

Article

Toward Sustainable Virtualized Healthcare: Extracting Medical Entities from Chinese Online Health Consultations Using Deep Neural Networks

Hangzhou Yang and Huiying Gao *

School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China; hangzhou@bit.edu.cn

* Correspondence: huiying@bit.edu.cn; Tel.: +86-10-68918919

Received: 29 August 2018; Accepted: 12 September 2018; Published: 14 September 2018



Abstract: Increasingly popular virtualized healthcare services such as online health consultations have significantly changed the way in which health information is sought, and can alleviate geographic barriers, time constraints, and medical resource shortage problems. These online patient–doctor communications have been generating abundant amounts of healthcare-related data. Medical entity extraction from these data is the foundation of medical knowledge discovery, including disease surveillance and adverse drug reaction detection, which can potentially enhance the sustainability of healthcare. Previous studies that focus on health-related entity extraction have certain limitations such as demanding tough handcrafted feature engineering, failing to extract out-of-vocabulary entities, and being unsuitable for the Chinese social media context. Motivated by these observations, this study proposes a novel model named CNMER (Chinese Medical Entity Recognition) using deep neural networks for medical entity recognition in Chinese online health consultations. The designed model utilizes Bidirectional Long Short-Term Memory and Conditional Random Fields as the basic architecture, and uses character embedding and context word embedding to automatically learn effective features to recognize and classify medical-related entities. Exploiting the consultation text collected from a prevalent online health community in China, the evaluation results indicate that the proposed method significantly outperforms the related state-of-the-art models that focus on the Chinese medical entity recognition task. We expect that our model can contribute to the sustainable development of the virtualized healthcare industry.

Keywords: medical entity extraction; deep neural networks; online health consultations; conditional random fields; virtualized healthcare; long short-term memory

1. Introduction

Healthcare has drawn considerable attention in recent years, and increasing numbers of patients are engaging in online health communities (OHCs) for health information exchange [1–3]. Online health communities are becoming an essential channel for users to search for health information and share their experiences of medical treatments [4]. According to Health Information National Trends Survey 2017, about 80 percent of adults in the U.S. search for health-related information online [5]. In China, there were around 195 million people using online medical services by the end of 2016 [6]. With the rapid growth of healthcare service delivery, a number of new models have been developed recently, including online health consultations [7]. Patients not only interact with their peers, but also consult doctors about their diseases through online communities [8], which forms a new communication channel between patients and doctors. This new form of online patient–doctor communication has greatly changed the traditional delivery model of healthcare service.

Online communication between patients and physicians can potentially alleviate the medical resource shortage problem and eliminate geographic barriers and time constraints to some extent [9].

Online health consultations generate large amounts of valuable health-related information [10]. The wide spread of periodic general health examinations [11] also contributes to the fast growth of the medical datasets available. The vast development of information and communication technologies dramatically improves the deposition and exchange of health-related data, which facilitates healthcare Big Data analytics. As one of the Sustainable Development Goals (SDGs), sustainable healthcare is dedicated to ensuring healthy lives and promoting well-being for all people. Medical-related entity extraction from online health consultations can contribute to the sustainability of visualized healthcare in the following aspects. First, the extracted entities in online health consultations can facilitate the procedure of online patient–doctor communication by automatically recognizing and classifying the critical health concepts in patient- and doctor-generated text. The high efficiency of online health consultation helps improve the convenience and flexibility, saving costs and time for healthcare service delivery [7,12]. It can potentially support users to manage their health conditions electronically and thus attain more promising health outcomes and reduce future health risks [13]. OHCs can also benefit from entity extraction by attracting more participants to engage in the information exchange platforms. Second, medical entity recognition is an essential task in clinical information extraction and medical knowledge discovery [14], and can facilitate a number of healthcare-related applications such as disease surveillance [15] and adverse drug reaction detection [16]. Early detection of disease activity can reduce the impact of certain diseases such as seasonal influenza with a rapid response [17]. Adverse drug reactions are among the top causes of morbidity and mortality and have been drawing considerable public attention [18]. Disease surveillance and adverse drug reaction detection using social media data can enhance public health monitoring and ensure a healthier life [19].

In this study, we aim to recognize several types of medical entities, namely, medical problems, medical tests, and treatments [20], which are critical health concepts in medical knowledge discovery. Medical problem entity recognition aims to identify the states of diseases or symptoms in text to extract the health conditions of a patient, such as “breast cancer” and “fever”. The medical test entity recognition task seeks to find the medical examinations mentioned in text, including laboratory tests and physical examinations such as “blood test” and “CT scan”. Treatment entity recognition attempts to extract the mentions of therapy in medical text including drug names and surgery procedures, such as “glucose” and “heart transplantation”. For example, in the post “My right face was slightly swollen and accompanied by fever. . . . I didn’t feel better after taking glucocorticoid. After a blood test and other thorough examination, it was diagnosed as a facial lymphoma and now I’m ready for chemotherapy”. In this post, “slightly swollen”, “fever”, and “facial lymphoma” are medical problem entities; “blood test” is a medical test entity; and “glucocorticoid”, “chemotherapy” are treatment entities.

Extensive studies have been conducted to extract medical-related entities. Lexical-based methods recognize an entity by matching to the most similar or identical terms in a dictionary [21], which makes lexical-based approaches particularly useful for practical information extraction [22]. In the medical field, the most widely used controlled terminology dictionaries include UMLS (Unified Medical Language System) [23], ICD (International Classification of Diseases) [24], and SNOMED CT (Systematized Nomenclature of Medicine–Clinical Terms) [25]. However, the short terms in the dictionary would result in false positives and significantly degrade the overall accuracy, and spelling variations which are quite common in the social media context make lexical-based approaches less usable. Machine learning approaches have been widely adopted for entity recognition because of their excellent environmental adaptability. The commonly used algorithms in entity recognition tasks include Maximum Entropy (ME) [26], Support Vector Machine (SVM) [27], Hidden Markov Model (HMM) [28], and Conditional Random Fields (CRF) [16,29,30]. Despite their excellent performance in some studies, the machine-learning-based models usually need tough feature engineering work. In recent years, the rapid improvement of deep learning techniques has brought new opportunities for natural language processing (NLP) studies including entity extraction [31–33], and has significantly

contributed to overcoming the above problem. Due to their capacity for automatically learning effective features from word embedding, deep neural network (DNN)-based models such as recurrent neural networks (RNNs) have been employed in state-of-the-art models [31,34]. As a unique RNN architecture, Long Short-Term Memory (LSTM) and its variant bidirectional LSTM (BiLSTM) have been utilized in entity extraction tasks and have shown encouraging performance [31,34,35].

Although medical entity extraction has been widely studied, existing approaches have several limitations when applied to the Chinese social media context. First, most traditional machine learning approaches require complicated feature engineering work [16,23,27–30]. Feature engineering relies on handcrafted rules and language-specific knowledge, which is inherently tough and time-consuming [34]. Second, most current works are designed for the English language context and ignore the uniqueness of Chinese. Compared with English, there is no blank space between words, and seldom do morphological changes exist in Chinese, making existing entity extraction approaches challenging in the context of Chinese. Third, unlike clinical notes that are described by healthcare professionals, the content in social media could be extensively informal, with features such as lexical variants, internet slang, typos, and grammatical errors. Previous approaches that used clinical notes as a data resource may fail to recognize out-of-vocabulary (OOV) terms, resulting in unsatisfying entity extraction performance [36].

Recognizing the significance of medical entity extraction and the limitations of existing related works, this study proposes a novel DNN-based approach to extracting medical entities in the context of Chinese social media that can overcome the aforementioned problems. This study intends to enhance the sustainability of online healthcare services and public health monitoring by improving the performance of health concept extraction in online health consultations. Recent developments in DNN have achieved great success in many areas, providing new opportunities for natural language processing (NLP) research [31,33]. Specifically, we aimed to design a model that can automatically capture the context features of text to avoid tough feature engineering work and is effective in medical entity recognition in Chinese social media text. Considering the uniqueness of Chinese, we also evaluate the effect of recognition granularity on the performance of the entity extraction.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed medical entity extraction model, followed by the evaluation procedure in Section 3. The experimental results are presented in Section 4. Section 5 discusses the evaluation results and reviews the practical implications of our model for the healthcare system. Lastly, we conclude our major research findings and research limitations in Section 6.

