

Exploring Word Segmentation and Medical Concept Recognition for Chinese Medical Texts

Yang Liu^{♠◇}, Yuanhe Tian[♥], Tsung-Hui Chang^{♠◇}, Song Wu^{♣△}, Xiang Wan[◇], Yan Song^{♠◇†}

[♠]The Chinese University of Hong Kong (Shenzhen), [♥]University of Washington

[♣]PingHu Hospital of Shenzhen University

[△]Shenzhen Hospital of Shanghai University of Traditional Chinese Medicine

[◇]Shenzhen Research Institute of Big Data

[♠]yangliu5@link.cuhk.edu.cn [♥]yhtian@uw.edu [♣]wusong@szu.edu.cn

[◇]wanxiang@sribd.cn [♠]{changtsunghui, songyan}@cuhk.edu.cn

Abstract

Chinese word segmentation (CWS) and medical concept recognition are two fundamental tasks to process Chinese electronic medical records (EMRs) and play important roles in downstream tasks for understanding Chinese EMRs. One challenge to these tasks is the lack of medical domain datasets with high-quality annotations, especially medical-related tags that reveal the characteristics of Chinese EMRs. In this paper, we collected a Chinese EMR corpus, namely, ACEMR, with human annotations for Chinese word segmentation and EMR-related tags. On the ACEMR corpus, we run well-known models (i.e., BiLSTM, BERT, and ZEN) and existing state-of-the-art systems (e.g., WMSeg and TwASP) for CWS and medical concept recognition. Experimental results demonstrate the necessity of building a dedicated medical dataset and show that models that leverage extra resources achieve the best performance for both tasks, which provides certain guidance for future studies on model selection in the medical domain.¹

1 Introduction

Medical language processing (MLP), i.e., natural language processing (NLP) for the electronic medical record (EMR), has drawn significant attention over the past few decades (Rector et al., 1991; Friedman et al., 2004; Stevenson et al., 2012; Kolecck et al., 2021). EMR normally records the entire process of a patient’s examination, diagnosis, and treatment by clinicians in the hospital, and contains a large amount of medical information, which, if extracted properly, can be used to train a machine learning model as an automated tool for auxiliary diagnosis and treatment, forming the foundation of wise information technology of medicine.

[†]Corresponding author.

¹The resources in this paper are released at <https://github.com/cuhksz-nlp/ACEMR>.

Chinese word segmentation (CWS) and medical concept recognition are two important and related tasks for Chinese MLP, which received much attention in previous studies (Xing et al., 2018; Wang et al., 2019). The first task (i.e., CWS) aims to segment Chinese text (i.e., character sequence) into words, which is a necessary step for MLP because the meaning of many medical terms cannot be simply inferred by its component characters. For example, it is hard to infer the meaning of “扁桃体” (*tonsil*) from its components “扁” (*flat*), “桃” (*peach*), and “体” (*body*). The second task (i.e., medical concept recognition) assigns an EMR-related tag (e.g., *Organism* and *Group*) to the segmented words. It is worth noting that the medical concept in this paper includes not only the standard medical named entities but also other categories that are useful for medical text analysis. For example, “Time” is a medical concept that can be used to represent the disease history; “Probability” is a possible medical concept tag for “考虑” (*consider*), in EMR.

To perform CWS and medical concept recognition in Chinese EMR, researchers face a challenge that existing training data for the tasks is either publicly unavailable or of poor quality. Although one possible solution is to apply models trained in the general domain to the medical text, these models always fail to have good performance because there are many domain-specific medical terms that rarely occur in the general domain. To address these challenges, we collect and annotate a new Chinese EMR corpus, named ACEMR, where texts from 500 EMRs (7K sentences) are annotated with CWS and medical concept recognition labels. In addition, we test several state-of-the-art models for CWS and medical concept recognition on the collected ACEMR corpus. Experimental results show the necessity of constructing an informative Chinese medical corpus and provide certain guidance for the model selection in medical domain.

