SiaKey: A Few-shot Learning Method for Health-Related Natural Language Processing

Anonymous EACL submission

Abstract

Clinical Natural Language Processing (NLP) has become an important sub-field of Natural Language Processing, which can identify patients' diseases and support doctors' decisions. With the development of Deep Learning, more and more NLP tasks are solved by state-of-art neural networks. However, these supervised models usually need a huge amount of annotated data to train and are computationally expensive. The annotation process is difficult and time-consuming for clinical data that exists in large-scale electronic health records (EHRs) and online posts, where only specialists with professional clinical knowledge could annotate them manually. On the other hand, the widely-used method that fine-tuning Pre-trained Language Models (PLMs) always has poor performance on few-shot training data. Thus, we proposed a novel Few-Shot Learning (FSL) approach called SiaKey System that combines Siamese Networks, Keyphrases Extraction, and Domain Knowledge. We tested few-shot settings including 5-shot, 10-shot, 15-shot, and 20-shot learning on health-related online posts classification tasks. The experiment results improved by 10 percent with the title and keyphrases application and reached over 60% accuracy, which proves the effectiveness of our system to capture features of texts and better performance compared to BioBERT.

1 Introduction

The abundant annotated data is considered the necessary element to train deep neural networks for clinical Natural Language Processing (NLP). However, the annotation process could be very expensive due to the huge amount of related data and the high demand for professional clinical knowledge. Even though there are several good quality and publicly available clinical datasets, such as i2b2 datasets (Uzuner et al., 2011), MIMIC-III datasets (Johnson et al., 2016), and BioNLP datasets (Nédellec et al., 2013), more annotated datasets from online patients' posts in social media are needed as a supplementary to electronic health records (EHRs) to build the complete clinical NLP systems. Considering the fact that state-of-the-art supervised deep learning neural networks always have poor performance when the training data is in shortage, implementing a clinical NLP system with few or no annotated data becomes the focus in clinical informatics research.

The fine-tuning approach is applied to solve the traditional neural networks' disadvantage on NLP tasks. Pre-trained Language Models (PLMs) (Qiu et al., 2020) are deep neural network models trained on unlabelled large-scale datasets, such as Wikipedia or PubMed data. This training process is called pre-training, which always takes a long time and vast amount of computational resources. Once this pre-training process is completed, these sophisticated language models always have general ability for NLP tasks. However, many NLP tasks are domain-specific and the PLMs need more domain texts to implement specified applications. The process that retraining the PLMs on a task-specific labeled dataset is called fine-tuning. The entire paradigm is called transfer learning, as the model learns a general context during pre-training, then this knowledge is transferred for a specific downstream NLP task by fine-tuning it in a low-data collection. For example, the state-of-the-art medical and clinical PLMs include BioBERT (Lee et al., 2020) and Clinical BERT (Alsentzer et al., 2019), which have been trained on millions of EHRs and unlabeled clinical text datasets like MIMIC. This pre-training process enables them to learn general medical and clinical linguistic characteristics. Later, this general knowledge could be transferred for specific downstream tasks like biology text mining or clinical Named Entity Recognition (NER) (Sun and Yang, 2019) tasks, by finetuning the PLMs with a lesser amount of task-specific annotated data (Sivarajkumar and Wang, 2022).

The training approach with a small amount of annotated data collection is called few-shot learning (FSL). Unfortunately, the fine-tuning PLMs procedure also requires a considerable amount of data and has unsatisfied performance on FSL. Siamese Neural Networks (SNN) are similar neural networks where input vectors are passed through it to extract features, which are later passed through the triplet loss in the few-shot learning process. These networks are initially used for computer vision tasks but the same idea can be extended to text classification. To the best of our knowledge, applying SNN to clinical NLP is still underrated. It could be trained to compute embeddings, which is the crucial element of NLP that extracts features and transfer the input text into vectorized forms. Then these vectors can be used for different NLP tasks, such as Named-Entity Recognition (NER) and Sentiment Understanding.

