Introducing Information Extraction to Radiology Information Systems to Improve the Efficiency on Reading Reports

Zhe Xie[1,2] Yuanyuan Yang[1] Mingqing Wang[1] Ming Li[3] Haozhe Huang[4] Dezhong Zheng[1,2] Rong Shu[1,2] Tonghui Ling[1,2]

[1] Laboratory for Medical Imaging Informatics, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, People's Republic of China
[2] Shanghai Institute of Technical Physics, University of Chinese Academy of Sciences, Beijing, People's Republic of China
[3] Department of Radiology, Fudan University Huadong Hospital, Shanghai, People's Republic of China
[4] Department of Interventional Radiology, Fudan University Shanghai Cancer Center, Shanghai, People's Republic of China

Address for correspondence Tonghui Ling, MS, Laboratory for Medical Imaging Informatics, Shanghai Institute of Technical Physics of Chinese Academy of Sciences, 500 Yu Tian Road, Shanghai 200083, People's Republic of China (e-mail: lingtonghui@mail.sitp.ac.cn).

Methods Inf Med 2019;58:94–106.

Abstract

Background Radiology reports are a permanent record of patient's health information often used in clinical practice and research. Reading radiology reports is common for clinicians and radiologists. However, it is laborious and time-consuming when the amount of reports to be read is large. Assisting clinicians to locate and assimilate the key information of reports is of great significance for improving the efficiency of reading reports. There are few studies on information extraction from Chinese medical texts and its application in radiology information systems (RIS) for efficiency improvement.

Objectives The purpose of this study was to explore methods for extracting, grouping, ranking, delivering, and displaying medical-named entities in radiology reports which can yield efficiency improvement in RISs.

Methods A total of 5,000 reports were obtained from two medical institutions for this study. We proposed a neural network model called Multi-Embedding-BGRU-CRF (bidirectional gated recurrent unit-conditional random field) for medical-named entity recognition and rule-based methods for entity grouping and ranking. Furthermore, a methodology for delivering and displaying entities in RISs was presented.

Results The proposed neural named entity recognition model has achieved a good *F*1 score of 95.88%. Entity ranking achieved a very high accuracy of 99.23%. The weakness of the system is the entity grouping approach which yield accuracy of 91.03%. The effectiveness of the overall solution was proved by an evaluation task performed by two clinicians based on the setup of actual clinical practice.

Conclusions The neural model shows great potential in extracting medical-named entities from radiology reports, especially for languages, that lack lexicons and natural language processing tools. The pipeline of extracting, grouping, ranking, delivering, and displaying medical-named entities could be a feasible solution to enhance RIS functionality by information extraction. The integration of information extraction and RIS has been demonstrated to be effective in improving the efficiency of reading radiology reports.

Keywords
- ► radiology reports
- ► radiology information systems
- ► named-entity recognition
- ► information extraction
- ► neural networks

received
January 3, 2019
accepted after revision
June 3, 2019

DOI https://doi.org/10.1055/s-0039-1694992.
ISSN 0026-1270.

Introduction

Radiology examination, as a noninvasive method to look inside the body, plays a vital role in patient health care including disease prevention, diagnosis, and treatment. A radiology report written by radiologist documents the crucial findings of the examination and the analyses of the findings. Radiology report is the most important means of communication between the radiologist and referring clinician.[1] According to the investigation,[2] about 87% of the referring clinicians thought the radiology report is an indispensable tool for medical practice. Radiology report as a permanent record of patient's health information can be used not only for clinical function, but also for billing, accreditation, quality improvement, research, and teaching.[3] Reading radiology reports is common for clinicians and radiologists whether in clinical routine or retrospective research. However, it is laborious and time-consuming when the amount of reports to be read is large. Although radiology report usually produced in a certain format and structure, due to the complexity of human disease and the diversity of natural language expression, it takes time to get the crucial information in the report and there is a possibility of missing and misreading when time is limited. Therefore, it is of great significance to assist medical doctors to assimilate the information scattered in radiology reports efficiently and precisely. Taking the hospitals that we are familiar with in China, for example, when a patient comes to hospital with some complaint or for a follow-up visit, clinicians should read the patient's historical records to know the clinical history, make a comparison, and get a full picture of the patient. It is a trade-off to speed up the report reading and ensure the quality of medical care. Therefore, a method of reading radiology reports with better efficiency supported by RIS can make patients get higher medical care in limited time and relieve the pressure in hospital service.

Crucial information is randomly distributed in radiology report, and it can be in any part of the report. Clinician should thoroughly read the radiology report to locate and acquire the crucial information. It could be useful if RISs can extract and highlight the medical terms with crucial information in the report before clinicians read it. Furthermore, crucial information is also unevenly distributed among radiology reports. Statistically, we analyzed 3,000 radiology reports randomly selected from hospital A and hospital B in the timespan of December 1, 2014 to May 1, 2017 to show the distribution of the crucial information among reports which is shown in ►Appendix Fig. A1. We defined five categories of medical-named entity to represent the key information of the radiology report which was detailed in section of "Data Preparation". As we can see, the distribution of medical entities among radiology reports is like a Gaussian distribution with mean of 9 except the value at 0 which represents a report with no abnormal findings. However, in the current RIS, for a list of radiology reports, clinicians need to click to open every report to thoroughly read it, whether there is clinically important information in it. With a tool that can extract and visually emphasize the key information of the report, clinicians can selectively spend more time on the parts contain more information and ignore the trivial parts, improving clinicians' efficiency of reading radiology reports.