2. Method

This study proposes a novel DNN-based model named CNMER (Chinese Medical Entity Recognition) to extract medical entities from Chinese OHCs. Figure 1 depicts an overview of our approach. After data collection, preprocessing was performed, and a subset of the processed data was randomly selected for data annotation. The remaining unlabeled dataset was utilized as the text corpus for unsupervised training on the domain word and character embeddings. With the part-of-speech (POS) feature and position feature, the trained embeddings were then used to formulate the character representation as the input for the BiLSTM-CRF.

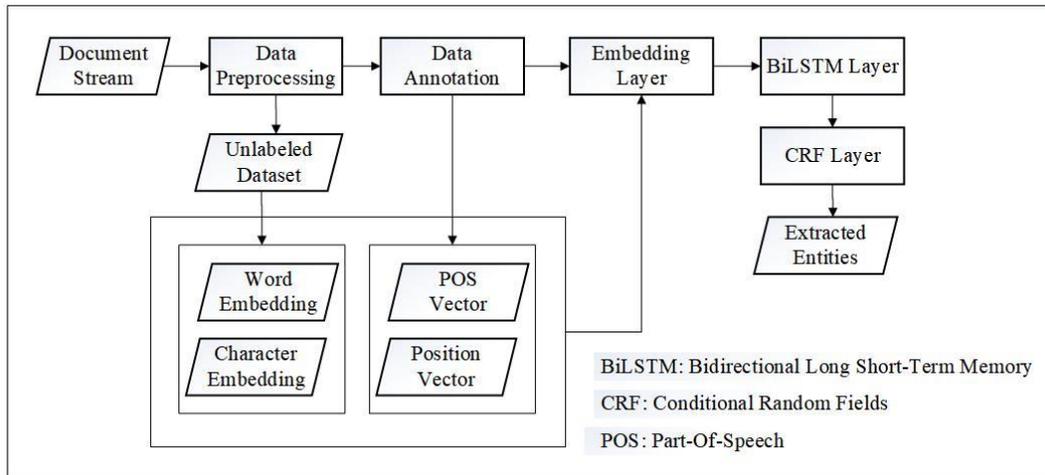


Figure 1. An overview of the CNMER (Chinese Medical Entity Recognition) model.

As shown in Figure 2, the BiLSTM-CRF architecture consists of an embedding layer, a BiLSTM layer, and a CRF layer. The embedding layer maps each character in a sentence using the predefined numerical representation vector. The BiLSTM layer includes forward LSTM and backward LSTM, and takes the representation vectors of the character sequence as input and returns another sequence by considering both left and right context information. The CRF layer makes final tagging decisions based on the output of the BiLSTM layer using the CRF model.

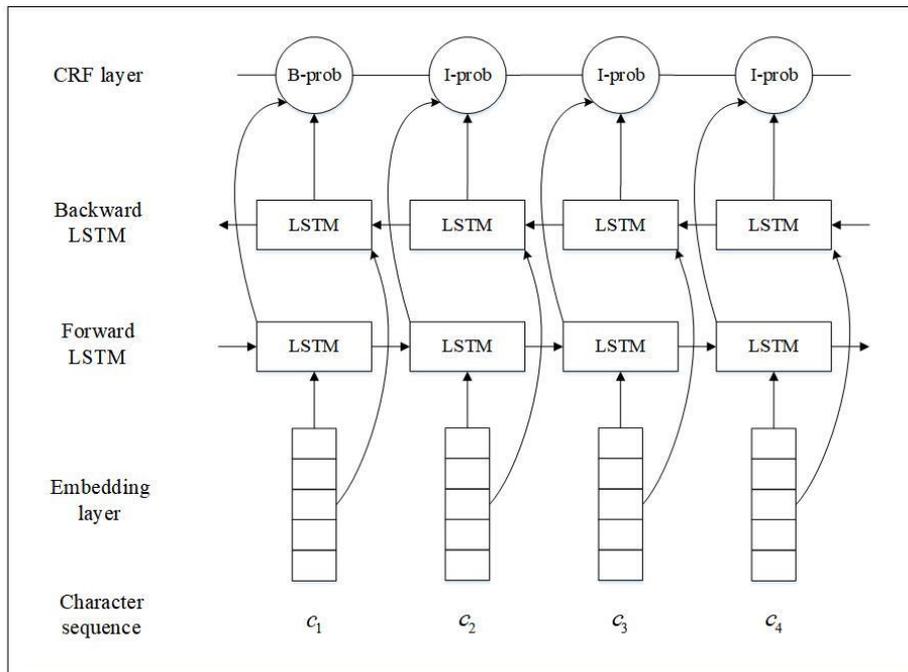


Figure 2. An overview of the BiLSTM-CRF architecture.

2.1. Data Preprocessing and Annotation

The communications between physicians and patients in OHCs generate abundant health-related text. In this study, we exploited the online consultation text as the data resource. First, data preprocessing was performed to remove irrelevant contents such as private information, html tags, and other invalid characters. We also filtered out the consultations that were shorter than five characters. Unlike English text, words are not separated by blank spaces in Chinese sentences; thus, word segmentation was conducted to split each word in a sentence. We utilized Jieba, an open source NLP application in the Python language

to segment sentences into words and perform POS tagging, for which a total of 40 types of POS tags were predefined. In this study, we employed Chinese Unified Medical Language System (CUMLS) [37], a repository of biomedical terminologies developed by Chinese Academy of Medical Sciences to help improve the performance of Chinese word segmentation for the health-related corpus. CUMLS integrates more than ten biomedical sources such as biomedical thesauri, classifications, and text words of biomedical literature. Specifically, CUMLS includes 100,000 medical terms. Using CUMLS as the supplementary dictionary, terms in consultations that are matched with the repository can be extracted and segmented as a single word automatically.

After data preprocessing, we randomly selected a small subset from the obtained corpus as the source of data annotation. An annotation protocol was developed before annotation. To obtain the annotation dataset, two expert annotators were recruited to independently label the entity boundaries and types in sentences. Another expert annotator was asked to check any disagreements and make the final judgement. In this study, we labeled entities using the “BIO” tagging formalism, where the “B” category represents the beginning of an entity, the “I” category represents the continuity of an entity, and “O” denotes all other characters. As an illustration, for a medical problem entity which consists of four characters in total, namely c_1, c_2, c_3, c_4 , the annotators are supposed to tag the character sequence as “B-prob, I-prob, I-prob, I-prob”.

2.2. Embedding Layer

Conventional machine learning approaches lack the ability to process natural data in their raw form and require careful engineering and designing work to extract effective features from raw data such as plain text [38]. The input of machine learning approaches is usually represented in the form of a fixed-length feature vector. For text input, bag-of-words is one of the most common used features. Although they have been widely used, bag-of-words features have certain disadvantages: they fail to capture the order of words in text and they miss the semantic information of words. For example, the word “sickness”, “illness”, and “hospital” are represented with the same distance by bag-of-words, although “sickness” should be closer to “illness” than to “hospital”, semantically.

Distributed representations of words in the form of a vector space can group similar words and facilitate many natural language processing tasks toward better performance. Representation learning approaches can automatically detect the information needed and represent it at a higher and more abstract level. A word embedding maps a word to a numerical vector in a low-dimensional vector space which can capture semantic or syntactic properties of the word; semantically similar words are expected to be assigned similar vectors [34]. The learned word representations explicitly encode many linguistic regularities and patterns, and many of these patterns can be represented as linear translations [39]. For example, the result of calculation $\text{vec}(\text{“Beijing”}) - \text{vec}(\text{“China”}) + \text{vec}(\text{“Japan”})$ is closer to $\text{vec}(\text{“Tokyo”})$ than to any other learned word vectors, where “vec” represents the learned embedding vector of a word. This study uses the skip-gram method for both word- and character-level embedding training [39], which predicts the words that are most likely to appear around the focused word. Given a sequence of training words w_1, w_2, \dots, w_T , the model is trained by maximizing the average log probability

$$\frac{1}{T} \sum_{t=1}^T \sum_{-s \leq j \leq s, j \neq 0} \log p(w_{t+j} | w_t), \quad (1)$$

where s is the size of training context, and w_{t+j} are the words surrounding the focused w_t . The basic skip-gram formulation defines $p(w_{t+j} | w_t)$ using the Softmax function

$$p(w_o | w_i) = \frac{\exp(v'_{w_o} v_{w_i})}{\sum_{w=1}^W \exp(v'_{w} v_{w_i})}, \quad (2)$$

where v_w and v'_w denote the “input” and “output” vector representations of w , respectively, and W represents the total number of words in the vocabulary [39]. We use word2vec, an open source tool

developed by Google, to train the character and word embedding [39]. We trained a 100-dimensional embedding for both characters and words based on the unlabeled dataset [33].