Domain knowledge refers to the specific knowledge of a field, and it can provide support and benefit to our domain research. However, domain knowledge is always ignored in clinical NLP studies. The related work usually just focused on collected and analyzed contents of EHRs that only contain symptom descriptions from health specialists. Whereas, online health-related posts on social media should also take into consideration, which possesses patients' emotions, feelings, and other irrelevant words. This task is much more challenging than content in EHRs. Therefore, we propose a keyphrases extraction strategy to extract medical entities from both the title and text, then add these keywords to the title and text as input. The result shows that this innovative method has positive effects on training and achieved better performance than contents.

2 Related Work

Pre-trained Language Models Various language model architectures have been proposed to solve different NLP tasks. The popular first-generation PLMs incorporate GloVe (Pennington et al., 2014) and Word2Vec (Goldberg and Levy, 2014). The features of first-generation PLMs are word embedding based on occurrences of different words in documents. Although these models are efficient in comprehending semantic information, they usually ignored the linguistic meaning of words and the underlying contexts behind the embeddings.

The second-generation PLMs are BERT (Devlin et al., 2018), GPT-2 (Radford et al., 2019), and T5 (Raffel et al., 2020). The main improvement of these PLMs is that they take the context information into consideration, which could reveal and understand complicated concepts of words. Consequently, these models have achieved state-of-the-art performance for diverse NLP tasks.

Siamese Networks There have been studies assessing the effectiveness of Siamese Networks (SNN) for image classification. Zhang et al. (2021) used SNN to capture the spatial information for object tracking tasks via multiscale spatial attentions. And Hunt et al. (2021) applied SNN for the classification of electrograms.

In the context of FSL, SNN has been used by Koch et al. (2015) for one-shot image recognition, which was based on the convolutional architecture to retrieve discriminative features from only one single example of each new class. Droghini et al. (2018) employed SNN for few-shot human-fall detection objectives by using audio signals. Nevertheless, none of these studies used SNN-based FSL for NLP. There is one recent study by Oniani et al. (2022) that explored SNN for FSL in Clinical NLP and demonstrated good performance on text classification and Named-Entity Recognition (NER). However, their dataset only contains clinical narratives from professional experts, which is much more straightforward and doesn't need domain knowledge such as keyphrase extraction to abstract features. To the best of our knowledge, none of the studies referenced above are using SNN and domain-specific information in machine learning.

3 Methods

In this section, we first present how we build the Siamese Networks on the FSL of NLP. Then, we propose a keyphrase extraction method for additional text features to improve the few-shot performance.

3.1 Sentence Embeddings

Considering the speed and efficiency demand of FSL, we used the universal-sentence-encoder (Cer et al., 2018), which is a sentence encoding module of tensorflow-hub. It encodes sentences into high-dimensional embeddings which can be further employed for semantic similarity, text classification, and other NLP tasks. The embeddings vector is 512 in length, irrespective of the length of the input, which is the biggest advantage compared to BERT encoder which has length restrictions. We will use this pre-trained universal sentence encoder to encode our sentences to get a better representation of sentences leveraging transfer learning.

3.2 Siamese Networks Architecture

The primary concept of Siamese Networks is to compute the triple loss between the anchor ($\boldsymbol{A}$), positive ($\boldsymbol{P}$), and negative ($\boldsymbol{N}$) input text as the loss function. After getting the sentence embeddings from the universal sentence encoder, we will preprocess these embeddings by several dense and normalization layers before feeding the final triple loss layer. This preprocess can help decrease the variance and dimension of input embeddings. Figure 1 explains our whole system architecture. Then the triplet loss takes three input embeddings of an anchor ($\boldsymbol{A_i}$), positive ($\boldsymbol{P_i}$) and negative ($\boldsymbol{N_i}$) data samples. The anchor and positive embeddings are of the same class and negative embedding is of a different class. We try to project the embeddings such that the distance between the anchor and negative samples $\boldsymbol{d(A_i, N_i)}$ is alpha ($\boldsymbol{\alpha}$) more than the distance between the anchor and positive samples $\boldsymbol{d(A_i, P_i)}$. $\boldsymbol{\alpha}$ is also defined as the margin point, because we are used to defining non-negative loss functions in neural networks. The loss would be zero if the difference between $\boldsymbol{d(A_i, N_i)}$ and $\boldsymbol{d(A_i, P_i)}$ is greater than the margin. Otherwise, this difference in distance is considered as the triplet loss and back-propagated through the whole Siamese Network. The mathematical definition of the triple loss is shown below.