Structured reports with standardized terminologies and templates would be a feasible solution for computer to recognize and highlight the crucial information. However, the structured report has not yet been universally accepted by radiologists.[4] Currently, many of the radiology reports are still written in narrative manner. Therefore, information extraction technology is needed to extract key information from the narrative radiology report. Information extraction is a task of natural language processing (NLP) for automatically extracting structured information includes entities, relations, and events from unstructured text. Free-text format is the biggest obstacle to extract key information from radiology reports due to the ambiguity and variability of the natural language, the complexity of the human disease, and the different expressions of different radiologists and medical institutions. However, most of the radiology reports follow a common pattern of expression, usually including the period of treatment, anatomical structure, clinical observation, diseases, and their modifiers. These common components capture the semantics and key information of the report. Many works have been done in extracting clinically significant information from medical texts. The methods can be roughly divided into three categories: lexicon-based, rule-based, and machine-learning approaches. According to the recent survey, more than half of the clinical information extraction systems use lexicon-based and rule-based methods.[5] Metageneralis introduce a lexicon-based method to extract and normalize lexicon units from magnetic resonance imaging (MRI) reports.[6] Lakhani et al uses a rule-based method to detect critical results from radiology reports.[7] Over the past decade, researchers have paid more attention to machine-learning technique. Hassanpour and Langlotz employed conditional Markov models and conditional random fields (CRF) from Stanford Named-Entity Recognizer toolkit for named-entity recognition (NER) in radiology reports and showed that the machine-learning approach have more generalizability than the dictionary-based approach.[8] Another research study involved using an ensemble CRF-based method which outperformed the vanilla CRF model for information extraction from mammography reports.[9] Some people use Support Vector Machine (SVM) to recognize named entity in Chinese clinical text.[10,11] Though they have been successfully used, traditional machine-learning methods are relying heavily on handcrafted features. Moreover, there are few research has been done on information extraction of Chinese medical texts according to the survey.[12] For the Chinese medical texts, due to the lack of medical lexicon such as UMLS,[13] MeSH,[14] SNOMED CT,[15] RadLex,[16] and clinical information extraction tools such as cTAKES,[17] MetaMap,[18] MedLEE,[19] developing a lexicon-based and rule-based clinical information extraction system almost starts from scratch. With the boom of deep learning, neural models show great potential in clinical information extraction without using local dictionaries and handcrafted rules. Lyu et al introduces a variation of recurrent neural network (RNN) called bidirectional long short-term memory units (BLSTM) to recognize biomedical-named entities.[20] Lakhani et al uses convolutional neural network (CNN) to extracting ICD-O-3 topographic codes from pathology reports.[7] Gao et al uses a hierarchical attention networks for information extraction from pathology reports.[21] Miao et al using a neural model in[22] to extract all Breast Imaging-Reporting and Data System (BI-RADS) finding categories from breast ultrasound reports in Chinese and proved that neural model is better than rule-based and traditional machine-learning methods.[23] Currently, neural network model is a promising approach to extract information from narrative Chinese medical texts.

In this manuscript, we propose a neural model to extract medical entities from radiology reports and employ a rule-based method to group and rank the extracted medical entities. Finally, a method of applying the proposed information extraction method in RIS was presented to enable medical doctors reading radiology reports in a more efficient way.

Objectives

The purpose of this study was to explore methods for extracting, grouping, ranking, delivering, and displaying medical-named entities in radiology reports which can yield efficiency improvement in RIS.

Methods

Data Preparation

The radiology reports used in this work are reports of lung computed tomography (CT) scan, retrieved from two level three hospitals in Shanghai, China, namely, hospital A and hospital B. For the imbalance information distribution of the reports, we retrieved 2,800 informative reports and 200 noncritical reports from these two hospitals to better represent the key information. Informative report means it contains plenty of entities, on the contrary, noncritical report means there is no entity in it. These reports are deidentified manually before retrieval from hospital which includes the anonymization of patient name, the removal of reporting date, and the encryption of patient's ID and report ID. After that, a total of 3,000 reports are randomly shuffled. Before annotating the reports, we defined five types of entities which are current_stage (STAG), lesion_size (SIZE), lesion_description (DESP), lesion_site (SITE), and disease (DISE). A doctoral student in biomedical informatics who knows both linguistics and basic clinical knowledge has annotated all the reports, and two clinicians have examined the annotations from the perspective of clinical practice. The controversial annotations were amended and confirmed by discussion later. Examples of medical-named entities that we annotated are shown on ►Table 1. Two hundred randomly selected reports' entity groups and orders were annotated and confirmed in the same way. We consider entities describing the same lesion as a group. The annotations of entity groups are illustrated in ►Fig. 1. All the data were annotated by BRAT annotation tool.[24] We divided the 3,000 radiology reports into the training set, development set, and test set in a ratio of 10:2:3. In addition, 3,000 reports together with another randomly selected 2,000 unlabeled reports were used to train character embeddings in an unsupervised manner. One hundred radiology reports were extracted for the evaluation of RIS integration. The workflow of data preparation is shown in ►Fig. 2.

Table 1 Examples of medical-named entities that we defined in radiology reports

Type	Examples
Current_stage	左肺上叶切除术后, (after left upper lobectomy); 纵膈肿瘤术后, (after mediastinal tumor surgery)
Tumor_size	0.6 cm, 6–7 mm, 3 mm × 5 mm, 79 mm × 57 mm, 0.9 cm × 1.7 cm
Tumor_description	肿瘤 (tumor), 结节 (nodule), 分叶 (lobulation), 钙化 (calcification), 毛刺 (blur), 磨玻璃 (GGO)
Tumor_site	右肺上叶 (superior lobe of right lung) 左肺下叶外基底段 (basal segment of left lower lobe)
Disease	肺癌 (lung cancer), MT (malignant tumor) 血管瘤 (hemangioma), AIS, MIA, AAH

Abbreviations: AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; GGO, Ground Glass Opacity; MIA, minimally invasive adenocarcinoma; MT, malignant tumor.

Named-Entity Recognition

Like many other studies,[9,20,25–27] we regard the NER in radiology report as a sequence-labeling problem. The IOB (inside, outside, and beginning) schema was adopted for named entity labeling. The IOB schema means labeling tokens inside a named entity as I-type, tokens outside entity as O-type, and the beginning token of entity as B-type. Take lesion_site as an example, the first token of the entity will be labeled as B-SITE; other tokens of the entity will be labeled as I-SITE; the tokens which is not a constituent of any entity will be labeled as O. IOBES with additional "S" to represent singleton entities and "E" to represent the end of entities is another frequently-used labeling schema. Some research showed that IOBES is more expressive,[28,29] but other results showed it doesn't have a significant improvement over IOB.[22] Because the labeling schema is not the focus of our model, we just randomly chose

Fig. 1 Visualization of entity group annotations. DESP, lesion_description; SITE, lesion_site; SIZE, lesion_size.