基本信息 (Basic Information)	Patient’s name, gender, age, reason for admission, time of admission
病历特点 (Case Characteristic)	Detailed symptoms of the patient before admission, past history, physical examination results and auxiliary examination results
初步诊断 (Preliminary Diagnosis)	The type of disease initially judged, usually several disease names
鉴别诊断 (Differential Diagnosis)	According to the main complaint of the patient, distinguish it from other diseases and exclude the possible diagnosis of other diseases
治疗计划 (Treatment Plan)	Medical examinations to be done in the next step, and preliminary treatment plan

Table 1: The major five parts of information contained in one First Course Record in Chinese EMRs.

2 Related Work

NLP for medical text has draw many attentions in the recent years (Xue et al., 2012; Xu et al., 2015; Li et al., 2019; Tian et al., 2019, 2020a; Song et al., 2020; Wang et al., 2020; Chen et al., 2020b), especially for the EMR texts. Among different tasks to process Chinese EMR texts, CWS and medical concept recognition are two fundamental ones that draw much attentions from previous studies. Due to the dramatic performance drop when applying the model trained from open source corpus on the medical field, previous studies (Xu et al., 2014, 2015; Li et al., 2015; Zhang et al., 2016; He et al., 2017) always construct Chinese medical datasets themselves and test their models on the datasets. However, most constructed datasets used for CWS are relatively small, where there are only roughly 100 Chinese EMRs. Besides, the medical concept types in most existing datasets are limited to named entities (e.g., “Disease” and “Symptoms and Signs”), which fails to consider other medical concept types (e.g., “Time”) in EMRs that are potentially helpful for Chinese EMR texts analysis.

3 The ACEMR Corpus

3.1 Data Collection

We collected 500 Chinese EMRs from five departments (i.e., Respiratory, Gastroenterology, Urology, Gynecology, and Cardiology) of a local hospital, where one EMR specifically means the *First Course Record* in the inpatient record for one pa-

Class	Sub-class	Count
物体 Thing	Organism (Ogm)	150
	Group (Gr)	3,059
	Health Device (HD)	433
事件 Event	Health Behavior (HB)	3,093
	Events (E)	4,442
身体 Body	Body Parts (BP)	19,004
	Body Substance (BS)	1,103
	Body Function (BF)	5,179
异常 Abnormality	Signs or symptoms (SOS)	21,263
	Disease (Di)	3,543
检查 Examination	Examination Project (EP)	3,201
治疗 Treatment	Treatment Project (TP)	1,579
	Clinical Drug (Drug)	728
概念 Concept	Time (T)	4,514
	Qualitative (Ql)	14,510
	Space (Sp)	7,626
	Presence (Pre)	8,748
	Absence (Ab)	13,642
	Probability (Prob)	388
	Cause and Effect (CE)	1,359
Total	–	107,943

Table 2: The list of all medical concepts and counts.

tient. *First Course Record* refers to the first course record written by the treating physician or on-duty physician within eight hours after the patient is admitted to the hospital. It contains seven fields, namely department, ward, basic information, case characteristics, preliminary diagnosis, differential diagnosis, treatment plan, where the last five fields are illustrated in Table 1. We extract the texts in those fields and clean them by anonymizing the text and removing invalid or garbled characters.

3.2 CWS and Medical Concept Annotation

Four specialists participated in the development of the annotation guideline, where two of them are junior doctors, and the other two are PhD students in NLP. For CWS guideline, we refer to the segmentation guidelines of the Chinese Treebank (Xia, 2000) for the general domain as well as the annotation guideline proposed by He et al. (2017) for the medical domain. For medical concept annotation guideline, we refer to the medical taxonomy defined by unified medical language system (UMLS) semantic groups (Lindberg et al., 1993) and define 7 major medical concept classes with 20 sub-classes, which are elaborated in Table 2. Compared to existing medical taxonomies, our proposed medical concept classes are simple and clear with fine-grained medical concept focusing on the characteristics of Chinese EMR texts. Note that,

	Count	Length		
		Avg.	Max.	Min.
Char/Types	326,098/1,595	-	-	-
Word/Types	205,304/4,144	2.54	13	1
Sentences	7,370	43.63	311	4

Table 3: The statistics of the ACEMR corpus.

for segmentation, we do not segment one word if it is a defined medical concept.