$$\boldsymbol{L(A,P,N)=\frac{1}{N}(\sum_{i=1}^{N} max(d(A_i,P_i)-d(A_i,N_i)+\alpha,0)} \tag{1}$$

where $\boldsymbol{d(X_i, Y_i)} = ||\vec{\boldsymbol{x_i}} - \vec{\boldsymbol{y_i}}||_2^2$

3.3 Medical Entity Extraction

In our health-related tasks, we consider keyphrases as critical medical entities in text. For medical entity extraction, we utilized Scispacy (Neumann et al., 2019), which extracts the entities in the text and links them with UMLS [1], medical metathesaurus. However, UMLS contains both medical as well as non-medical terms for example emotions, body anatomy, and activities. As we are interested in only health or medical-related terms mentioned in the post by the user we filtered these entities based on semantic types their semantic types. The entities related to drug, disease, treatment procedures, and diagnosis. Scispacy provides all the entities in the text, but for our purpose, we only wanted top N entities that are more representative of the post. So we ranked the entities based on EmbedRank (Bennani-Smires et al., 2018) and utilized the top N entities which have more similarity with the user-generated text. For this purpose, we utilized BioBERT embeddings (Lee et al., 2020). As a service tool, this API can be accessed at: http://ngrok.luozm.me:8395/keyphrase/kweigt

3.4 Domain Knowledge

When we analyzed deeper into the posts of each class, it was found that titles always enclose much valuable information and serve as a hint or summarization of the following post content. Therefore, the function of title is similar to manual keyphrases extraction, which is from the patient themselves' understanding and could assist our model to retrieve critical features from the body of their post, especially when the post text is long and filled with unrelated words. Finally, we decided to integrate the manual and automatic keyphrases extraction methodology in our experiments.

4 Datasets

For our experiments, we consider the Drug Abuse Data - Reddit Dataset (Ghosh et al., 2020), which was published in the paper "Utilizing Social Media for Identifying Drug Addiction and Recovery Intervention" in 2020. The dataset contains posts from drug addiction-related Subreddits which are topic-specific communities within the Reddit online social media (https://www.reddit.com/). It annotated 3151 posts out of all the posts collected as one of the 5 classes (Stage of addiction): 'Addicted', 'E(early)-Recovery', 'M(maintaining)-Recovery', 'A(advanced)-Recovery', 'Others'. In our NLP classification task, we used the 'title' (Each post is associated with a title), 'body'(The main descriptive part of the post), and 'label_classification'(The label given to a post for the classification task) three fields. To test our few-shot learning methods, we randomly picked 200 samples for each of the five classes to build our test datasets with