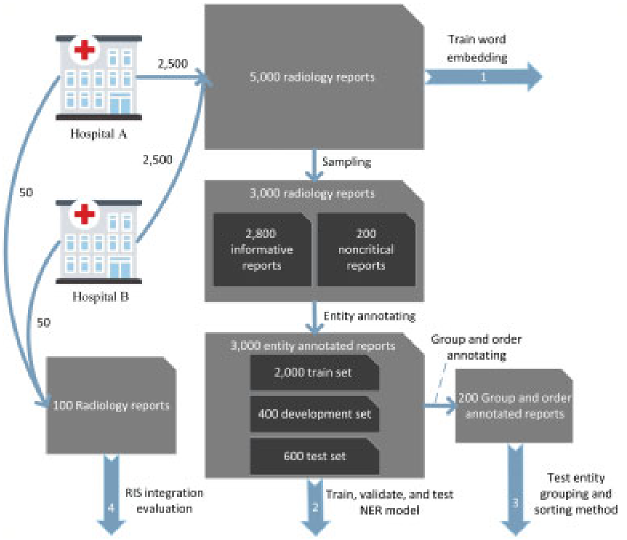

Fig. 2 The workflow of data preparation. Hospital icon made by Freepik from *www.flaticon.com*. NER, named-entity recognition; RIS, radiology information systems.

IOB as our labeling schema. Our model consists of three main parts: Chinese text embedding, bidirectional gated recurrent unit (BGRU[30]) sequence-to-sequence encoder network, and conditional random field. The detailed architecture is described further.

Chinese Text Embedding

The first step of NER is mapping each word in the vocabulary to a vector. This mapping constructs a numerical input representation for the downstream model. In the simplest way, one-hot encoder can represent the vocabulary with one-dimension per word, but its dimensionality would be very large, and the vectors can't capture the semantic and syntactic similarity of the words. Some methods, such as neural networks,[31] dimensionality reduction,[32] and probabilistic models[33] have been proposed to tackle these problems. We used a hierarchical embedding method to convert tokens into vectors of real numbers. The composition hierarchy of Chinese words can be divided into three levels: word level, character level, and radical level, which are illustrated in ►**Fig. 3**. The hybrid embedding method we used can generate a dense and distributed representation of a token at different level. Compared with plain word embedding method and traditional methods such as bag of words and one-hot encoding, this hybrid embedding method shows better performance according to some literatures.[34–36] We adopted the widely used method called word2vec[31] to construct embedding layers. Word2vec contains two models called continuous bag of words (CBOW)) and skip-gram. CBOW predicts the current word based on the context, whereas skip-gram predicts surrounding words given the current word. They both use a concise neural network to learn the distributed representation of words.

Unlike English, there is no spacing between Chinese words. Therefore, we employ Jieba Chinese word segmentation tool[37] to segment Chinese characters into words. Before being fed into an embedding layer, the segmentation boundary was

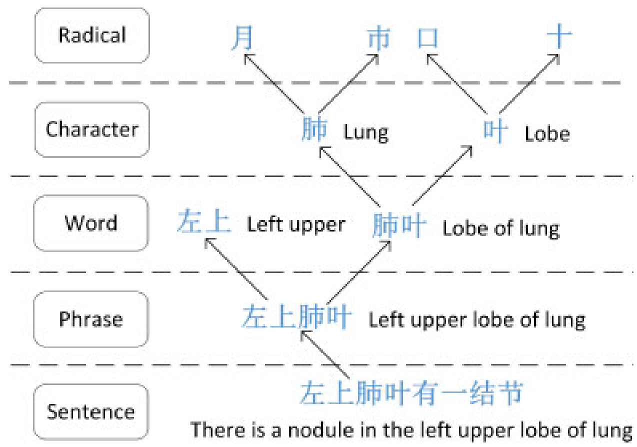

Fig. 3 The composition hierarchy of Chinese text.

encoded by BIES (Begin, Inside, End, Single) schema. The first character of the word was represented by B, the internal characters were represented by I, the end character was represented by E, and the single-character word was represented by S. At the character level, we utilize a large set of unlabeled radiology reports to pretrain character embeddings which sequentially fine-tuned on the training set. Chinese character is made up of several radicals. Characters with similar semantic meanings usually contain the same radicals, particularly for characters in medical text. For example, 月 is a radical represent flesh or moon, so many anatomic sites comprise it, such as 肺 (lung), 肝 (liver), 肠 (intestines), 肾 (kidney), and 胃 (stomach). Also, 疒 is a radical means illness, so it is a constituent part of many disease related characters, such as 癌 (cancer), 瘤 (tumor), 疮 (sore), and 疟 (malaria). Accordingly, it is reasonable to put the radical information into token embeddings for medical text. We get radicals of characters from *http://tool.httpcn.com/zi/* and use word2vec to get radical embeddings. Then a CNN is employed to train a fixed-size radical-level representation for each character. Finally, we concatenate these three embeddings to generate hybrid embeddings.

Bidirectional Gated Recurrent Unit

RNN is a series of neural networks which usually used to model sequence data. With recursion, traditional RNN is hard to prevent vanishing and exploding gradient problem and tend to be biased toward its most recent inputs.[38] GRU or LSTM as a variant of RNN has been proposed to combat these problems to learn long-term dependency of the sequence. GRU unit contains two gates to update and reset the information. Each gate controls the proportion of information forgotten and transmitted to the next time step. For a piece of medical report text or a sequence of tokens, the simple GRU only computes a representation of the left context of each character. However, the ignored right context is equally important for characters' representation in sentences. A proven method to tackle this shortage is the bidirectional RNN.[39] BGRU uses a forward GRU to learn the left context representation, $\overrightarrow{h_t}$, and a backward GRU to learn the right context representation, $\overleftarrow{h_t}$. Then concatenating the left and

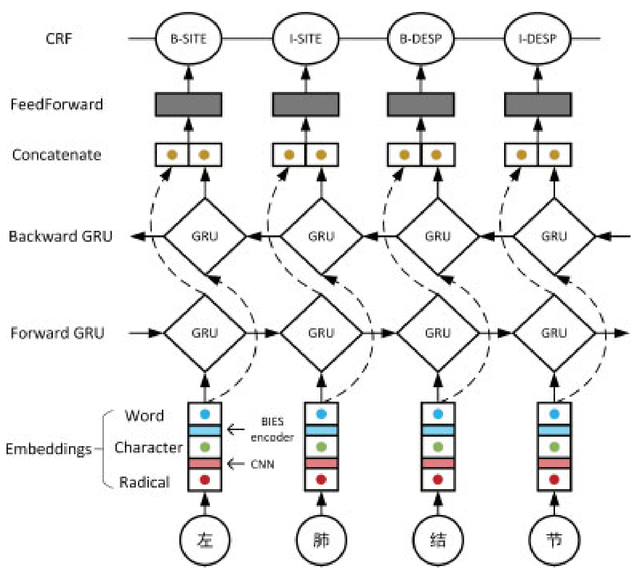

Fig. 4 The architecture of the Multi-Embedding-BGRU-CRF model. B-DESP, beginning-lesion_description; B-SITE, beginning-lesion_site; CNN. convolutional neural network; CRF, conditional random fields; GRU, gated recurrent unit; I-DESP, inside- lesion_description; I-SITE, indise-lesion_site.

right context to get the final representation: $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$ which proved to be very useful for sequence-labeling task.