According to the annotation guideline, we ask the two junior doctors to annotate the 500 EMRs independently and resolve their disagreements by discussion. The consistency of labeling between two annotators is evaluated by the F value (Hripcsak and Rothschild, 2005). The specific method is to treat the labeling result of one annotator (A1) as the standard answer, and calculate the F value of the labeling result of the other annotator (A2). The annotation agreement evaluated by the F value between two annotators of CWS and medical concept tagging are 0.9409 and 0.9360, respectively. We name the annotated corpus as Annotated Chinese Electronic Medical Record (ACEMR) and report its statistics in Table 3, where the lengths are computed based on Chinese characters. In addition, the number of medical concepts in ACEMR is also reported in the last column of Table 2.

Table 4 shows two annotated example sentences, where Chinese words are split by white spaces². The medical concept tag attached to a specific word is highlighted in red color (“/” is a delimiter between a word and its medical concept tag).

3.3 The Corpus Properties

ACEMR is an informative Chinese medical dataset. It contains 500 Chinese EMR texts that are annotated with CWS labels and medical concepts from 20 sub-classes. Due to space limitations, among 20 sub-classes, we introduce three sub-classes (i.e., *Group*, *Health Behavior*, and *Qualitative*) in the following texts. *Group* includes the patient’s gender, age, and name. It generally appears at the beginning of Chinese EMRs as part of the basic information, indicating the group the patient belongs to. In addition, it can also act as a participant in medical and health activities (i.e. patients and doctors). *Health Behavior* means medical-related behaviors. It mainly includes examination behaviors, diagnostic behaviors, and broad non-specific treat-

²If a Chinese word is translated into multiple English words, we use “*” in the English translation to mark its boundary in Table 4. E.g., “3天” is translated into “*3 days*”.

患者/Gr 老年/Gr 女性/Gr , 慢性/Ql 病程/Di , 急性/Ql 加重/SOS 。患者/Gr 主/Ql 因/CE " 反复/Ql 咳嗽/SOS 、 咳痰/SOS , 加重/SOS 3天/T " 入院/E 。

Patient/Gr elderly/Gr female/Gr , chronic/Ql course/Di , acute/Ql exacerbation/SOS . The main/Ql cause/CE of the patient/Gr was " repeated/Ql cough/SOS and sputum/SOS , which became worse/SOS for *3 days*/T " and was *admitted to the hospital*/E .

Table 4: An example of annotated medical sentence in ACEMR with the corresponding English translations. The abbreviations of tags are used for annotation.

ment behaviors. E.g., “予” (*given*), “入院治疗” (*admission to hospital for treatment*). *Qualitative* emphasizes a qualitative description of something, rather than a direct measurement and can be used to describe the body, abnormalities, etc. E.g., “胃肠型感冒” (*gastrointestinal cold*) where “胃肠型” (*gastrointestinal*) are *Qualitative* medical concepts.

4 Methods

A good text representation is highly important in achieving a promising performance in many NLP tasks (Song et al., 2017; Liu and Lapata, 2018; Song and Shi, 2018). Therefore, we select several well-known models for CWS and medical concept recognition tasks and test them on ACEMR corpus.

