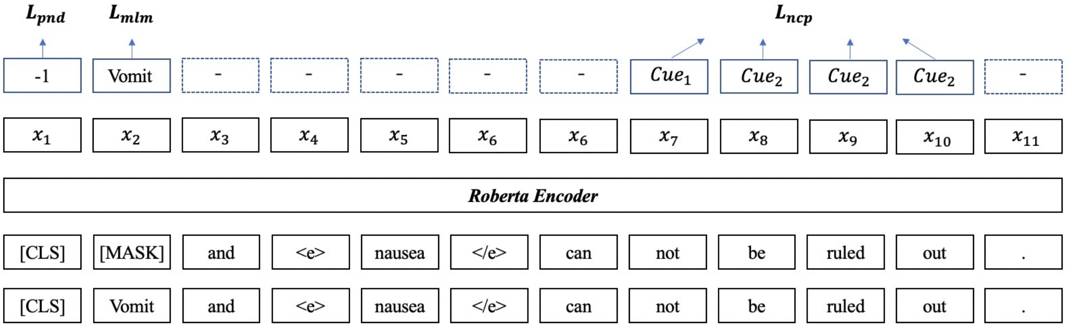

Figure 1: The framework of NAP, Negation-Aware Pre-training. The NAP includes 3 pre-training tasks: common token masking, negation cue detection and pseudo negation detection. Common token masking is inherited from prior pre-training work. Negation cue detection is a token classification task, which aims to embed negation knowledge at token level. Pseudo negation detection is similar to final negation detection task, but the training data is generated based on heuristic rule. Pseudo negation detection could help model capture the complex relation between negation cue and target mention.

for negation knowledge infusion. For each token in input sequence, the task aims to guide the model predict its negation polarity (negation or not) based on its representation x_i output by transformer encoder. Through this way, the pre-trained model could learn the negation knowledge at token level. The objective of negation cue prediction is the same as classical token classification task and is denoted as L_{ncp}. Pseudo negation detection is another crucial key to negation-aware pre-training. Compared to negation cue prediction task, this task could further help model to capture the semantic relation between negation cues and target mentions.The objective L_{pnd} for a single sample is defined as follows:

$$\hat{y} = sigmoid(Wx_1 + b)$$
$$L_{pnd} = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$$

Here x_1 denotes the output vector of first token from transformer encoder. All the above 3 pre-training objectives are jointly optimized. Thus, overall pre-training objective L is:

$$L = L_{mlm} + L_{ncp} + L_{pnd}$$

3.2 Pseudo label Training

Although the negation knowledge is infused through pre-training tasks, the pre-trained model is still lack of the ability to understand complicated semantic information and negation relation. Since the training data and corresponding labels in pre-training stage are generated by simple heuristic rules, most training sample are easy to learn and some samples with wrong labels may harm the model. Thus, test domain samples with high-quality labels are needed to guide the pre-trained model learn more useful information. We leverage the source model to predict the probability of each test sample to be negated or non-negated. Then inspired by *Self-entropy Descent* (SED) proposed in (Li et al., 2020), we conduct *Confidence Threshold Search* to generate high-quality test domain samples.

Confidence Threshold Search Self-entropy could be used as a metric to measure the prediction uncertainty (Kim et al., 2020), i.e. $H(x) = -\sum p(x)log(p(x))$. The lower the self-entropy the more confident the prediction is. We set the probability threshold to be non-negated as 0.999 empirically, then search the probability to be negated from 0.985 to 0.975. The step size is set to be 0.001. The generated pseudo labels are used to fine-tune the source model and then evaluate the mean self-entropy of the dataset after training. When the mean self-entropy descends and hits the first local minimum, we take the corresponding probability as an appropriate threshold for generating labels. Samples do not reach the threshold will be excluded.

4 Source Model and Data

Source Domain Data The source domain data is from SHARP Seed dataset which consists of de-identified clinical notes from Mayo Clinic. In the SHARP data, clinical events are marked with a boolean polarity indicator, with values of either asserted or negated. There are 10259 samples provided including 902 negated instances for source model training.

Source Model Since the data privacy limitation, SHARP Seed dataset cannot be distributed. Thus organizers provide a ”span-in-context” negation detection model trained on SHARP Seed dataset as the source model. The source model is RoBERTa-based and could achieve promising result on the SHARP Seed dataset.