Conditional Random Field

It is intuitive to stack a classifier on top of the BGRU model to assign a label for each token in a sentence. However, in NER task, there is strong dependency across output labels. For example, it is illegal for I-SIZE following B-DISE. Considering the hard constraints on the output labels, instead of individual modeling, the labeling decision for each token, we employ a CRF layer to jointly model the output labels of the whole sentence. CRF layer can perceive correlations between labels and construct consistent interpretations. Combining bidirectional RNN and CRF, to solve the sequence-labeling problem, is proved to be effective in many literatures.[22,26,27] For sequence labeling, the CRF we mention here refers to linear chain CRF.

Multi-Embedding-bidirectional gated recurrent unit-conditional random field

The combination of the three parts mentioned above which we denoted as Multi-Embedding-BGRU-CRF is the model we used in NER. The architecture of the model was shown in ►Fig. 4. Given an input sentence, $X = (X_1, X_2, \cdots, X_n)$, we concatenate the word level representation R_w, character embedding R_c, and radical embedding R_r of each Chinese character in a sentence to get the final token representation R_i. Then we get a sequence of vectors $R = (R_{i1}, R_{i2}, \cdots, R_{in})$ represent the sentence. We feed the vector sequence into a forward GRU $\overrightarrow{f_{\grave{e}}}(\cdot)$ and a backward GRU $\overleftarrow{f_{\grave{e}}}(\cdot)$ and concatenate their output before sent to feed forward layers. Finally, a CRF layer was adopted to jointly decode the output labels based on feedforward layers' output.

Entity Grouping and Ranking

For displaying the entities systematically, we have grouped the entities described the same lesions and ranked the groups according to the lesion size. Only sized lesions and the entities related to them are grouped, other entities are considered less important and ignored in this step. This is because entities in a radiology report are probably plentiful and the displayed space is limited, we only show the most important entities. A simple rule-based approach was employed to group and rank the entities that extracted by NER step. An entity group represents an information frame of a lesion which contains all the entities related to the same lesion. One example of using the proposed rule-based method for entity grouping was illustrated in ►Fig. 5. The grouping rules are derived from an exploration of several pieces of reports. We knew the numbers of each type of entity and the relative position of entities for one lesion through the exploration. We found that only the tumor_size has one-to-one relationship to lesion. One lesion may have several lesion_description and not every lesion is followed by a disease; One lesion_site may corresponds to multiply lesions; current_stage is not specific to one lesion. So we regard tumor_size as the start point of the group entities search. And according to the relative distribution of each type of entity in group, we designed the search orders and directions of rest entities. The method can be described as follows: first, we scan all entities to find out entities with type

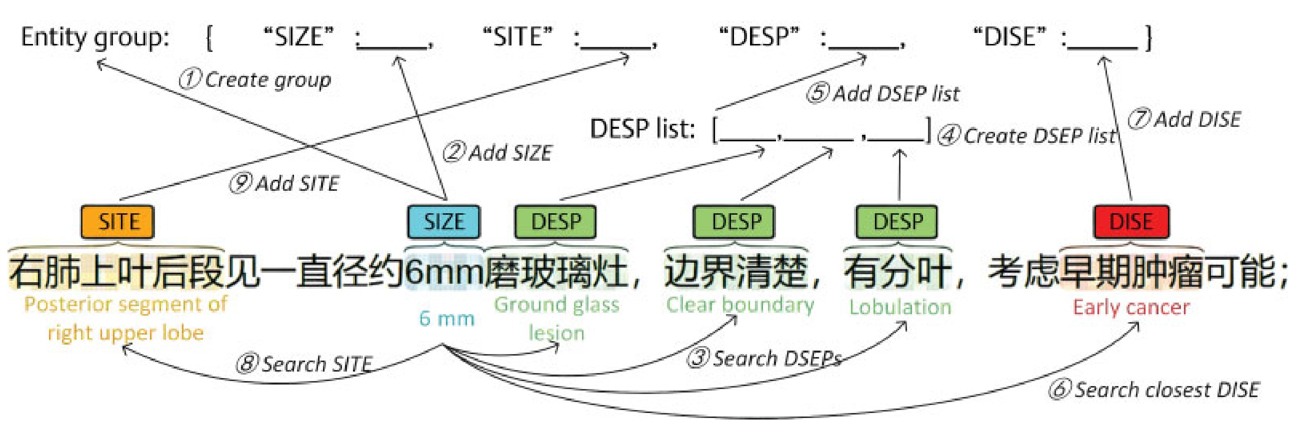

Fig. 5 The illustration of the rule-based entity grouping method. DESP, lesion_description; DISE, disease; SITE, lesion_site; SIZE, lesion_size.

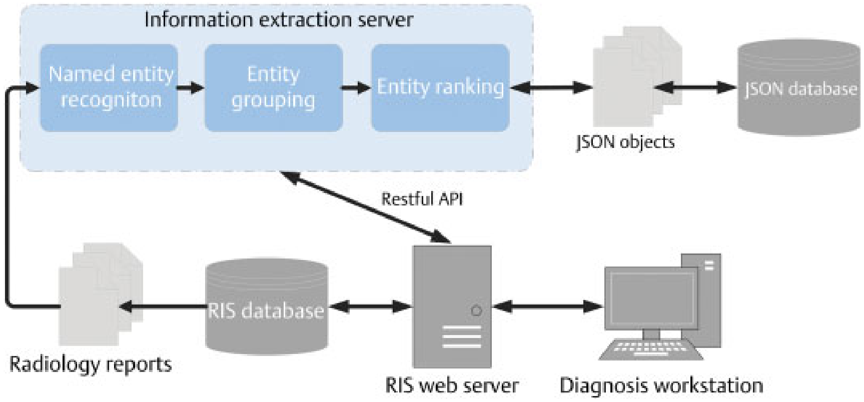

Fig. 6 The integration of information extraction and RIS. API, application program interface; JSON, JavaScript Object Notation; RIS, radiology information systems.