4.1 CWS for Chinese EMR

For CWS, we follow the convention in previous CWS studies (Sun and Xu, 2011; Song et al., 2012; Song and Xia, 2013; Chen et al., 2015; Zhang et al., 2016; Qiu et al., 2019) to regard it as a sequence labeling task with the “BIES” scheme. We select four well-know models, namely, BiLSTM, BERT (Devlin et al., 2019), ZEN (Diao et al., 2020), and WMSeg (Tian et al., 2020d) with softmax and CRF decoder. Herein, BERT and ZEN are pre-trained language models that have achieved state-of-the-art performance in many NLP tasks (Liang et al., 2020; Tian et al., 2020c; Yu et al., 2020; Nie et al., 2020; Luoma and Pyysalo, 2020; Chen et al., 2020a; Helwe et al., 2020; Tian et al., 2021a,b). WMSeg is CWS model that leverages key-value memory networks (KVMN) (Miller et al., 2016) to incorporate wordhood information to improve model performance, which achieves state-of-the-art performance on many CWS benchmark datasets.

4.2 Medical Concept Recognition

Similarly, for medical concept recognition, we regard it as a character-based sequence labeling task and perform it in a similar way with named entity

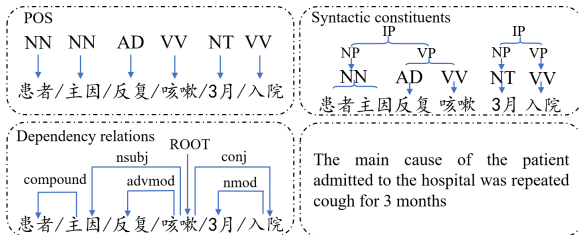


Figure 1: The auto-generated syntactic information (i.e., POS labels, dependency relations, and syntactic constituents) of a sentence with English translation.

Dataset	Word Counts	Word Types
Train	169,047	3,833
Test	36,257	1,529

Table 5: The statistics (i.e., word count and word type) of the training and test sets of ACEMR.

recognition, where the medical concept tags for the input characters follow the “BIOES” scheme. For example, “支气管” (“virus”) has a medical tag sub-class “BP”, and thus the tags for the three characters are “B-BP”, “I-BP”, and “E-BP”, respectively. We try BiLSTM, BERT, ZEN, as well as TwASP (Tian et al., 2020b) with the CRF decoder for medical concept recognition. TwASP is a model that leverages the auto-generated syntactic information (e.g., the POS tags (POS), the dependency relations (Dep.), and the syntactic constituents (Syn.)) through a two-way attention mechanism to improve model performance for sequence labeling tasks. To obtain the syntactic information of the input sentence required by TwASP, we use Stanford CoreNLP Toolkits (Manning et al., 2014) to obtain the POS tags, the dependency tree, and the constituent syntax tree. Figure 1 shows an example sentence (with English translation) and the three types of the auto-generated syntactic information.

5 Experiments

In the experiments, we use two datasets. The first is the in-domain ACEMR corpus introduced in Sec. 3; the second is CTB6 (Xue et al., 2005), which is a benchmark CWS dataset of the general domain text. We split the ACEMR corpus into training/test sets and report the statistics in Table 5. For all experiments, we use precision (Prec.), recall, and F1 scores to evaluate different models.

5.1 Performance on Medical CWS

For medical CWS, we try BiLSTM, BERT, ZEN, and WMSeg³. For BiLSTM, we use pre-trained

³<https://github.com/SVAIGBA/WMSeg>

Methods	Prec.	Recall	F1
<i>CTB Only</i>			
WMSeg	77.60	76.85	77.22
<i>*ZEN is the base model</i>			
<i>CTB+ACEMR</i>			
BiLSTM	98.01	98.09	98.05
+ CRF	98.22	98.30	98.26
+ Tencent Embedding	98.75	98.68	98.72
BERT	98.32	98.65	98.48
+ CRF	98.40	98.66	98.53
+ KVMN	98.55	98.78	98.69
ZEN	98.51	98.89	98.70
+ CRF	98.70	98.81	98.76
+ KVMN	98.86	98.84	98.85
<i>ACEMR Only</i>			
ZEN	99.01	99.00	99.00
+ CRF	98.99	98.91	98.94
+ KVMN	99.03	99.04	99.03