[1] https://uts.nlm.nih.gov/uts/umls/home

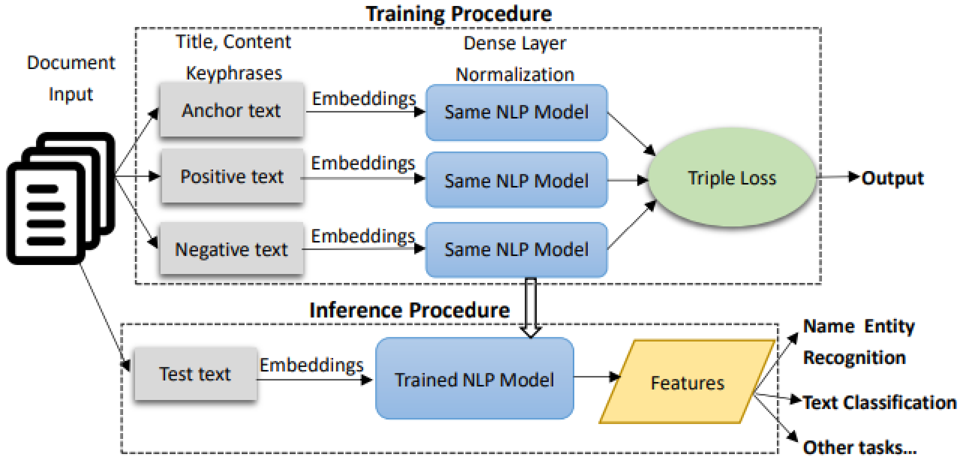

Figure 1: Overview of our SiaKey system. The NLP Model is a simple neural network with several dense layers. (see Appendix A for a full list of hyper-parameters)

Title	One month off suboxone.
Body	Hello everyone. Here we are. One month free and clear. Feels good man!
Label_classification	E-Recovery

Table 1: Datasets sample.

1000 samples in total. In addition, we randomly selected 20 samples for each of the five classes out of the test datasets to build our FSL training datasets. Specifically, we conducted 5-shot(S_5), 10-shot(S_{10}), 15-shot(S_{15}) and 20-shot(S_{20}) training process. The relationship between these training sets is: $S_5 \subset S_{10} \subset S_{15} \subset S_{20}$. Therefore, the larger shots training set always contains fewer shots training sets, which guarantees the fairness of comparing different shots training results. One training sample is shown in Table 1.

5 Experiments and Results

In order to evaluate our SiaKey system, we trained and tested text classification tasks on the Drug Abuse Dataset. Specifically, we first used BioBERT to fine-tune all samples except the test training set as the upper bound of our experiments. Then, we assess the feasibility of combining the title with text as the input and compared the results between them. Finally, we trained our model with various combinations of title, text, and keyphrases(extracted from the title, text, or both) to analyze their performance.

All the machine learning processes on this classification task after gaining text features from our Siakey system are the simple K-nearest neighbors(KNN) algorithm with K=5. The FSL was done on 1 GPU (NVIDIA gtx1080), with a batch size of 32. We trained the model for 50 epochs with 10 steps for each epoch. The optimizer is ”Adam” with a learning rate of $1e{-3}$. Although we used the metrics Precision, Recall, and F1 score of all five classes to evaluate the text classification task, we decided to show the final performance via Accuracy and Average-F1 metrics due to the number of results.

5.1 Baselines Experiments

BioBERT was employed as the baseline for fine-tuning on the PLMs method, which was trained on all 2150 available samples except the test cases in our datasets. The result in Table 2 shows the good performance of fine-tuning on sufficient data volumes and functions as an upper bound for our FSL approach.

Model	Accuracy	Average-F1
BioBERT	80.83%	80.74%

Table 2: BioBERT fine-tuning.

5.2 Few-shot Learning Experiments with Title and Text

We trained our SiaKey system on 5-shot, 10-shot, 15-shot, and 20-shot FSL processes, with different inputs such as title, text, and title+text. Meanwhile, we compared them with BioBERT which is poorer even than the worst result in each evaluation. Thus, we only showed the result of BioBERT on title+text input in Table 3.