of lesion_size. Then, we create a group for each lesion_size entity and add the lesion_size entity to the group (illustrated in figure as ①, ②). Second, for each lesion_size entity, start with its position, we search lesion_description entities along with both directions of the sentence until encounter an entity that is not a lesion_description. Second, we get a lesion_description list in which all entities describe the same lesion with corresponding lesion_size entity. We added the entity list to the group (illustrated in figure as ③, ④, ⑤). Third, we put the disease entity closest to the lesion_description entities into the group if exists (illustrated in figure as ⑥, ⑦). Fourthly, the nearest lesion_site entity in the left side of lesion_size entity was added to the group (illustrated in figure as ⑧, ⑨). After that, we used regular expressions to get the three-dimension values, two-dimension size values, or diameter values from lesion_size entity and then used these size values to rank the entity group. The entity group with the greater size value has higher display priority. In addition, the current_stage entity that, generally, describes patients' current treatment progress information is not specified to a single lesion, so it does not belong to any group.

Integration into Radiology Information Systems

Our original RIS consists of the web server, database server, registration workstation, exam workstation, and diagnosis workstation. Based on it, we added an information extraction server that deployed the algorithm which we described in the previous section and a JSON (JavaScript Object Notation) database to store the extracted entities. Furthermore, the web server has been changed to enable RIS to show and highlight the extracted entities. The method of integration of information extraction and RIS is shown in ►Fig. 6. When a radiologist finished a radiology report in diagnostic workstation and uploaded it to the web server, in addition to sending the report to the RIS database for storage, the web server sends the report to the information extraction server via RESTful API (application program interface based on representational state transfer technology)[40] to extract medical entities from the report. Then all the extracted entities are packed into a JSON object and saved to a JSON database which implemented by PostgreSQL. When a clinician operates the diagnostic workstation to obtain reports, the web server retrieves medical entities from information extraction server via RESTful API and then adds them to the render of web pages.

Web pages also modified to accommodate the display of the entities. We exhibit entities in report list page and report view page, respectively. Different types of entities are displayed in different colors. In report list page, the entities are displayed as a field of the list. At the beginning, the current_stage entities are shown to cast light on treatment stage of the patient. Then, other entities are shown in groups. In each group, the entities are shown in the order of lesion_site, lesion_description, lesion_size, and disease. The newly designed report list page is shown in ►Appendix Fig. A2. A report may have multiple groups of entities and clinicians can configure the web page to display a part of groups according to their preference and screen size. In the report view page, all the extracted entities are highlighted in the report with different colors and additional listed according to their order in the text. The newly designed report view page is shown in ►Appendix Fig. A3.

Evaluation Method

The evaluation of our approaches was conducted in two steps. We separately evaluated the accuracy of the information extraction method and the effectiveness of the information extraction method using in RIS for improving clinician's reading efficiency of radiology reports. For the information extraction method, we use precision, recall, and $F1$ score as the measurement to evaluate the NER model. A simple neural model similar to[22] which only contains no pretrained character embedding, BGRU, and CRF is employed as a baseline. Precision, recall, and $F1$ score are formulated as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

, where TP, FP, and FN denote the numbers of true positives, false positives, and false negatives, respectively. $F1$ score is the harmonic mean of precision and recall. Also, the effectiveness of the submodules, the pretrain method, and the RNN unit was evaluated. The above model is implemented and evaluated in AllenNLP,[41] an open-source NLP research library. Apart from that, performance was calculated on data from different hospitals to evaluate the generalization ability of the model. Two hundred randomly selected radiology reports and their entities were used to evaluate the accuracy of the entity grouping and ranking method. All the hyper parameters of the evaluated models are fine tuned to get the best performance. All the measures are calculated on the average of five times run to make the result reliable.

We designed an evaluation task to assess the effectiveness of the information extraction method used in RIS for improving the efficiency of reading radiology reports. We acquired 100 radiology reports from the RIS database of two hospitals. Two clinicians were asked to identify the size, site, and morphological descriptions include ground glass, calcification, boundary sharpness, burr, lobulation, and cavity of the primary lesion which are important for cancer diagnosis and treatment planning. The clinicians were required to carry out the task separately on the information extraction applied RIS and original RIS. They are asked to fill a test form for each radiology report. The time that clinicians spend on the task and the accuracy they achieved on the two RISs are recorded automatically. The accuracy was calculated by comparing the finished test forms, to the forms carefully filled by another two clinicians, which were treated as the right answer.

Results

For NER, the result is evaluated on test set which contains randomly selected 600 test radiology reports as illustrated in ►Fig. 2. As can be seen in ►Table 2, we measured precision, recall, and *F*1 score of present models and the baseline, respectively. Our model Multi-Embedding-BGRU-CRF has achieved 30% error reduction compared with baseline. The pretrain evaluation in ►Table 2 shows the performance of different pretrained setups. Character embeddings pretrained by radiology reports are much better than without pretrain and pretrained by generic corpus like Wikipedia,

Table 2 Measurements of our model compared with baseline, model with LSTM unit and other pretrain options in the NER task

Model	Precision	Recall	*F*1 score
Baseline	94.18 ± 0.13	94.11 ± 0.07	94.15 ± 0.10
Multi-Embedding-BGRU-CRF	95.88 ± 0.18	95.87 ± 0.15	95.88 ± 0.14
Multi-Embedding-BLSTM-CRF	95.72 ± 0.15	95.58 ± 0.13	95.65 ± 0.14
Pretrain evaluation			
Wiki pretrained	95.31 ± 0.11	95.50 ± 0.03	95.40 ± 0.07
Without pretrain	95.36 ± 0.06	95.60 ± 0.19	95.48 ± 0.10

Abbreviations: BGRU-CRF, bidirectional gated recurrent unit-conditional random field; BLSTM, bidirectional long short-term memory units; NER, named-entity recognition.

which give the insight that training word vectors in a specific domain on a large-scale unlabeled data can improve the performance of domain-specific NLP tasks, such as medical NER. Surprisingly, no pretrained results are better than Wikipedia pretrained results. This may be due to the huge difference between the language of the radiology reports and the language of the common articles. Another famous RNN unit,– LSTM also has been evaluated. The result shows LSTM is slightly worse than GRU. Possibly because GRU has fewer parameters, so it can converge faster than LSTM and avoid over-fitting on small datasets, and our dataset is indeed small for deep learning. We sequentially add submodules to the baseline model to evaluate the contribution of submodules to our model and the results are shown in ►Fig. 7. From this perspective, pretrain, radical embedding, and word embedding play a similar role in improving performance. Pretrain shows 10% error reduction to the baseline; radical embeddings and word segmentation embeddings both show 12% error reduction when been added to the model.