Table 6: CWS performance for different composition of training data where +CRF, +KVMN, +Tencent Embedding represent the use of CRF layer, memory network (WMSeg) and Tencent Embedding respectively.

character embeddings from Tencent Embedding⁴ (Song et al., 2018), with the training epoch, batch size, and learning rate set to 50, 32, and 0.001, respectively. For BERT, ZEN, and WMSeg, we use the official settings (e.g., 768 dimensional hidden vectors with 12 multi-head self-attentions for BERT), where the number of training epoch is 50, the batch size is 16, and the learning rate is 1e-5.

The experimental results of CWS are presented in Table 6 with three different settings (i.e., *CTB Only*, *CTB+ACEMR*, and *ACEMR Only*). The *CTB Only* setting displays the results of WMSeg model (with ZEN encoder) when it is trained on CTB6 only and evaluated on the ACEMR test set. The inferior results confirm the big gap between the texts and guidelines in general and medical domains, which indicates the challenge to perform transfer learning from the general domain to the medical domain. The *CTB+ACEMR* setting shows the results of all models trained on the combination of ACEMR and CTB6 datasets, where all models have a high improvement compared with the WMSeg model trained on CTB6 only, emphasizing the necessity of constructing an annotated dataset in medical domain. Compared with BERT and ZEN baseline, adding the KVMN module at the top of the BERT/ZEN encoder to leverage wordhood information (which is exactly the architecture of

⁴We use the official release from <https://ai.tencent.com/ailab/nlp/zh/embedding.html>.

Methods	Prec.	Recall	F1
BiLSTM-CRF	95.65	95.41	95.53
BERT-CRF	97.62	97.84	97.73
ZEN-CRF	97.00	97.87	97.82

Table 7: The results on three different well-known models on medical concept recognition.

Concept	F1	Count	OOV
<i>Top 3 sub-classes</i>			
Probability (Prob)	100.00	372	0.000
Group (Gr)	99.84	2,440	0.177
Absence (Ab)	99.74	10,964	0.053
<i>Bottom 3 sub-classes</i>			
Treatment Project (TP)	90.91	1,380	0.291
Clinical Drug (Drug)	88.20	648	0.462
Body Substance (BS)	74.24	976	0.400
Total	97.82	-	-

Table 8: The top and bottom 3 results of ZEN-CRF on each sub-classes of medical concept recognition, where the number of medical concepts belonging to each sub-class in training set and the out-of-vocabulary (OOV) rate in test set are reported in last two columns.

WMSeg) can improve the performance on CWS. In addition, models with ZEN encoder achieve higher performance than the ones with BERT, which may result from the fact that ZEN leverage n-gram information during pre-training and thus can obtain a better contextual representation. Moreover, if we train the model on ACEMR only (i.e., the *ACEMR only* setting), models with ZEN encoder can be further improved. This observation is not surprising because the texts in CTB6 from the general domain could introduce noise into the model.

5.2 Performance on Concept Recognition

For medical concept recognition (MCR) task where the gold CWS results are given, the results from BiLSTM, BERT, and ZEN encoder with CRF decoder are reported in Table 7, where ZEN-CRF achieves the highest performance. In addition, we rank the F1 scores of all sub-class labels obtained by ZEN-CRF and present the results of the top and bottom 3 ones in Table 8, where the number of medical concepts belonging to each sub-class in the training set as well as the rate of out-of-vocabulary (OOV) medical concepts in the test set is also reported. It is observed that the model does not perform well on sub-classes with fewer training instances and higher OOV rate (e.g., *Body Substance*), which suggests that the OOV issue is a challenge for Chinese medical concept recognition.