For Chinese online health-related text, word segmentation is a challenging task [40] which could result in unsatisfying performance for word-based entity extraction methods. To address this issue, character-based entity recognition was proposed [41]. The character representation has been recognized as an important factor that impacts the entity recognition performance [32,33]. However, the semantic information of a character varies according to context, while the same character in different contexts is usually represented by the same embedding. Therefore, the direct use of character-level embedding for the various contexts will lead to inaccurate character feature representation [33]. In this study, we propose to combine character embedding with the context word embedding as a part of the character embedding vector. Thus, the character representation incorporates not only the features of the focal character, but also the context information of the related word.

The POS feature and the position of a character in the context word [42] were also incorporated into our model as they carry critical context information for the focused character. According to the tagging scheme in Jieba, we predefined a list of POS tags and mapped each tag to a 40-dimensional one-hot vector to represent the POS feature of the context word. To represent the position feature of a character, we used a 4-dimensional one-hot vector to present the positional information of a character in the context word: single-character word, the beginning of a word, the middle of a word, or the end of a word. All the embeddings and vectors were then concatenated together as a single vector, and finally we obtained a 244-dimensional numerical representation for each character as input for the BiLSTM network. Figure 3 illustrates an example of character representation used in our model, where d_c represents the dimension of the character embedding and d_w indicates the dimension of the word embedding. In the example, the entity is divided into two parts during word segmentation, namely, w_1 which consists of c_1, c_2, c_3, c_4 , and w_2 , which consists of c_5 . Therefore, the representation vector of the character c_1 consists of four parts, which are the character embedding of c_1 , the word embedding of w_1 , the POS feature vector of the word w_1 , and the position feature (i.e., “the beginning of a word”) vector of the character c_1 in the word w_1 .

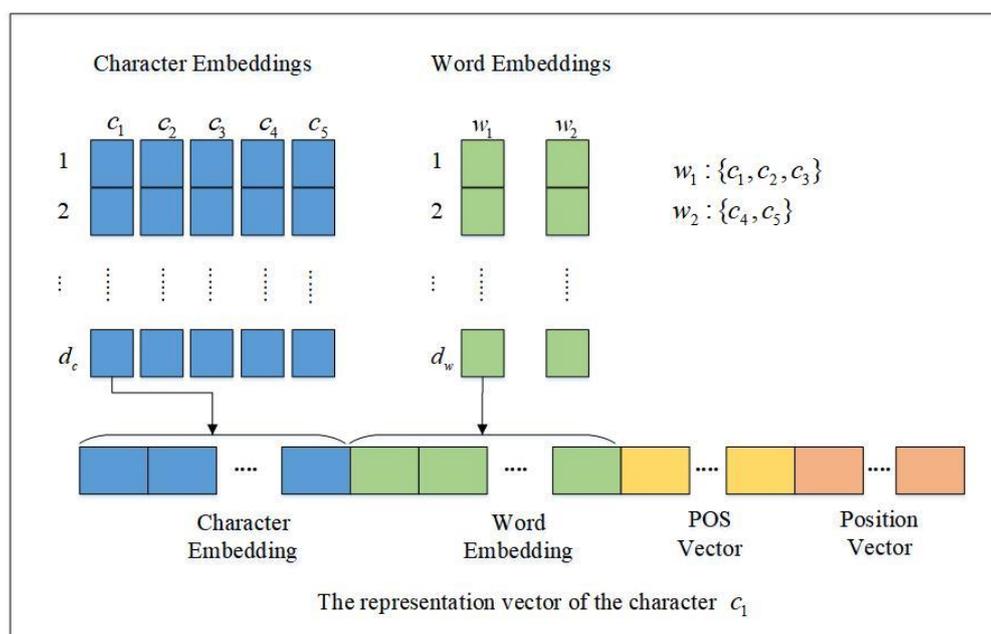


Figure 3. The structure of a character representation vector.

2.3. BiLSTM Layer

A typical neural network contains a set of input units, multiple hidden layers that contain hidden units, a set of output units that stands for tags, and the connections between those units [43]. The model

is trained using an algorithm named “back-propagation” to adjust the weights of connections between units, so that any input tends to generate the corresponding output. The relationship between inputs and outputs that a neural network learns can be regarded as a mapping, and neural networks with multiple hidden layers are believed to be good at learning mappings.

Deep neural networks are neural networks with a large number of hidden layers. A deep neural network system is usually regarded as a classification system that decides what category (e.g., entity type) a given input (e.g., word) is mapped to. Theoretically, given infinite data, a deep learning system is capable of representing any deterministic mapping for any given inputs and corresponding outputs [43]. However, due to the finite amount of data available in real-world applications, deep learning systems have to generalize beyond the training data.

Compared with human beings, deep learning systems lack the ability to learn abstractions from explicit and verbal definitions. Instead, they rely on the large amount of training examples to learn these rules. In the context of entity recognition, given the definition of a medical entity, humans can easily tell whether a word is a medical entity and what type the entity is. However, deep learning models have to learn this “definition” through large numbers of annotated examples. In a DNN, the final tagging result of a given input character in terms of medical recognition depends on many features, such as the POS information, the positional information, and the context words. The hidden layers in a DNN are considered as complex feature transformations in the networks and produce the most abstract features for the final output layer; this is a critical process in learning the implicit rules embedded in the training set.

The RNN is an extension of the traditional feedforward neural network, and can handle variable-length input sequences. An RNN contains a recurrent hidden state, and the activation of the hidden state depends on that of the previous time. Nevertheless, RNNs fail to capture long-term dependencies as the gradient tends to either vanish or explode during training.

The LSTM is a special kind of RNN that is designed to avoid the long-term dependency issue by joining with a gated memory cell [44]. Typically, an LSTM unit consists of an input gate i_t , an output gate o_t , a forget gate f_t , a memory cell c_t , and a hidden state h_t . The LSTM incorporates these structures called gates to optionally remove or add information; they contain a sigmoid neural net layer and a pointwise multiplication operation. The sigmoid layer outputs values between 0 and 1 to indicate how much of each component should be reserved, in which a value of 0 denotes “let nothing through” and 1 denotes “let everything through”. The LSTM computes the output by iterating the following equations:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i), \quad (3)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f), \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c), \quad (5)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o), \quad (6)$$

$$h_t = o_t \odot \tanh(c_t), \quad (7)$$

where σ means the sigmoid function; \odot denotes pointwise multiplication; W_i , W_f , W_c , and W_o (with subscripts x , h , and c) are the weight matrices for input x_t , hidden state h_t , memory cell c_t , and output o_t , respectively; and b_i , b_f , b_c , and b_o denote the bias vectors. The BiLSTM is composed of a forward LSTM and a backward LSTM to capture both past and future information, which are two separated networks with different parameters.

The entity extraction task can be modeled by deep learning methods as a sequence labeling task. In OHCs, there are many long sentences in patient-contributed content, and the semantic meaning of a focused character can be shaped by the characters before and after it over a long distance. In the text sequence of online consultations, users report their health conditions in detail and the mentions of each medical entity could rely on long-distance information in the text. Based on these intuitions,

we utilized BiLSTM to extract medical named entities, as BiLSTM can learn long-distance dependencies and the bidirectional information of a character at the same time.