►Table 3 shows the measures by entity type. The amount of different type of entities in the test set varies greatly. But the imbalance distribution of entity type seems to have little impact on performance. This may be due to the fact that the entity types with larger amount often have more diverse

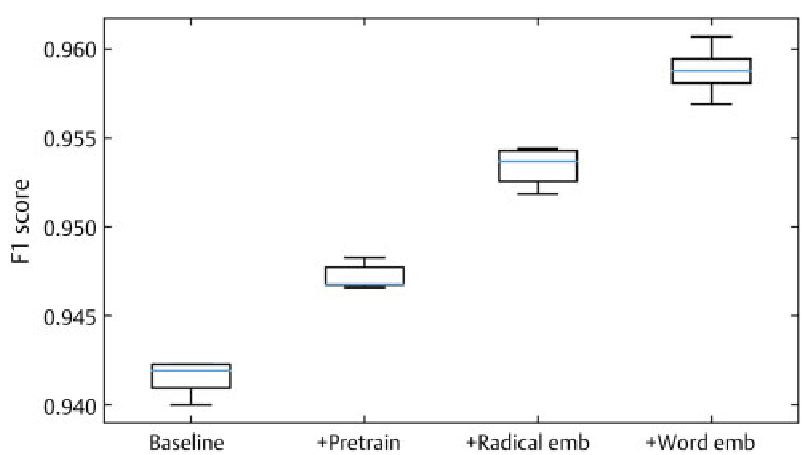

Fig. 7 Measures of submodules' contribution to the final model.

Table 3 Measures for each type of entities: counts are the amount of the corresponding type counted in test set

Type	Count	Precision	Recall	*F*1 score
STAG	145	96.33	97.24	96.78
SIZE	446	98.22	99.01	98.62
DESP	3,133	95.67	96.00	95.84
SITE	1,910	95.60	95.27	95.43
DISE	241	95.56	92.86	94.19

Abbreviations: DESP, lesion_description; DISE, disease; SITE, lesion_site; SIZE, lesion_size; STAG, lesion_stage.

expressions, which means it is harder for a model to learn though it has more training data. And we believe that the NER accuracy of any specific entity type depends more on the complexity and diversity of its own expression. For example, the entity lesion_size which has a relatively fixed form is easier for model to learn and gets a higher *F*1 score. To go deep into the error of our model, we calculate the confusion matrix. The confusion matrix we calculated is shown in ►**Fig. 8**. The Y axis indicates which type the token should be classified. The X axis shows the type that the model predicted for the token. The confusion matrix reveals the misclassification of each entity types. All the values that are not located at the diagonal are the count of the misclassified token. We can get a detailed insight of the inaccuracy of the model from it. From the confusion matrix, we can know that most of the tokens are not named entities. The lesion_description and lesion_site are the two entities most confusing to the model. Furthermore, the measures by hospitals are shown in ►**Table 4**. The present model's performance on two hospitals' data is basically consistent. This indicates data from different sources can be learned jointly in our model. Also, we have tried to train the model from data from hospital B and evaluated it on data from hospital A. The resulted *F*1 score is only 82.50% which is much worse than before which

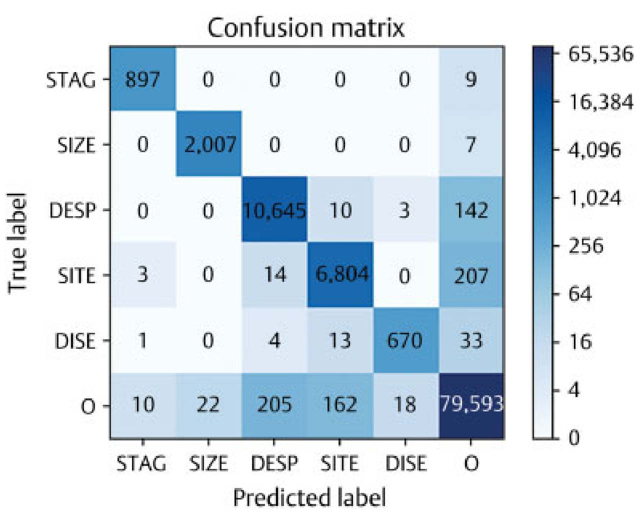

Fig. 8 Confusion matrix of NER for radiology reports. O denotes the tokens that are not medical-named entities. DESP, lesion_description; DISE, disease; SITE, lesion_site; SIZE, lesion_size; STAG, lesion_stage.

Table 4 Model performance by hospital

Hospital	Precision	Recall	*F*1 score
Hospital A	95.88	95.58	95.73
Hospital B	95.83	96.17	96.00

shows different data distribution of the two hospitals. The intuition that we trained the model from the data combined from two hospitals is to make the model more robust to different sources of data and avoid over-fitting regarding the small size of data from each hospital. And we also tried to train models for two hospitals, respectively. The resulted *F*1 score of hospital A and hospital B is 96.73 and 95.42% which shows the combined training improved the performance of worse one at the cost of damaging the better one. Therefore, the combined training model is more balanced and we believe that it learned more generic features of radiology reports compared with the separated training.

Two hundred reports were manually selected and annotated with entity group and entity order for the grouping and ranking accuracy measurement. The results are shown in ►**Table 5**. Entity ranking has got a higher accuracy than entity grouping. This is because getting values of size from lesion_size entities only requires capturing the syntactic structure of the lesion_size entity, but grouping entities need to capture semantic correlations between the entities in reports which are tough for rule-based methods to deal with. The limited rules can only tackle the most frequent cases and the expression of the radiology reports is quite diverse. Though writing more rules to cover more situations improves the accuracy, it is very time-consuming and as the number of rules goes up, it is hard to write a new rule without affecting the existing ones. It also makes the system more complex and difficult to maintain.