Methods	Prec.	Recall	F1
BERT-CRF	97.62	97.84	97.73
TwASP (POS)	97.74	98.04	97.89
TwASP (Dep.)	97.85	98.02	97.94
TwASP (Syn.)	97.65	97.93	97.79
ZEN-CRF	97.00	97.78	97.82
TwASP (POS)	97.77	98.00	97.85
TwASP (Dep.)	97.64	98.01	97.90
TwASP (Syn.)	97.52	97.74	97.63

Table 9: The results of TWASP on medical concept recognition with auto-generated POS labels, dependencies (Dep.), and syntactic constituents (Syn.).

In addition, we run TwASP⁵ with three different types of auto-generated syntactic information (i.e., POS labels, dependency relations, and syntactic constituents). The results are reported in Table 9, where we find that MCR can benefit from syntactic information and obtain improvement in most cases, although BERT-CRF and ZEN-CRF baselines have already achieve outstanding performance.

6 Conclusion

In this paper, we collect a new Chinese medical corpus, named ACEMR, which contains 500 EMRs from a local hospital, and annotate the corpus with CWS and medical concept labels. ACEMR features in the rich types of medical concept, in which 20 sub-classes of medical concepts are annotated. We test several state-of-the-art models for CWS and medical concept recognition on the annotated ACEMR. The results on CWS show that models trained on general domain dataset (i.e., CTB6) cannot perform well on medical domain, which confirms the necessity of constructing the ACEMR corpus. Furthermore, WMSeg with wordhood information and TwASP with auto-generated syntactic information outperforms strong baselines on word segmentation and medical concept recognition, respectively, which demonstrates the benefit of leveraging extra resources (i.e., wordhood information and syntactic information) for CWS and medical concept recognition.

Acknowledgements

This work is supported by Chinese Key-Area Research and Development Program of Guangdong Province (2020B0101350001) and also partially supported by NSFC under the project "The Essential Algorithms and Technologies for Standardized Analytics of Clinical Texts" (12026610).