Target Domain Data The target domain data is from MIMIC-III version 1.4 dataset (Johnson et al., 2016). MIMIC-III version 1.4 is a large, freely-available database comprising de-identified health-related data associated with over forty thousand patients who stayed in critical care units of the Beth Israel Deaconess Medical Center between 2001 and 2012. Based on rules, we extract about $50,000$ pseudo samples from the file *NOTEEVENTS.csv* to perform negation-aware pre-training. The offical test dataset including 9580 samples is also extracted from the file *NOTEEVENTS.csv*. To further validate the effectiveness of the proposed method, we further create a custom test dataset which includes 500 negated samples and 500 non-negated samples. In the custom test dataset, various negation cues and negation style are included.

5 Experiment

5.1 Negation-aware Pre-training Stage

We leverage the source model provided by organizers as our initial model in pre-training stage. Besides, several layers are added to perform masked language modeling and negation cue prediction. During pre-training, the learning rate is set as 0.00001, and batch size is set to be 64. We conduct pre-training in 3 epochs.

5.2 Pseudo Label Training Stage

We assume that a sample including no negation cue is definitely non-negated. Thus, we assign these sample without negation cues non-negated label. These samples will not be included in the training phase. Since the provided test samples are extracted directly from *NOTEEVENTS.csv* file, the format of each sample is messy, we conduct sentence split with NLTK toolkit (Loper and Bird, 2002) for each sample and only keep the sentence with target mention.

Then confidence threshold search is conducted to generate high-quality labels for the rest of test data. Finally, the confidence threshold for negated and non-negated samples are selected as 0.983 and 0.999 respectively. In other words, if a test sample assigned negated label, the probability to be negated generated by the source model should be higher than 0.983. Similarly, if a test sample assigned non-negated label, the probability to be non-negated should be higher than 0.999.

During training stage, we use the source model provided by organizers but initialize the transformer encoder with the corresponding part from pre-training model, because the transformer encoder from the pre-training model could capture various negation knowledge from input sentences. The learning rate is kept as 0.00001, and batch size is 32. The number of epoch is set to be 5. In the first 2 epoch, the parameters of *ClassificationHead* in model is frozen.

6 Results

The result is evaluated using the standard precision, recall and F1 scores as used in most published work. We achieve the best performance on the official blind test dataset.

6.1 Result on Official Test Dataset

We compare our method with two baselines, and the result is shown in Table 2. The result of source model without any domain adaptation is inferior. Masked language modeling trained on target domain data could improve the performance. However, its effect is not significant, because masked language pre-training focus more on low-level language information rather than high-level semantic knowledge about negation. With the help of NAP, the adapted model could improve the recall score with a large margin. This indicates that our negation-aware pre-training method could help embed negation knowledge into the sequence representation, and facilitate the domain adaptation from source domain to target domain.

Although the proposed adaptation method achieves superior result in the competition, there still exists a problem which harms the recall performance: the adapted model is not sensitive to long-term negation dependency. For example, in the case *He denies chest pain, dyspnea, dizziness/lightheadedness, or <e> abdominal pain <e\>*, though the mention *abdominal pain* is connected with negation cue *denies* via *or*, the model still output non-negated prediction. This phenomenon may be caused by pre-training, since in pseudo negation detection pre-training task, the training data are generated only based on simple

Methods	F1	Precision	Recall
Source	0.685	0.921	0.545
Source + MLM	0.724	0.905	0.603
Source + NAP	**0.822**	**0.902**	**0.756**

Table 2: Results of different methods on official test set. MLM, NAP denote masked language modeling (common token masking), negation-aware pre-training respectively. Our experiments are conducted on the data cleaned, so the result of source model without adaptation is better than the one organizers provide.

Methods	Precision	Recall
Source	0.761	0.502
Source + MLM	0.783	0.526
Source + NAP	0.894	0.833

Table 3: Results of different methods on customized test set.

rules: the clinical mention is a single noun and the distance between mention and negation cue is limited to no more than 3 tokens. To handle this problem, useful and effective generation method should be further explored. In addition, dependency relation between tokens may be introduced into both pre-training and training stage to solve this problem in negation detection task.