One hundred radiology reports were adopted to evaluate the integration of information extraction and RIS. The result was compared with the RIS without information extraction. The time consumption and the correctness of two clinicians in the evaluation task are shown in ►**Table 6**. Compared with the original RIS, RIS with information extraction significantly reduce report reading time by half and even with a little higher accuracy in the task. This demonstrated that RIS with information extraction can improve the efficiency of clinician in reading radiology reports without error increase. According to the clinicians' feedback, RIS with information extraction assist reading reports in three ways. First, it helps clinicians quickly skip over the normal reports which with less crucial information. Second, by highlighting the extracted entities in different colors in report view page, clinicians can quickly grasp the focus of the report. Third, based on the annotated entities in reports, much information that clinicians are asked to fill can be

Table 5 The accuracy of entity grouping and ranking

Task	Accuracy
Grouping	91.03
Ranking	99.23

Table 6 The RIS integration evaluation: seconds used is the total time spent in reading 100 reports

		Seconds used	Accuracy
RIS without IE	Clinician 1	1,516	94.38
	Clinician 2	1,606	96.50
	Average	1,561	95.44
RIS with IE	Clinician 1	784	98.13
	Clinician 2	904	96.88
	Average	844	97.51

Abbreviations: IE, information extraction; RIS, radiology information systems.

preidentified and clinicians only need to confirm it, rather than search for this information from scratch.

Discussion

Effectiveness of the Approach

We introduced an information extraction approach for extracting key information from Chinese radiology reports. The proposed approach cascades a neural NER model and a rule-based entity grouping and ranking method. We used 5,000 radiology reports to train and test the information extraction model and get a quiet good result. Our approach does not rely on professional and comprehensive medical dictionaries like many traditional methods. Thus, our model does not have an out-of-vocabulary (OOV) problem because of the word vectors are trained from many unlabeled data in an unsupervised manner. Besides, with the ability of learning long-term dependency of GRU, our deep neural model can learn the context-sensitive semantic structures which are hard for rule-based method and even traditional learning-based method. For example, it is straightforward to distinguish the diameter of lesions, the diameter of lymph nodes, and the depth of pleural effusion for our model and this is hard for traditional method to differentiate because these tokens are identity as a character string. We only need to treat them differently while annotating data. The model can learn the semantic difference between them according to the context. Though the problem can be tackled by writing regular expressions to match the contexts in rule-based NER systems, when there are too many contexts to match, the problem becomes unmanageable. Also, manually constructed features for matching context relies heavily on the researchers' linguistics and feature engineering expertise. Moreover, our model's robustness for data from different source has been proved through the performance measurement by hospital. We attribute this to the strong learning capability of deep neural network model. The rule-based method of grouping and ranking entities is quite simple and effective. In many other information extraction works,[12] a relation extraction is usually cascaded on NER to get a structured information frame. The presented entity grouping method can be regarded as a simple rule-based relation extraction procedure. And we recognized that the learning-based method for relation extraction yields better performance currently.[42] The obstacles to using learning-based method is the expensive time and labor cost of manually tagging entity relation for the training data. A methodology of delivering and displaying medical-named entities are proposed to apply the information extraction approach to RIS. To evaluate the application of the approaches, two clinicians were asked to complete the evaluation task we designed. The result shows that the RIS equipped with our information extraction method can improve the efficiency of reading radiology reports.

Limitations

The proposed approach is trained and evaluated on the CT reports of lung. Only the CT reports of lung can be analyzed by our information extraction server. The effectiveness of the proposed approach for reports of other body parts and imaging modalities needs to be further evaluated. Compared with NER and entity ranking, entity grouping has poorer performance, so it is the main obstacle to improve the overall performance of the system. The performance of entity grouping can be further improved by using cutting-edge learning-based method, such as neural relation extraction.[43] Though we demonstrated the effectiveness of the integration of information extraction and RIS, RIS vary from vendor to vendor; different clinicians, especially those from different hospitals, have different working procedures and personal preference of using RIS, the delivery and display of the medical-named entities in RIS should be more configurable. The application was only evaluated in the RIS deployed in hospital A. Further study is required in larger cohorts to determine whether the proposed approach can be widely applied in different RIS for different clinicians.

Conclusions

This article presented an information-extraction solution for improving the efficiency of reading report in RIS. The neural network model shows great potential in extracting medical-named entities from radiology reports, especially for languages that lacking medical dictionaries and NLP tools. A rule-based entity grouping and ranking method has been introduced to organize the extracted medical entities. The way in which entities are delivered and displayed in RIS can be referenced by other researchers. The integration of information extraction and RIS has been demonstrated to be effective in improving clinicians' efficiency of reading radiology reports. The pipeline of extracting, grouping, ranking, delivering, and displaying medical-named entities can provide inspiration for researchers who are engaged in medical information extraction applications. The application of the extracted and organized medical-named entities in other medical systems such as EMRS (Electronic Medical Record System), PIS (Pathology Information System), and HIS (Hospital Information System) can be further discussed.

Conflict of Interest
None declared.

Acknowledgments

The authors have special thanks for Professor Jiangguo Zhang, the former director of their lab. The authors would like to thank the colleagues of Hospital A and Hospital B for this research. This project was supported by the grants from the Ministry of Science and Technology of China under project 22001177YYFFCC00111122990021.