2.4. CRF Layer

When it comes to the context of entity recognition in text, it is always beneficial to consider the correlations between the sequential labels as there are many tagging constraints in natural language sentences. However, the widely used Softmax method predicts the final labels independently, and using Softmax as the top inference layer to extract medical entities will probably break these constraints.

CRF is the most successful model that can control the structure prediction of tagging results. Therefore, CRF was employed to predict the final label sequence in the proposed model. CRF is a probabilistic framework and is usually adopted for sequential data including text [45]. The basic idea of CRF is to use a series of potential functions to estimate the conditional probability of the output label sequence given the input sequence. More specifically, CRF uses an undirected graphical model to calculate the conditional probability $p(y|x, w)$ of a label sequence y given an input sequence x , where w denotes the parameters in the model. $\Psi(x, y)$ denotes the feature vector and $Z(w, x)$ is the cumulative sum of $p(y|x, w)$ over all the possible y :

$$p(y|x, w) = \frac{\exp(w^T \Psi(x, y))}{Z(w, x)}. \quad (8)$$

The model is trained over a given training set $(Y, X) = \{x_i, y_i\}, i = 1 \dots N$, by maximizing the conditional likelihood:

$$w = \arg \max_w p(Y|X, w). \quad (9)$$

For the input sequence x and the trained parameters w , the final prediction of a trained CRF is the label sequence y^* that maximizes the model:

$$y^* = \arg \max_y p(y|x, w). \quad (10)$$

CRF predicts the optimal sequence of labels using a Viterbi algorithm for the input sequence. In our model, the final output of the entity recognition task imposes several hard constraints; for example, "I-cure" cannot follow "B-prob". The CRF layer considers the interactions between successive labels and can automatically learn these constraints from training data to ensure the validity of the final entity tagging results.

3. Evaluation

3.1. Datasets

The dataset used in the experiment was collected from the Good Doctor website (www.haodf.com). Established in 2006, the Good Doctor website is one of the largest online patient–doctor communication platforms. The platform enables patients to consult physicians about their health-related concerns by providing personal healthcare information in the manner of online posts, by telephone, and even by teleconference. Currently there are over 180,000 certificated doctors registered on the platform providing professional medical consultation services, attracting around 10,000 online health consultations from patients or their caregivers every day. In online health consultations, patients provide their basic health conditions and ask questions to physicians. A sample of an online health consultation on the Good Doctor website is provided in Figure 4. In the section "condition description" of medical consultations, patients describe their medical information such as symptoms, medical testing, treatment, medicine, cause of disease, and family medical history, which contains abundant health-related concepts. Therefore, we select the "condition description" sector of medical consultations as our entity tagging target.

h***
 Status: Before visiting
 2018-02-09
 Consultation title: Neuropathic tumor
 Disease: Tumor
 Length of illness: More than half a year
 Pregnancy: Not pregnant
 Allergy history: None (filled in on 2018-02-09)
 Condition description:
 In February 2017, I was diagnosed with bronchial cyst in Wuxi Private Hospital. In August 2017, I was re-examined in the same hospital. The doctor just said that I am fine and should prevent getting a cold, and did not do any treatment. In February 2018, I returned to Baokang for examination, and was diagnosed with neurogenic cysts by ct. There was no abnormality in blood tests.
 Questions:
 The county hospital said that surgery is required, and it is required to transfer to the central hospital. It is recommended to go after the year. I would like to ask when it will be appropriate next year.
 Hospital departments visited: Internal medicine, People's Hospital of Baokang County
 Medical history: None (filled in on 2018-02-09)
 Medical check information: Case data is only available for the doctor and the patient after logging in.

Figure 4. A sample of an online health consultation on the Good Doctor website.

We collected the “condition description” section of consultations that were posted on the Good Doctor platform from 1 January 2014 to 30 April 2017 using a crawler programmed in Python. After data preprocessing, we finally obtained around 8.6 million unlabeled medical consultations across all departments from the Good Doctor website for embedding training. After training using word2vec based on the collected consultation text corpus, we finally obtained 852,497 unique words and 10,336 unique characters in the word embedding table and the character embedding table, respectively. We collected another medical consultation dataset from the oncology departments of the Good Doctor website of consultations that were posted in May 2017 for manual annotation. Each selected consultation contains at least one medical-related entity and should be longer than 20 characters. The consultations in the oncology department were selected for evaluation in this study as cancer is one of the leading causes of morbidity and mortality worldwide [46]. After data annotation, we obtained 536 labeled medical consultations as our final annotated dataset. The Cohen’s kappa value for inter-annotator reliability is 0.96, which indicates a near-perfect agreement [47]. The statistics of the annotated dataset are shown in Table 1. The collected consultation datasets and trained word and character embeddings were deposited in Harvard’s Dataverse [48].

Table 1. Statistics of the annotated dataset.

| Statistics | Numbers |
|---|---------|
| Number of sentences | 536 |
| Average number of characters in each sentence | 163 |
| Number of mentioned problems | 3870 |
| Number of mentioned tests | 987 |
| Number of mentioned treatments | 1608 |

3.2. Metrics

In this study, precision (P), recall (R), and F-measure (F) were adopted as the performance evaluation metrics. More specifically, precision represents the portion of entities that are correctly recognized, while recall denotes the portion of correctly recognized entities among all correct entities; indicating the overall performance of precision and recall, the F-measure is calculated as the harmonic average of precision and recall. The values of precision, recall, and F-measure are all real values between 0 and 1, with higher values indicating better performance.

3.3. Baseline Models

To evaluate the performance of our proposed approach on medical entity recognition in Chinese OHCs, we assessed our model against the following baseline systems: a CRF-based model that uses words as tag units [49], a CRF model that uses characters as tag units [50], and a DNN-based model using Character–Word Mixed Embedding (CWME) [32]. The reasons for selecting those three works as the baseline models in this study are as follows. First, the CRF-based methods have been widely adopted in sequence labeling problems such as POS and word segmentation, and have achieved promising performance in entity extraction tasks [31]. Second, the CWME method is also based on DNN and focuses on the Chinese social media context; it has been reported to perform well in entity extraction. Third, the first baseline model uses words as the basic tagging unit [49], while the second baseline model uses characters as the basic tagging unit [50]. We chose these two baseline models to evaluate the impact of recognition granularity on the performance of entity extraction, which has not been fully investigated in the context of Chinese health-related social media.

3.4. Model Settings

A 10-fold cross-validation procedure was utilized to run our proposed model and the baseline methods. The annotated dataset was split into three parts: six folds for training, two for validation, and the remaining two for testing. We tested the CRF baseline methods using the same additional features that were incorporated into our proposed model to establish fair comparisons. In this study, we utilized an open source tool named CRF++ to construct the CRF baseline models as it is fast and customizable [51]. TensorFlow was utilized to construct DNN-based models [52], and the adapted codes were uploaded and made available in an open repository [53]. We selected Adam, an optimization algorithm that can be used to update network weights iteratively based on training data, to update the parameters [54]. To avoid overfitting, the hidden layer size was set to 150 [55]. The initial learning rate was set to 0.001 and the dropout rate was set to 0.1. During model training, the predefined character embeddings were fine-tuned based on the training data [35]. To achieve better results, the hyperparameters were tuned based on the performance for different combinations of hyperparameter values using the validation dataset.

4. Results

4.1. Evaluation Results

Based on the results from the 10-fold cross-validation, a general performance comparison of the proposed model with the baseline models across different medical entity types is presented in Table 2. As shown in Table 2, based on our experimental data, we observe that our proposed method attained considerably higher recall and a better F-measure in general, while attaining relatively lower precision for most entity types compared with the CRF-based baseline models. In general, our model obtained 7.36% improvement over the word-based CRF model, and 2.31% improvement over the character-based CRF model in terms of overall F-measure. In general, our model attained better overall performance in terms of precision, recall, and F-measure over the CWME approach. For the two CRF baseline models, the character-based method generated a substantially better overall performance over the word-based method.

From the perspective of different entity types, the DNN-based models achieved moderately lower performance in terms of precision compared with the CRF-based models across almost all entity types. In contrast, the DNN-based models substantially outperformed the CRF-based models in terms of recall for all three entity types. We notice that CNMER obtained a relatively higher recall than CWME for the entity types of medical problems and treatment. The character-based CRF model extensively outperformed the word-based CRF model in terms of recall for all entity types. For the F-measure, we observe that CNMER achieved better performance over the other three baseline models for the entity types of medical problems and treatment.