⁵<https://github.com/SVAIGBA/TwASP>

References

- Guimin Chen, Yuanhe Tian, and Yan Song. 2020a. Joint Aspect Extraction and Sentiment Analysis with Directional Graph Convolutional Networks. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 272–279.
- Xinchu Chen, Xipeng Qiu, Chenxi Zhu, Pengfei Liu, and Xuanjing Huang. 2015. Long Short-Term Memory Neural Networks for Chinese Word Segmentation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1197–1206.
- Zhihong Chen, Yan Song, Tsung-Hui Chang, and Xiang Wan. 2020b. Generating Radiology Reports via Memory-driven Transformer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1439–1449, Online.
- Jacob Devlin, Ming Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Shizhe Diao, Jiaxin Bai, Yan Song, Tong Zhang, and Yonggang Wang. 2020. ZEN: Pre-training Chinese Text Encoder Enhanced by N-gram Representations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 4729–4740.
- Carol Friedman, Lyudmila Shagina, Yves Lussier, and George Hripcsak. 2004. Automated Encoding of Clinical Documents based on Natural Language Processing. *Journal of the American Medical Informatics Association*, 11(5):392–402.
- Bin He, Bin Dong, Yi Guan, Jinfeng Yang, Zhipeng Jiang, Qiubin Yu, Jianyi Cheng, and Chunyan Qu. 2017. Building a Comprehensive Syntactic and Semantic Corpus of Chinese Clinical Texts. *Journal of biomedical informatics*, 69:203–217.
- Chadi Helwe, Ghassan Dib, Mohsen Shamas, and Shady Elbassuoni. 2020. A Semi-Supervised BERT Approach for Arabic Named Entity Recognition. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, pages 49–57.
- George Hripcsak and Adam S Rothschild. 2005. Agreement, the F-measure, and Reliability in Information Retrieval. *Journal of the American medical informatics association*, 12(3):296–298.
- Theresa A Koleck, Nicholas P Tatonetti, Suzanne Bakken, Shazia Mitha, Morgan M Henderson, Maureen George, Christine Miaskowski, Arlene Smaldone, and Maxim Topaz. 2021. Identifying Symptom Information in Clinical Notes Using Natural Language Processing. *Nursing research*.
- Xiaozheng Li, Huazhen Wang, Huixin He, Jixiang Du, Jian Chen, and Jinzhun Wu. 2019. Intelligent Diagnosis with Chinese Electronic Medical Records based on Convolutional Neural Networks. *BMC bioinformatics*, 20(1):1–12.
- Yu-Bing Li, Xue-Zhong Zhou, Run-Shun Zhang, Ying-Hui Wang, Yonghong Peng, Jing-Qing Hu, Qi Xie, Yan-Xing Xue, Li-Li Xu, Xiao-Fang Liu, et al. 2015. Detection of Herb-Symptom Associations from Traditional Chinese Medicine Clinical Data. *Evidence-Based Complementary and Alternative Medicine*, 2015.
- Chen Liang, Yue Yu, Haoming Jiang, Siawpeng Er, Ruijia Wang, Tuo Zhao, and Chao Zhang. 2020. Bond: Bert-Assisted Open-Domain Named Entity Recognition with Distant Supervision. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Donald Lindberg, Betsy Humphreys, and Alexa McCray. 1993. The Unified Medical Language System. *Methods of information in medicine*, 32(4):281.
- Yang Liu and Mirella Lapata. 2018. Learning Structured Text Representations. *Transactions of the Association for Computational Linguistics*, 6:63–75.
- Jouni Luoma and Sampo Pyysalo. 2020. Exploring Cross-sentence Contexts for Named Entity Recognition with BERT. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 904–914.
- Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland.
- Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. 2016. Key-Value Memory Networks for Directly Reading Documents. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1400–1409.
- Yuyang Nie, Yuanhe Tian, Xiang Wan, Yan Song, and Bo Dai. 2020. Named Entity Recognition for Social Media Texts with Semantic Augmentation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1383–1391, Online.
- Xipeng Qiu, Hengzhi Pei, Hang Yan, and Xuanjing Huang. 2019. Multi-Criteria Chinese Word Segmentation with Transformer. *arXiv preprint arXiv:1906.12035*.
- Alex Rector, Wanda Nowlan, and Shazia Kay. 1991. Foundations for an Electronic Medical Record. *Methods of information in medicine*, 30(03):179–186.