Model-Shots	Input	Accuracy	Average-F1
SiaKey-5	Title	46.80%	46.31%
SiaKey-5	Text	45.50%	45.32%
SiaKey-5	Title+Text	**49.10%**	**48.63%**
BioBERT-5	Title+Text	21.33%	13.71%
SiaKey-10	Title	52.50%	52.01%
SiaKey-10	Text	50.40%	50.51%
SiaKey-10	Title+Text	**53.20%**	**53.23%**
BioBERT-10	Title+Text	45.67%	46.74%
SiaKey-15	Title	59.30%	59.18%
SiaKey-15	Text	51.40%	51.70%
SiaKey-15	Title+Text	**59.80%**	**59.72%**
BioBERT-15	Title+Text	49.17%	48.55%
SiaKey-20	Title	**64.50%**	**64.60%**
SiaKey-20	Text	53.10%	52.99%
SiaKey-20	Title+Text	59.60%	59.52%
BioBERT-20	Title+Text	57.83%	58.30%

Table 3: FSL experiments with Title and Text.

5.3 SiaKey System Experiments with Title, Text, and Keyphrases

This section studies our SiaKey system performance with various input text combinations of title, text, and keyphrases from 5-shot to 20-shot training sets. More specifically, keyphrases are extracted from the title, text, or both of them. For better comparison visualization, we would only attach the training input of ”Title” and ”Title+Text”in Table 3 into Table 4. In addition, the p-value of the t-paired tests between Title and Title+Text+Keyphrases(Title, Text, Both) of the same shot are all less than 0.05, which means we have enough proof to claim that their performances are different.

6 Discussion

6.1 Performance Comparisons Between BioBERT and FSL Strategy

The results in Table 3 reveal that the performance of BioBERT on this FSL text classification task is far behind our SiaKey system. For the fine-tuning process with samples of less than 20-shots, BioBERT has difficulties capturing features from text effectively. This is according to our expectations because the fine-tuning procedure always takes more than 20 samples for each class. However, it is evident that with the increasing number of shots, the difference between BioBERT and our SiaKey system is decreasing. And BioBERT would surpass our system when there are enough input samples, which can be shown as the upper bound in Table 2.

6.2 The experimental results of Title and Keyphrases

When we only connect the title with text as the input, it is obvious that the variety Title+Text has better performance than the title or text itself in the 5, 10, and 15 shots. Furthermore, the results of Title are even better than the results of Text. Therefore, the manual summarization advantage of the title is beneficial to our task. On the other hand, when keyphrases from the title and text are appended to the original input, the performance increase again and the best result of less than 20-shots is the Title+Text+Keyphrases(Both) combination. Thus, it is confident to assert that the automatic summarization advantage of keyphrases from both the title and text is apparent. The last interesting phenomenon is that the Title itself has the best outcome in 20-shots evaluation, even better than all other mixtures of text and keyphrases. One possible supposition for this is that the effectiveness of keyphrases is declining when input samples increase, even too many keyphrases could become a burden for tasks. Consequently, Title itself serves as the most concise extraction and evolves the most promising.

6.3 The Advantage and Interpretability of Siakey System

For interpreting the reason why titles and keyphrases could improve our model, we investigated the right and wrong predictions of our model and select one as an example to discuss here. The full information on this sample could be found in Appendix B and we will show the effectiveness of the title and keyphrases in this subsection. If only the content was taken for input according to common approaches, it would be difficult to extract information in terms of this 484-length text containing large parts of the emotions and feelings of the patient himself. The title ”22 days today. Went from begging for subs to deciding to cold turkey completely” is critical for summarizing what this patient wants to express in his following post content. In addition, the keyphrases ”cold turkey” from the title further indicates that he is in the ”A(advanced)-Recovery” class and wanted to abruptly complete cessation of the drug. Moreover, the keyphrases ”pregabalin” and ”restless legs” from the content also suggest the advanced