Table 2. Performance comparison of CNMER and baseline methods ¹.

| Models | Problem (%) | | | Test (%) | | | Treatment (%) | | | All (%) | | |
|--------|-------------|-------|-------|----------|-------|-------|---------------|-------|-------|---------|-------|-------|
| | P | R | F | P | R | F | P | R | F | P | R | F |
| CRF_W | 66.30 | 55.17 | 60.22 | 70.05 | 56.18 | 62.33 | 71.23 | 55.57 | 62.41 | 68.02 | 55.42 | 61.07 |
| CRF_C | 69.75 | 63.30 | 66.36 | 71.61 | 63.10 | 67.08 | 70.84 | 60.05 | 64.98 | 70.28 | 62.43 | 66.12 |
| CWME | 67.25 | 67.72 | 67.46 | 68.17 | 68.70 | 68.40 | 65.83 | 66.95 | 66.35 | 67.00 | 67.65 | 67.31 |
| CNMER | 67.46 | 69.80 | 68.55 | 67.78 | 68.18 | 67.95 | 69.62 | 67.44 | 68.47 | 67.97 | 68.96 | 68.43 |

¹ CRF_W, the word-based CRF baseline model; CRF_C, the character-based CRF baseline model; CWME, deep neural network (DNN) baseline model based on Character-Word Mixed Embedding; CNMER, the model proposed in this study.

For further explanation, a *t*-test was conducted on the 10 general results of the 10-fold cross-validation. Based on our experimental datasets, the evaluation results indicate that the recall of our model (mean = 68.96%) is significantly higher than those of CRF_W (mean = 55.42%) ($t = -24.485$, $p < 0.01$), CRF_C (mean = 62.43%) ($t = -12.297$, $p < 0.01$), and CWME (mean = 67.65%) ($t = -2.745$, $p < 0.05$). In terms of F-measure, our model (mean = 68.43%) statistically outperforms CRF_W (mean = 61.07%) ($t = -16.241$, $p < 0.01$), CRF_C (mean = 66.12%) ($t = -6.369$, $p < 0.01$), and CWME (mean = 67.31%) ($t = -2.983$, $p < 0.05$).

To further evaluate the contribution of the predefined character representation, we conducted a comparison of the proposed models over different character representations. As shown in Table 3, CNMER generally outperforms the models that use “Random” or “CW” as the character representation, where “CW” denotes the model that uses concatenated character and context word embedding as the character representation. The *t*-test results also indicate that CNMER substantially outperforms “Random” in terms of precision ($t = -6.265$, $p < 0.01$), recall ($t = -9.352$, $p < 0.01$), and F-measure ($t = -9.821$, $p < 0.01$). Meanwhile, CNMER outperforms “CW” in terms of precision ($t = -2.596$, $p < 0.05$) and F-measure ($t = -4.720$, $p < 0.01$).

Table 3. Performance comparison of the results across different DNN models and entity types ¹.

| Models | Problem (%) | | | Test (%) | | | Treatment (%) | | | All (%) | | |
|--------|-------------|-------|-------|----------|-------|-------|---------------|-------|-------|---------|-------|-------|
| | P | R | F | P | R | F | P | R | F | P | R | F |
| Random | 62.92 | 65.06 | 63.94 | 65.77 | 65.48 | 65.54 | 63.41 | 61.33 | 62.33 | 63.42 | 64.16 | 63.76 |
| CW | 66.61 | 70.45 | 68.46 | 65.47 | 69.42 | 67.37 | 67.24 | 66.70 | 66.93 | 66.53 | 69.34 | 67.90 |
| CNMER | 67.46 | 69.80 | 68.55 | 67.78 | 68.18 | 67.95 | 69.62 | 67.44 | 68.47 | 67.97 | 68.96 | 68.43 |

¹ Random, the model that uses random embedding as character representation; CW, the model that only uses concatenated character and context word embedding as the character representation; CNMER, the model proposed in this study.

4.2. Extracted Medical Entities

Our proposed model tags each single character instead of segmented words to recognize medical entities, and the representation of each character is designed according to the uniqueness of Chinese health-related social media. Using CNMER, we can effectively extract medical-related entities including medical problems, tests, and treatment from the informal text of Chinese social media. Table 4 presents some examples of the extracted medical entities from the Good Doctor website. Some of those entities are colloquial such as “Face is itching” and “There is a malignant tumor in the left lung”. There are even extracted entities that are misspellings in Chinese such as “Hilar mediastinal lymph node metastasis”. However, these entities are rarely recognized and classified correctly by existing related models. The informal medical entities are quite common in Chinese social media, yet they are rare in medical dictionaries and it is tough to manually capture their unique features, making it challenging for most existing models.

Table 4. Examples of extracted medical entities from online health consultations.

| Entity Types | Examples |
|--------------|--|
| Problem | Face is itching |
| | Hilar mediastinal lymph node metastasis |
| | The size of tumor is about two centimeters |
| | There is a malignant tumor in the left lung Central poorly differentiated lung adenocarcinoma |
| Test | Liver puncture |
| | Enhanced CT scan |
| | X-ray examination |
| | DNA genetic testing Enhanced nuclear magnetic resonance images of brain |
| Treatment | Erlotinib |
| | Surgical removal of lesions |
| | Six-cycle course of chemotherapy |
| | Traditional Chinese medicine for blood circulation Radiation therapy for brain and spinal cord tumors |

5. Discussion

In spite of the comparatively weak performance in terms of precision, the experimental results reveal the substantial ability of our proposed model over existing approaches in medical entity extraction, indicating the advantage of our designed model. The further character representation evaluation implies that using pretrained embeddings based on the domain corpus can dramatically improve the performance of medical entity recognition over randomly initialized ones, and incorporating position and POS features can further improve the overall performance. The evaluation also suggests the advantage of character-based methods over word-based methods in the Chinese social media context. By the inclusion of context word embedding with character embedding as the representation of the text input, our model can effectively extract medical-related entities in Chinese OHCs without complex feature engineering.

The Bidirectional LSTM architecture is capable of learning long-term dependencies from both forward and backward directions to capture further context features, which results in better performance of the DNN model over traditional machine learning models in terms of overall recall and F-measure. The inclusion of context word embedding with the character embedding can partly capture the context information and avoid use of the same character embedding vector in varied contexts. Therefore, a character with the same location tag might be assigned a different representation vector, while a character in a different context would be represented with the same embedding in CWME; this could be the reason why CNMER generally outperforms the CWME approach. The overall higher recall over traditional machine learning approaches and better F-measure over the three baseline models demonstrates that our model is more appropriate for medical entity extraction in online medical consultations.

In the setting of online consultations, physicians need to process abundant unstructured text information, among which medical entities are the most critical part for efficient health assistance. The rapid development of information and communication technology in recent years has greatly changed the manner of health service delivery in modern society. Compared with real-world face-to-face visits, e-mediated patient–doctor communication has certain unique characteristics that touch the critical components of the relationship between patients and physicians [56] and would potentially affect the sustainability and effectiveness of online communities.

Confirmation bias means that one is more inclined to the evidence that supports their existing beliefs, expectations, and hypothesis in hand [57]. In online health consultations, users are anonymous to physicians and the user-generated content is the only cue for physicians to infer the health conditions of patients. With the limited information available in online consultations, physicians need to evaluate the health conditions of patients and make professional medical suggestions. However, adequate

and accurate evidence are required in medical services. During this process, confirmation bias could occur in two possible ways. First, before reporting their conditions to healthcare professionals online, some patients may have their own prior judgement for the medical problem and thus unconsciously describe their conditions with bias. Second, due to the limited information available, even the most seasoned healthcare practitioner can be prejudiced occasionally and be led to misdiagnose a problem by confirmation bias [58]. CNMER has been proven effective in extracting health-related concepts from Chinese OHCs, and these medical concepts are essential components for health professionals to provide feedback. In the context of online medical consultations, the principal contents submitted by users are highlighted with the extracted medical entities. Users can check and edit what they have written, and physicians can effectively examine the posts without ignoring the critical information in the text, which could potentially help to alleviate the effect of confirmation bias.