6.2 Result on customized Test Dataset

Negated samples in the official test dataset only include *deny*, *none*, *no*, *not* and *without*, so we also conduct experiments on the customized test data we manually created from test domain, which contains various negation cues and two negation styles (active or passive voice). As shown in Table 3, compared to the proposed method, the recall of source model and source model with MLM is much lower, because many negated samples with *never*, *resolve*, *free of*, *absent* and *exclude* can not be recognized correctly. Meanwhile, due to the lack of the ability to capture double-negative semantic, they both fail to distinguish false-positive samples from real positive ones. However, with the negation knowledge embedded through negation-aware pre-training, our method could handle both scenarios better.

7 Conclusion

In this paper, we model source-free domain adaptation as learning with pseudo label. We leverage mean self-entropy of dataset to search appropriate probability threshold for high-quality pseudo label generation. Besides, we propose negation-aware pre-training to integrate different types of negation knowledge to improve the generalization of negation representation. The result shows that the additional negation-aware pre-training is helpful for negation detection task. In the future, we will work towards more robust pseudo label generation method and effective pre-training task introducing more knowledge such as part-of-speech tag and dependency relation.

References

Parminder Bhatia, E Busra Celikkaya, and Mohammed Khalilia. 2019. End-to-end joint entity extraction and negation detection for clinical text. In *International Workshop on Health Intelligence*, pages 139–148. Springer.

Wendy W Chapman, Will Bridewell, Paul Hanbury, Gregory F Cooper, and Bruce G Buchanan. 2001. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics*, 34(5):301–310.

Alexandra Chronopoulou, Christos Baziotis, and Alexandros Potamianos. 2019. An embarrassingly simple approach for transfer learning from pretrained language models. *arXiv preprint arXiv:1902.10547*.

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*.

Dipesh Gautam, Nabin Maharjan, Rajendra Banjade, Lasang Jimba Tamang, and Vasile Rus. 2018. Long short term memory based models for negation handling in tutorial dialogues. In *FLAIRS Conference*, pages 14–19.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. 2020. Don’t stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

Yang Huang and Henry J Lowe. 2007. A novel hybrid approach to automated negation detection in clinical radiology reports. *Journal of the American medical informatics association*, 14(3):304–311.

Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-Wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. 2016. Mimic-iii, a freely accessible critical care database. *Scientific data*, 3(1):1–9.

Aditya Khandelwal and Suraj Sawant. 2019. Negbert: A transfer learning approach for negation detection and scope resolution. *arXiv preprint arXiv:1911.04211*.

Youngeun Kim, Sungeun Hong, Donghyeon Cho, Hyoungseob Park, and Priyadarshini Panda. 2020. Domain adaptation without source data. *arXiv preprint arXiv:2007.01524*.

Lydia Lazib, Yanyan Zhao, Bing Qin, and Ting Liu. 2019. Negation scope detection with recurrent neural networks models in review texts. *International Journal of High Performance Computing and Networking*, 13(2):211–221.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240.

Xianfeng Li, Weijie Chen, Di Xie, Shicai Yang, Peng Yuan, Shiliang Pu, and Yueting Zhuang. 2020. A free lunch for unsupervised domain adaptive object detection without source data. *arXiv preprint arXiv:2012.05400*.

Jian Liang, Dapeng Hu, and Jiashi Feng. 2020. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 6028–6039. PMLR.

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*.

Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

Zhong Qian, Peifeng Li, Qiaoming Zhu, Guodong Zhou, Zhunchen Luo, and Wei Luo. 2016. Speculation and negation scope detection via convolutional neural networks. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 815–825.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training.

Olivia Sanchez-Graillet and Massimo Poesio. 2007. Negation of protein–protein interactions: analysis and extraction. *Bioinformatics*, 23(13):i424–i432.

Sunghwan Sohn, Stephen Wu, and Christopher G Chute. 2012. Dependency parser-based negation detection in clinical narratives. *AMIA Summits on Translational Science Proceedings*, 2012:1.

Wei-Hung Weng, Yu-An Chung, and Schrasing Tong. 2020. Clinical text summarization with syntax-based negation and semantic concept identification. *arXiv preprint arXiv:2003.00353*.

Shiqi Yang, Yaxing Wang, Joost van de Weijer, and Luis Herranz. 2020. Unsupervised domain adaptation without source data by casting a bait. *arXiv preprint arXiv:2010.12427*.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*.

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*.