- Yan Song, Prescott Klassen, Fei Xia, and Chunyu Kit. 2012. Entropy-based Training Data Selection for Domain Adaptation. In *Proceedings of COLING 2012: Posters*, pages 1191–1200.
- Yan Song, Chia-Jung Lee, and Fei Xia. 2017. Learning Word Representations with Regularization from Prior Knowledge. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 143–152.
- Yan Song and Shuming Shi. 2018. Complementary Learning of Word Embeddings. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4368–4374.
- Yan Song, Shuming Shi, Jing Li, and Haisong Zhang. 2018. Directional Skip-Gram: Explicitly Distinguishing Left and Right Context for Word Embeddings. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 175–180.
- Yan Song, Yuanhe Tian, Nan Wang, and Fei Xia. 2020. Summarizing Medical Conversations via Identifying Important Utterances. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 717–729.
- Yan Song and Fei Xia. 2013. A Common Case of Jekyll and Hyde: The Synergistic Effect of Using Divided Source Training Data for Feature Augmentation. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 623–631.
- Mark Stevenson, Eneko Agirre, and Aitor Soroa. 2012. Exploiting Domain Information for Word Sense Disambiguation of Medical Documents. *Journal of the American Medical Informatics Association: JAMIA*, 19(2):235–240.
- Weiwei Sun and Jia Xu. 2011. Enhancing Chinese Word Segmentation Using Unlabeled Data. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 970–979.
- Yuanhe Tian, Guimin Chen, and Yan Song. 2021a. Aspect-based Sentiment Analysis with Type-aware Graph Convolutional Networks and Layer Ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*.
- Yuanhe Tian, Guimin Chen, and Yan Song. 2021b. Enhancing Aspect-level Sentiment Analysis with Word Dependencies. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3726–3739, Online.
- Yuanhe Tian, Weicheng Ma, Fei Xia, and Yan Song. 2019. ChiMed: A Chinese Medical Corpus for Question Answering. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 250–260, Florence, Italy.
- Yuanhe Tian, Wang Shen, Yan Song, Fei Xia, Min He, and Kenli Li. 2020a. Improving Biomedical Named Entity Recognition with Syntactic Information. *BMC Bioinformatics*, 21:1471–2105.
- Yuanhe Tian, Yan Song, Xiang Ao, Fei Xia, Xiaojun Quan, Tong Zhang, and Yonggang Wang. 2020b. Joint Chinese Word Segmentation and Part-of-speech Tagging via Two-way Attentions of Auto-analyzed Knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8286–8296.
- Yuanhe Tian, Yan Song, and Fei Xia. 2020c. Joint Chinese Word Segmentation and Part-of-speech Tagging via Multi-channel Attention of Character N-grams. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2073–2084.
- Yuanhe Tian, Yan Song, Fei Xia, Tong Zhang, and Yonggang Wang. 2020d. Improving Chinese Word Segmentation with Wordhood Memory Networks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8274–8285.
- Nan Wang, Yan Song, and Fei Xia. 2020. Studying Challenges in Medical Conversation with Structured Annotation. In *Proceedings of the First Workshop on Natural Language Processing for Medical Conversations*, pages 12–21, Online.
- Qi Wang, Yangming Zhou, Tong Ruan, Daqi Gao, Yuhang Xia, and Ping He. 2019. Incorporating Dictionaries into Deep Neural Networks for the Chinese Clinical Named Entity Recognition. *Journal of biomedical informatics*, 92:103133.
- Fei Xia. 2000. The Segmentation Guidelines for the Penn Chinese Treebank (3.0).
- Junjie Xing, Kenny Zhu, and Shaodian Zhang. 2018. Adaptive Multi-task Transfer Learning for Chinese Word Segmentation in Medical Text. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3619–3630.
- Dong Xu, Meizhuo Zhang, Tianwan Zhao, Chen Ge, Weiguo Gao, Jia Wei, and Kenny Q Zhu. 2015. Data-driven Information Extraction from Chinese Electronic Medical Records. *PLoS one*, 10(8):e0136270.
- Yan Xu, Yining Wang, Tianren Liu, Jiahua Liu, Yubo Fan, Yi Qian, Junichi Tsujii, and Eric I Chang. 2014. Joint Segmentation and Named Entity Recognition using Dual Decomposition in Chinese discharge summaries. *Journal of the American Medical Informatics Association*, 21(e1):e84–e92.

- Naiwen Xue, Fei Xia, Fu-Dong Chiou, and Marta Palmer. 2005. The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus. *Natural language engineering*, 11(2):207.
- Yajiong Xue, Huigang Liang, Xiaocheng Wu, Hai Gong, Bin Li, and Yuxia Zhang. 2012. Effects of Electronic Medical Record in a Chinese hospital: A Time Series Study. *International journal of medical informatics*, 81(10):683–689.
- Juntao Yu, Bernd Bohnet, and Massimo Poesio. 2020. Named Entity Recognition as Dependency Parsing. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6470–6476.
- Shaodian Zhang, Tian Kang, Xingting Zhang, Dong Wen, Noémie Elhadad, and Jianbo Lei. 2016. Speculation Detection for Chinese Clinical Notes: Impacts of Word Segmentation and Embedding Models. *Journal of biomedical informatics*, 60:334–341.