Trust is another critical concern in e-mediated patient–doctor interaction. The first and foremost function of trust is to reduce complexity [59]. Trust has been shown to affect a host of behaviors including patients' willingness to seek care, reveal sensitive information, and remain with a physician [60]. In the context of e-mediated communication, patients are anonymous to healthcare service providers, which further highlights the importance of trust. Patients' trust in their doctor and doctors' trust in their patients during online consultations play an essential role in dealing with the health issues of patients. For patients that seek online medical support, trust in their doctors can help to sustain well-being when coping with health risks. The continuous trust between patients and physicians in online health consultations is one of the key critical elements that ensure the sustainability of online healthcare service delivery. The higher level and status of a healthcare system have proven to be associated with more trust [61]. The extracted medical concepts facilitate efficient information processing and boost the information exchange in online consultations, improving the relationship between patient and doctor. Patient–doctor communication is more than transferring information about medical conditions from patient to doctor and medical knowledge from doctor to patient: it is about releasing the patient's feelings of stress, anxiety, and risk in health issues [56]. The significant positive relationship between trust and the perceived value of social interaction has been reported in a previous study [62]. An efficient, intelligent healthcare system employing an online platform can thus improve the trust between patients and doctors as the social exchange is perceived to be beneficial.

Despite the wide use of health insurance and other related programs, economic or time cost is usually inevitable for most healthcare consumers when dealing with their health problems. Healthcare consumers tend to reduce the cost without impairing the quality of care; they evaluate the return and corresponding cost of different healthcare services and make decisions according to their related knowledge and experience. Chronic diseases such as diabetes, cancer, cardiovascular disease, and chronic respiratory diseases have been a substantial economic burden for patients due to expenditure on long-term medical care, especially for those from low-income and middle-income countries such as China [63]. The introduction of online healthcare services provides users with alternative options to cope with these health concerns. OHCs have been reported as powerful platforms for chronic disease patients to tackle some of the challenges, with certain advantages including the exchange of medical knowledge, supporting self-management, and improving patient-centered care [64]. Online healthcare platforms not only provide modern patients an open communication channel with their physicians, but also facilitate patients gaining control over their lives and improving the quality of care by self-management [64]. While the sensitivity to cost of healthcare services varies [65], individuals can seek medical support with minimal time and cost restrictions in OHCs.

Health information technology has been widely adopted in recent years due to its capacity to improve the cost savings, efficiency, quality, and safety of medical service delivery. Among all the components, cost remains the primary barrier that impedes the adoption of health information technology [66], and cost–benefit analysis of healthcare system adoption is meaningful. For the proposed health system designed for OHCs, online platforms can employ the system on their websites, and both patients and doctors can utilize the system to enhance the healthcare service delivery.

As stated before, the engagement of the intelligent system can benefit the OHCs by attracting more doctors and patients to participate in healthcare information exchanges due to its advantages including diminishing confirmation bias, gaining trust, and reducing cost. Despite the potentially high cost of DNN systems at the moment, the rapid development of deep learning technologies and the booming of related web services make the system more applicable. It is economically feasible for online healthcare platforms to employ the system as further considerable benefits are expected.

The utilization of DNN in our model achieved more promising performance over conventional machine learning methods. From the perspective of practical implementation, DNN systems are known for their lack of transparency, and the prediction results are tough to explain. Consequently, there are concerns regarding the safety issue of employing the system as DNN-based models remain mysterious to their users. However, as our system are designed for extracting medical entities to facilitate information processing rather than providing health-related professional suggestions, the transparency of our model and the explanation of results are actually dispensable in real-world applications.

The sustainable employment of the proposed DNN-based healthcare system by online health platforms relies on the continuous benefits that are obtained from the system. The powerful capacity to extract medical concepts from the healthcare system can moderately improve the quality of information transmission between patient and doctors in OHCs, which reduces economic and time cost and enhances quality of life [67,68]. According to the medical advice provided by health professionals, users can cope with their health issues more properly and thus decrease their medical expenditures. The effective health-related information seeking powered by the proposed model can also minimize patients' future health risks by reducing medical uncertainty. The sustainable development of OHCs depends on the participation of health professionals, and doctors can gain social and economic returns by participating in OHCs [69]. For healthcare service providers, the system can assist them to improve the efficiency of medical information processing with higher accuracy. As increasing numbers of participants engage with and benefit from the system, OHCs can gain more profit and thus invest more in the development of intelligent healthcare systems, which in turn attracts more participants to the platforms. Medical concept discovery is the basis of healthcare knowledge discovery strategies such as disease surveillance and adverse drug reaction detection. Healthcare knowledge discovery from social media has been validated as viable in previous works [15,16], and can contribute to the sustainability of public health. Therefore, the adoption of the proposed system can directly or indirectly benefit various participants including health consumers, health service providers, and online healthcare platforms, contributing to the sustainability of the virtualized healthcare industry.

6. Conclusions

Our study contributes to the literature mainly in terms of the following points. Firstly, this work designs an effective DNN model that can automatically learn context features of text to replace complex and time-consuming handcrafted feature engineering work. The evaluation results demonstrate that the proposed model considerably outperforms traditional machine learning approaches and a strong DNN baseline model. Second, this paper investigates the medical entity extraction task in the context of Chinese social media, while prior research primarily focused on the English language context. Considering the uniqueness of health-related Chinese social media text, this study proposes concatenating character embedding with context word embedding, together with position and a POS features vector, to enhance the feature representation of characters in Chinese online medical text. As far as we know, this research is among the first works to focus on medical-related entity recognition in Chinese social media. Third, based on a large domain text corpus collected from a well-known Chinese OHC, this work builds a word embedding dataset and a character embedding dataset in the context of Chinese medical-related social media, which are available to the public online [48]. The learned distributed representations of words and characters capture both syntactic and semantic features, and can facilitate learning algorithms to achieve more promising performance in many NLP-related tasks, including sentiment analysis [70], text classification [71], and recommendation [72].

Previous studies have certain limitations when applied to the context of Chinese health-related social media. This study designed a BiLSTM-CRF-based model named CNMER to extract medical-related entities from Chinese OHCs. The model utilizes character embedding, word embedding, position, and POS feature vectors as the character representation and avoids tough feature engineering work. Despite the relatively unsatisfying results in terms of precision compared with the CRF-based methods, the proposed CNMER approach attained statistically better performance in terms of recall and F-measure over all three baseline models including a strong DNN model, which indicates that our model is more effective in extracting health-related entities from Chinese OHCs. The advantages of using characters as the basic tag units are also validated in this study. The proposed medical entity extraction system contributes to the sustainable development of virtualized healthcare as it benefits many stakeholders including health consumers, health service providers, and online healthcare platforms.

Besides the above achievements, the designed model has certain limitations. First, we only considered the recognition of three main types of medical-related concepts; other entity types such as body part, medical department, and time which are also essential for medical decision support were not investigated in this study. Second, only the focal character and word were considered when constructing a representation vector, while further context characters and words could also contribute to additional performance improvement; this was not explored in our study. Lastly, although the evaluation results indicate that our model outperforms the baseline approaches, the performance is still not satisfying enough for real-world applications. Medical entity extraction in Chinese social media remains a challenging task and deserves further investigation.

Author Contributions: Conceptualization, H.Y.; Data curation, H.Y.; Funding acquisition, H.G.; Methodology, H.Y.; Supervision, H.G.; Validation, H.Y.; Writing—original draft, H.Y.; Writing—review & editing, H.G.

Funding: This research was funded by National Key Research & Development Plan of China (grant number: 2017YFB1400101), National Natural Science Foundation of China (grant number: 71572013).