Shots	Input	Accuracy	Average-F1
5	Title	46.80%	46.31%
5	Title+Keyphrases(Text)	47.90%	46.44%
5	Text+Keyphrases(Text)	45.30%	44.75%
5	Title+Text	49.10%	48.63%
5	Title+Text+Keyphrases(Title)	50.50%	50.80%
5	Title+Text+Keyphrases(Text)	50.00%	49.64%
5	Title+Text+Keyphrases(Both)	**51.70%**	**51.54%**
10	Title	52.50%	52.01%
10	Title+Keyphrases(Text)	51.90%	50.59%
10	Text+Keyphrases(Text)	50.20%	49.99%
10	Title+Text	53.20%	53.23%
10	Title+Text+Keyphrases(Title)	50.90%	53.59%
10	Title+Text+Keyphrases(Text)	54.00%	52.48%
10	Title+Text+Keyphrases(Both)	**54.50%**	**54.20%**
15	Title	59.30%	59.18%
15	Title+Keyphrases(Text)	55.60%	55.44%
15	Text+Keyphrases(Text)	51.30%	51.32%
15	Title+Text	59.80%	59.72%
15	Title+Text+Keyphrases(Title)	59.90%	59.03%
15	Title+Text+Keyphrases(Text)	59.30%	58.50%
15	Title+Text+Keyphrases(Both)	**60.30%**	**59.75%**
20	Title	**64.50%**	**64.60%**
20	Title+Keyphrases(Text)	53.10%	53.19%
20	Text+Keyphrases(Text)	53.10%	53.19%
20	Title+Text	59.60%	59.52%
20	Title+Text+Keyphrases(Title)	61.00%	60.97%
20	Title+Text+Keyphrases(Text)	60.50%	60.48%
20	Title+Text+Keyphrases(Both)	60.50%	60.11%

Table 4: SiaKey system experiments with Title, Text, and Keyphrases.

recovery stage of this patient. In conclusion, our Siakey system captured the key features via title and keyphrases and improved the performance of the whole system.

7 Conclusion

In this paper, we study the problem of few-shot learning aspect-based text classification task. We established the Siakey system and designed various experiments to investigate our model performance. Results indicate that our Siakey system performs better than the same tasks on BioBERT, the introduction of titles and keyphrases also improves the model by the way of content summarization. However, there are also limitations of our work that could be discussed in future work. First, only the BioBERT model is compared in our experiments, the future work could introduce comparisons between more PLMs. Second, the reason why the title itself can have better performance than other combinations after 20-shot FSL is not fully explained, other related work could analyze this problem. Third, the lack of public datasets specialized for FSL (Ge et al., 2022) makes it difficult to compare our work to other FSL strategies on benchmark datasets. Thus, system comparisons and benchmarking are one of the directions for future studies of FSL.

References

Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. 2019. Publicly available clinical bert embeddings. *arXiv preprint arXiv:1904.03323*.

Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl, and Martin Jaggi. 2018. Simple unsupervised keyphrase extraction using sentence embeddings. In *Proceedings of the 22nd Con-*

ference on Computational Natural Language Learning, pages 221–229.

Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.

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*.

Diego Droghini, Fabio Vesperini, Emanuele Principi, Stefano Squartini, and Francesco Piazza. 2018. Few-shot siamese neural networks employing audio features for human-fall detection. In *Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence*, pages 63–69.

Yao Ge, Yuting Guo, Yuan-Chi Yang, Mohammed Ali Al-Garadi, and Abeed Sarker. 2022. Few-shot learning for medical text: A systematic review. *arXiv preprint arXiv:2204.14081*.

Shalmoli Ghosh, Janardan Misra, Saptarshi Ghosh, and Sanjay Podder. 2020. Utilizing social media for identifying drug addiction and recovery intervention. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 3413–3422. IEEE.

Yoav Goldberg and Omer Levy. 2014. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.

Bram Hunt, Eugene Kwan, Derek Dosdall, Rob S MacLeod, and Ravi Ranjan. 2021. Siamese neural networks for small dataset classification of electrograms. In *2021 Computing in Cardiology (CinC)*, volume 48, pages 1–4. IEEE.

Alistair EW Johnson, Tom J Pollard, Lu Shen, Li-wei H Lehman, 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.

Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. 2015. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, page 0. Lille.

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.

Claire Nédellec, Robert Bossy, Jin-Dong Kim, Jung-Jae Kim, Tomoko Ohta, Sampo Pyysalo, and Pierre Zweigenbaum. 2013. Overview of bionlp shared task 2013. In *Proceedings of the BioNLP shared task 2013 workshop*, pages 1–7.

Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. 2019. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 319–327, Florence, Italy. Association for Computational Linguistics.

David Oniani, Sonish Sivarajkumar, and Yanshan Wang. 2022. Few-shot learning for clinical natural language processing using siamese neural networks. *arXiv preprint arXiv:2208.14923*.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. 2020. Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10):1872–1897.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

Sonish Sivarajkumar and Yanshan Wang. 2022. Healthprompt: A zero-shot learning paradigm for clinical natural language processing. *arXiv preprint arXiv:2203.05061*.

Cong Sun and Zhihao Yang. 2019. Transfer learning in biomedical named entity recognition: an evaluation of bert in the pharmaconer task. In *Proceedings of The 5th Workshop on BioNLP Open Shared Tasks*, pages 100–104.

Özlem Uzuner, Brett R South, Shuying Shen, and Scott L DuVall. 2011. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association*, 18(5):552–556.

Jianming Zhang, Benben Huang, Zi Ye, Li-Dan Kuang, and Xin Ning. 2021. Siamese anchor-free object tracking with multiscale spatial attentions. *Scientific Reports*, 11(1):1–14.

A NLP Model Architecture

Our neural networks layers sequences are: Input(shape=512), dense_layer1(shape=256), dropout_layer1(shape=256), batch_normalization layer(shape=256), dense_layer2(shape=64), dropout_layer2(shape=64), dense_layer3(shape=128), norm_layer(shape=128).

B Full Sample in Section 6.3

Sample title: ”22 days today. Went from begging for subs to deciding to cold turkey completely”;
Sample label_classification: ”A-Recovery”;
Keyphrases from title: ”cold turkey”;
Keyphrases from content: ”pregabalin”, ”restless legs”;
Sample body:
”Basically, the NHS is useless and in some cases dangerous.
I have taken this path before with benzos, got myself up to 20mg alprazolam a day and not even really feeling it, had to do a taper over two months and manage it all by myself. I think this really helped my opioid recovery.
22 days ago I took the plunge. I was about to go into debt, and like with the benzos, I didn’t want to be a prisoner, unable to leave without smuggling gear with me.
I went through the acute phase all by myself completely cold turkey as I didn’t want to scare my mother. After about five days I came clean to her, she cried, I cried.
We booked appointments, we went to A&E. They offered me 8mg of codeine. They didn’t even check my blood pressure. I attempted to laugh through the sobs and left.
I went to the GP. I asked for something for my restless legs. Due to my previously-admitted benzo addiction they wouldn’t prescribe me anything abusable, and as pregabalin and gabapentin are classed as abusable, they gave me nothing. I left.
I began counselling. I kept a journal to remind me of the pathos of this path. I got in the car and just drove, I went to the gym and just stumbled, but I did not fall. I did not leave.
I almost fell. I went to the ATM two days ago, had money in my pocket. I was so close to giving in. You can see it in my post history. But I left.
Yesterday I realised I am alone in this fight. Despite paying my taxes and national insurance I have been offered either no help or options that would have been a step back. This strengthend my resolve. I know I can do this, I’ve done it with benzos. I will do this myself. I am doing this myself.
Today I woke up and if I closed my eyes and concentrated really hard I could almost convince myself that I’m human. I went to the GP. They said they could recommend prescribing me subs. I refused. I left.
This is possible. It is not easy. But I’m sorry my sweet, terrible seductress, it’s time for me to leave. It’s time for me to leave.
Goodbye is such a definite word, an infinite word. But there are no waterfalls at the edge of the world. I hope it never comes to it but I’m not naf.
Brothers and sisters and all inbetween, you can do it. All things pass, and if you’re not careful your time will come too soon and the rent you leave in the lives of everyone around you will ruin them until they too pass. Your parents. Your siblings.
I love you all. If you want to talk private message me or comment here.
I love you all.”