Acknowledgments: The manuscript was approved by all authors for publication. We would like to thank all the anonymous reviewers for their valuable comments and suggestions which improved this paper. We would also thank the Editors and the Editorial Office for their professional work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yan, L.; Tan, Y. The Consensus Effect in Online Health-Care Communities. *J. Manag. Inf. Syst.* **2017**, *34*, 11–39. [[CrossRef](#)]
2. Jung, Y.; Hur, C.; Kim, M. Sustainable Situation-Aware Recommendation Services with Collective Intelligence. *Sustainability* **2018**, *10*, 1632. [[CrossRef](#)]
3. Wang, X.; Zhao, K.; Street, N. Analyzing and predicting user participations in online health communities: A social support perspective. *J. Med. Internet Res.* **2017**, *19*, e130. [[CrossRef](#)] [[PubMed](#)]
4. Kazmer, M.M.; Lustria, M.L.A.; Cortese, J.; Burnett, G.; Kim, J.H.; Ma, J.; Frost, J. Distributed knowledge in an online patient support community: Authority and discovery. *J. Assoc. Inf. Sci. Technol.* **2014**, *65*, 1319–1334. [[CrossRef](#)]
5. HINTS. HINTS-FDA Survey Instrument. Available online: http://hints.cancer.gov/question-details.aspx?PK_Cycle=8&qid=757 (accessed on 13 March 2018).
6. CNNIC. 39th Statistical Report on Internet Development in China. Available online: <http://www.cnnic.cn/hlwfzyj/hlwxzbg/hlwjbg/201701/P020170123364672657408.pdf> (accessed on 13 March 2018).
7. Jung, C.; Padman, R. Virtualized healthcare delivery: Understanding users and their usage patterns of online medical consultations. *Int. J. Med. Inf.* **2014**, *83*, 901–914. [[CrossRef](#)] [[PubMed](#)]
8. Li, M.; Mao, J. Hedonic or utilitarian? Exploring the impact of communication style alignment on user's perception of virtual health advisory services. *Int. J. Inf. Manag.* **2015**, *35*, 229–243. [[CrossRef](#)]
9. Yan, Z.; Wang, T.; Chen, Y.; Zhang, H. Knowledge sharing in online health communities: A social exchange theory perspective. *Inf. Manag.* **2016**, *53*, 643–653. [[CrossRef](#)]
10. Barrett, M.; Oborn, E.; Orlikowski, W. Creating value in online communities: The sociomaterial configuring of strategy, platform, and stakeholder engagement. *Inf. Syst. Res.* **2016**, *27*, 704–723. [[CrossRef](#)]

11. Vuong, Q.-H. Survey data on Vietnamese propensity to attend periodic general health examinations. *Sci. Data* **2017**, *4*, 170142. [[CrossRef](#)] [[PubMed](#)]
12. Remondino, M. Information Technology in Healthcare: HHC-MOTES, a Novel Set of Metrics to Analyse IT Sustainability in Different Areas. *Sustainability* **2018**, *10*, 2721. [[CrossRef](#)]
13. Lu, H.-Y.; Shaw, B.R.; Gustafson, D.H. Online health consultation: Examining uses of an interactive cancer communication tool by low-income women with breast cancer. *Int. J. Med. Inf.* **2011**, *80*, 518–528. [[CrossRef](#)] [[PubMed](#)]
14. Lei, J.; Tang, B.; Lu, X.; Gao, K.; Jiang, M.; Xu, H. A comprehensive study of named entity recognition in Chinese clinical text. *J. Am. Med. Inform. Assoc.* **2014**, *21*, 808–814. [[CrossRef](#)] [[PubMed](#)]
15. Kagashe, I.; Yan, Z.; Suheryani, I. Enhancing Seasonal Influenza Surveillance: Topic Analysis of Widely Used Medicinal Drugs Using Twitter Data. *J. Med. Internet Res.* **2017**, *19*, e315. [[CrossRef](#)] [[PubMed](#)]
16. Nikfarjam, A.; Sarker, A.; O'Connor, K.; Ginn, R.; Gonzalez, G. Pharmacovigilance from social media: Mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J. Am. Med. Inform. Assoc.* **2015**, *22*, 671–681. [[CrossRef](#)] [[PubMed](#)]
17. Ginsberg, J.; Mohebbi, M.H.; Patel, R.S.; Brammer, L.; Smolinski, M.S.; Brilliant, L. Detecting influenza epidemics using search engine query data. *Nature* **2009**, *457*, 1012. [[CrossRef](#)] [[PubMed](#)]
18. Pirmohamed, M.; James, S.; Meakin, S.; Green, C.; Scott, A.K.; Walley, T.J.; Farrar, K.; Park, B.K.; Breckenridge, A.M. Adverse drug reactions as cause of admission to hospital: Prospective analysis of 18 820 patients. *BMJ* **2004**, *329*, 15–19. [[CrossRef](#)] [[PubMed](#)]
19. Liu, Y.; Cheng, Y.; Yan, Z.; Ye, X. Multilevel Analysis of International Scientific Collaboration Network in the Influenza Virus Vaccine Field: 2006–2013. *Sustainability* **2018**, *10*, 1232. [[CrossRef](#)]
20. Uzuner, Ö.; South, B.R.; Shen, S.; Duvall, S.L. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 552–556. [[CrossRef](#)] [[PubMed](#)]
21. Gupta, S.; MacLean, D.L.; Heer, J.; Manning, C.D. Induced lexico-syntactic patterns improve information extraction from online medical forums. *J. Am. Med. Inform. Assoc.* **2014**, *21*, 902–909. [[CrossRef](#)] [[PubMed](#)]
22. Song, M.; Yu, H.; Han, W.-S. Developing a hybrid dictionary-based bio-entity recognition technique. *BMC Med. Inform. Decis. Mak.* **2015**, *15*, S9. [[CrossRef](#)] [[PubMed](#)]
23. Liu, J.; Zhao, S.; Zhang, X. An ensemble method for extracting adverse drug events from social media. *Artif. Intell. Med.* **2016**, *70*, 62–76. [[CrossRef](#)] [[PubMed](#)]
24. Coden, A.; Savova, G.; Sominsky, I.; Tanenblatt, M.; Masanz, J.; Schuler, K.; Cooper, J.; Guan, W.; de Groen, P.C. Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model. *J. Biomed. Inform.* **2009**, *42*, 937–949. [[CrossRef](#)] [[PubMed](#)]
25. Sanz, X.; Pareja, L.; Rius, A.; Rodenas, P.; Abdon, N.; Galvez, J.; Esteban, L.; Escriba, J.M.; Borrás, J.M.; Ribes, J. Definition of a SNOMED CT pathology subset and microglossary, based on 1.17 million biological samples from the Catalan Pathology Registry. *J. Biomed. Inform.* **2018**, *78*, 167–176. [[CrossRef](#)] [[PubMed](#)]
26. Saha, S.K.; Sarkar, S.; Mitra, P. Feature selection techniques for maximum entropy based biomedical named entity recognition. *J. Biomed. Inform.* **2009**, *42*, 905–911. [[CrossRef](#)] [[PubMed](#)]
27. Jiang, M.; Chen, Y.; Liu, M.; Rosenbloom, S.T.; Mani, S.; Denny, J.C.; Xu, H. A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 601–606. [[CrossRef](#)] [[PubMed](#)]
28. Sampathkumar, H.; Chen, X.-W.; Luo, B. Mining adverse drug reactions from online healthcare forums using hidden Markov model. *BMC Med. Inform. Decis. Mak.* **2014**, *14*, 91. [[CrossRef](#)] [[PubMed](#)]
29. Sun, C.; Yi, G.; Wang, X.; Lin, L. Rich features based Conditional Random Fields for biological named entities recognition. *Comput. Biol. Med.* **2007**, *37*, 1327–1333. [[CrossRef](#)] [[PubMed](#)]
30. Kovačević, A.; Dehghan, A.; Filannino, M.; Keane, J.A.; Nenadic, G. Combining rules and machine learning for extraction of temporal expressions and events from clinical narratives. *J. Am. Med. Inform. Assoc.* **2013**, *20*, 859–866. [[CrossRef](#)] [[PubMed](#)]
31. Xie, J.; Liu, X.; Zeng, D.D. Mining e-cigarette adverse events in social media using Bi-LSTM recurrent neural network with word embedding representation. *J. Am. Med. Inform. Assoc.* **2017**, *25*, 72–80. [[CrossRef](#)] [[PubMed](#)]
32. Xiang, Y. Chinese Named Entity Recognition with Character-Word Mixed Embedding. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 2055–2058.

33. Peng, N.; Dredze, M. Named entity recognition for chinese social media with jointly trained embeddings. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 548–554.
34. Unanue, I.J.; Borzeshi, E.Z.; Piccardi, M. Recurrent neural networks with specialized word embeddings for health-domain named-entity recognition. *J. Biomed. Inform.* **2017**, *76*, 102–109. [[CrossRef](#)] [[PubMed](#)]
35. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural Architectures for Named Entity Recognition. In Proceedings of the NAACL-HLT, San Diego, CA, USA, 12–17 June 2016; pp. 260–270.
36. Xu, Y.; Wang, Y.; Liu, T.; Liu, J.; Fan, Y.; Qian, Y.; Tsujii, J.; Chang, E.I. Joint segmentation and named entity recognition using dual decomposition in Chinese discharge summaries. *J. Am. Med. Inform. Assoc.* **2013**, *21*, e84–e92. [[CrossRef](#)] [[PubMed](#)]
37. Li, D.; Hu, T.; Li, J.; Qian, Q.; Zhu, W. Construction and Application of the Chinese Unified Medical Language System. *J. Intell.* **2011**, *30*, 147–151.
38. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [[CrossRef](#)] [[PubMed](#)]
39. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.
40. Duan, H.; Sui, Z.; Tian, Y.; Li, W. The cips-sighan clp 2012 chinese word segmentation on microblog corpora bakeoff. In Proceedings of the Second CIPS-SIGHAN Joint Conference on Chinese Language Processing, Tianjin, China, 20–21 December 2012; pp. 35–40.
41. Klein, D.; Smarr, J.; Nguyen, H.; Manning, C.D. Named entity recognition with character-level models. In Proceedings of the CoNLL-2003, Edmonton, AB, Canada, 31 May–1 June 2003; pp. 180–183.
42. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
43. Marcus, G. Deep learning: A critical appraisal. *arXiv*, 2018; arXiv:1801.00631.
44. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
45. Lafferty, J.; McCallum, A.; Pereira, F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning, Williamstown, MA, USA, 28 June–1 July 2001; pp. 282–289.
46. Ferlay, J.; Soerjomataram, I.; Dikshit, R.; Eser, S.; Mathers, C.; Rebelo, M.; Parkin, D.M.; Forman, D.; Bray, F. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **2014**, *5*, E359–E386.
47. Blackman, N.J.M.; Koval, J.J. Interval estimation for Cohen’s kappa as a measure of agreement. *Stat. Med.* **2000**, *19*, 723–741. [[CrossRef](#)]
48. Yang, H. Replication Data for: Toward Sustainable Virtualized Healthcare: Extracting Medical Entities in Chinese Online Health Consultations with Deep Neural Networks. Available online: <https://doi.org/10.7910/DVN/4GBJIU> (accessed on 1 September 2018).
49. Mao, X.; Dong, Y.; He, S.; Bao, S.; Wang, H. Chinese word segmentation and named entity recognition based on conditional random fields. In Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing, Hyderabad, India, 11–12 January 2008.
50. Song, S.; Zhang, N.; Huang, H. Named entity recognition based on conditional random fields. *Clust. Comput.* **2017**, 1–12. [[CrossRef](#)]
51. Kudo, T. CRF++: Yet Another CRF Toolkit. Available online: <http://crfpp.sourceforge.net/> (accessed on 13 March 2018).
52. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M. TensorFlow: A System for Large-Scale Machine Learning. In Proceedings of the 2016 OSDI, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
53. Yang, H. CNMER: A Model for Chinese Medical Named Entity Extraction. Github, 2018. Available online: <https://github.com/yhzbbit/CNMER> (accessed on 20 August 2018).
54. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.

55. Ling, W.; Dyer, C.; Black, A.W.; Trancoso, I.; Fernandez, R.; Amir, S.; Marujo, L.; Luis, T. Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 17–21 September 2015; pp. 1520–1530.
56. Andreassen, H.K.; Trondsen, M.; Kummervold, P.E.; Gammon, D.; Hjortdahl, P. Patients who use e-mediated communication with their doctor: New constructions of trust in the patient-doctor relationship. *Qual. Health Res.* **2006**, *16*, 238–248. [[CrossRef](#)] [[PubMed](#)]
57. Nickerson, R.S. Confirmation bias: A ubiquitous phenomenon in many guises. *Rev. Gen. Psychol.* **1998**, *2*, 175. [[CrossRef](#)]
58. Patel, V.L.; Kaufman, D.R.; Arocha, J.F. Emerging paradigms of cognition in medical decision-making. *J. Biomed. Inform.* **2002**, *35*, 52–75. [[CrossRef](#)]
59. Luhmann, N. *Trust and Power Chichester*; John Wiley and Sons, Inc.: Chichester, UK, 1979.
60. Hall, M.A.; Dugan, E.; Zheng, B.; Mishra, A.K. Trust in physicians and medical institutions: What is it, can it be measured, and does it matter? *Milbank Q.* **2001**, *79*, 613–639. [[CrossRef](#)] [[PubMed](#)]
61. Johansson, E.; Winkvist, A. Trust and transparency in human encounters in tuberculosis control: Lessons learned from Vietnam. *Qual. Health Res.* **2002**, *12*, 473–491. [[CrossRef](#)] [[PubMed](#)]
62. Singh, J.; Sirdeshmukh, D. Agency and trust mechanisms in consumer satisfaction and loyalty judgments. *J. Acad. Mark. Sci.* **2000**, *28*, 150–167. [[CrossRef](#)]
63. Abegunde, D.O.; Mathers, C.D.; Adam, T.; Ortegón, M.; Strong, K. The burden and costs of chronic diseases in low-income and middle-income countries. *Lancet* **2007**, *370*, 1929–1938. [[CrossRef](#)]
64. van der Eijk, M.; Faber, M.J.; Aarts, J.W.; Kremer, J.A.; Munneke, M.; Bloem, B.R. Using online health communities to deliver patient-centered care to people with chronic conditions. *J. Med. Internet Res.* **2013**, *15*, e115. [[CrossRef](#)] [[PubMed](#)]
65. Vuong, Q.-H.; Ho, T.-M.; Nguyen, H.-K.; Vuong, T.-T. Healthcare consumers' sensitivity to costs: A reflection on behavioural economics from an emerging market. *Palgrave Commun.* **2018**, *4*, 70. [[CrossRef](#)]
66. Goldzweig, C.L.; Towfigh, A.; Maglione, M.; Shekelle, P.G. Costs and benefits of health information technology: New trends from the literature. *Health Aff. (Millwood)* **2009**, *28*, w282–w293. [[CrossRef](#)] [[PubMed](#)]
67. Lorig, K.R.; Holman, H.R. Self-management education: History, definition, outcomes, and mechanisms. *Ann. Behav. Med.* **2003**, *26*, 1–7. [[CrossRef](#)] [[PubMed](#)]
68. Newman, S.; Steed, L.; Mulligan, K. Self-management interventions for chronic illness. *Lancet* **2004**, *364*, 1523–1537. [[CrossRef](#)]
69. Guo, S.; Guo, X.; Fang, Y.; Vogel, D. How doctors gain social and economic returns in online health-care communities: A professional capital perspective. *J. Manag. Inf. Syst.* **2017**, *34*, 487–519. [[CrossRef](#)]
70. Peng, H.; Ma, Y.; Li, Y.; Cambria, E. Learning multi-grained aspect target sequence for Chinese sentiment analysis. *Knowl.-Based Syst.* **2018**, *148*, 167–176. [[CrossRef](#)]
71. Zhu, G.; Iglesias, C.A. Exploiting semantic similarity for named entity disambiguation in knowledge graphs. *Expert Syst. Appl.* **2018**, *101*, 8–24. [[CrossRef](#)]
72. Pourgholamali, F.; Kahani, M.; Bagheri, E.; Noorian, Z. Embedding unstructured side information in product recommendation. *Electron. Commer. Res. Appl.* **2017**, *25*, 70–85. [[CrossRef](#)]