References

1 Society E; European Society of Radiology (ESR); Guidelines from the European Society of Radiology (ESR). Good practice for radiological reporting. Insights Imaging 2011;2(02):93–96
2 Bosmans JML, Weyler JJ, De Schepper AM, Parizel PM. The radiology report as seen by radiologists and referring clinicians: results of the COVER and ROVER surveys. Radiology 2011;259(01):184–195
3 Kahn CE Jr., Langlotz CP, Burnside ES, et al. Toward best practices in radiology reporting. Radiology 2009;252(03):852–856
4 Ganeshan D, Duong PT, Probyn L, et al. Structured Reporting in Radiology. Acad Radiol 2018;25(01):66–73
5 Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: A literature review. J Biomed Inform 2018;77:34–49
6 Barbosa F, Traina AJ, Muglia VF. Meta-generalis: A novel method for structuring information from radiology reports. Appl Clin Inform 2016;7(03):803–816
7 Lakhani P, Kim W, Langlotz CP. Automated detection of critical results in radiology reports. J Digit Imaging 2012;25 (01):30–36
8 Hassanpour S, Langlotz CP. Information extraction from multi-institutional radiology reports. Artif Intell Med 2016;66:29–39
9 Esuli A, Marcheggiani D, Sebastiani F. An enhanced CRFs-based system for information extraction from radiology reports. J Biomed Inform 2013;46(03):425–435
10 Xu H. Research on medical entity extraction based on supervised learning. 2016Q15
11 Wang HC, Zhao TJ. SVM-based biomedical name entity recognition. Harbin Gongcheng Daxue Xuebao/journal Harbin Eng Univ 2006;27:570–574
12 Yang J-F, Yu Q-B, Guan Y, Jiang Z-P. An overview of research on electronic medical record oriented named entity recognition and entity relation extraction. Zidonghua Xuebao/Acta Autom Sin 2014;40(08):1537–1562
13 Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32 (Database issue):D267–D270
14 Lipscomb CE. Medical subject headings (MeSH). Bull Med Libr Assoc 2000;88(03):265–266
15 Q16International Health Terminology Standards Development Organisation (IHTSDO). SNOMED CT Technical Implementation Guide; 2014
16 Langlotz CP. RadLex: a new method for indexing online educational materials. Radiographics 2006;26(06):1595–1597
17 Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010; 17(05):507–513
18 Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp 2001: 17–21
19 Friedman C, Hripcsak G, DuMouchel W, Johnson SB, Clayton PD. Natural language processing in an operational clinical information system. Nat Lang Eng 1995;1(01):83–108
20 Lyu C, Chen B, Ren Y, Ji D. Long short-term memory RNN for biomedical named entity recognition. BMC Bioinformatics 2017; 18(01):462
21 Gao S, Young MT, Qiu JX, et al. Hierarchical attention networks for information extraction from cancer pathology reports. J Am Med Inform Assoc 2017 (e-pub ahead of print); Doi:10.1093/jamia/ocx131
22 Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C. Neural Architectures for Named Entity Recognition. In: proceedings of NAACL-HLT 2016. San Diego, CA; 2016
23 Miao S, Xu T, Wu Y, et al. Extraction of BI-RADS findings from breast ultrasound reports in Chinese using deep learning approaches. Int J Med Inform 2018;119:17–21
24 Stenetorp P, Pyysalo S, Topi G, Ohta T, Ananiadou S, Tsujii J. BRAT: a Web-based Tool for NLP-Assisted Text Annotation. In: proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Avignon, France; 2012
25 Nandhakumar N, Sherkat E, Milios EE, Gu H, Butler M. Clinically significant information extraction from radiology reports. In: Proceedings of the 2017 ACM Symposium on Document Engineering–DocEng. New York, NY: ACM Press; 2017:153–162
26 Ma X, Hovy E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In: proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany; 2016
27 Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. Available from: https://arxiv.org/pdf/1508.01991.pdf. Accessed June 20, 2019
28 Dai HJ, Lai PT, Chang YC, Tsai RTH. Enhancing of chemical compound and drug name recognition using representative tag scheme and fine-grained tokenization. J Cheminform 2015;7(Suppl. 1, Text mining for chemistry and the CHEMDNER track):S14
29 Ratinov L, Roth D. Design challenges and misconceptions in named entity recognition. In: Proceedings of the 13th Conference on Computational Natural Language Learning–CoNLL. Boulder, CO; 2009
30 Cho K, van Merrienboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar; 2014
31 Mikolov T, Corrado G, Chen K, Dean J. Efficient estimation of word representations in vector space. Available from: https://arxiv.org/pdf/1301.3781.pdf. Accessed June 21, 2019
32 Levy O, Goldberg Y. Neural word embedding as implicit matrix factorization. Available from: https://papers.nips.cc/paper/5477-neural-word-embedding-as-implicit-matrix-factorization.pdf. Accessed June 20, 2019.
33 Globerson A, Chechik G, Pereira F, Tishby PN. Euclidean embedding of co-occurrence data. J Mach Learn Res 2007;7: 2265–2295
34 Sun Y, Lin L, Yang N, Ji Z, Wang X. Radical-enhanced Chinese character embedding. In: Loo CK, Yap KS, Wong KW, Teoh A, Huang K, eds. Neural Information Processing: 21st International Conference. Kuching, Malaysia: Springer; 2014:279–286
35 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; 2015
36 Yin R, Wang Q, Li R, Li P, Wang B. Multi-granularity Chinese word embedding. In: proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX; 2016
37 Jieba Chinese text segmentation.Q17
38 Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 1994;5 (02):157–166
39 Graves A, Schmidhuber J. Framewise phoneme classification with bidirectional LSTM networks. In: Proceedings of the International Joint Conference on Neural Networks. 2005

40 Fielding RT, Taylor RN. Principled design of the modern Web architecture. ACM Trans Internet Technol 2002;2(02): 115–150

41 Gardner M, Grus J, Neumann M, et al. AllenNLP: A Deep Semantic Natural Language Processing Platform. In: proceedings of workshop for NLP Open Source Software. Melbourne, Australia; 2018

42 Pawar S, Palshikar GK, Bhattacharyya P. Relation Extraction: A Survey. Available from: https://arxiv.org/pdf/1712.05191.pdf. Accessed June 21, 2019.

43 Lin Y, Shen S, Liu Z, Luan H, Sun M. Neural Relation Extraction with Selective Attention over Instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016

Appendix

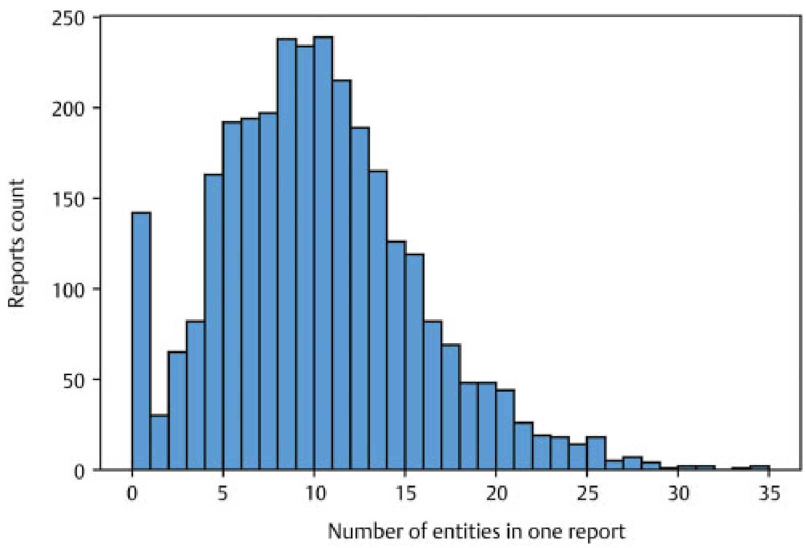

Fig. A1 The distribution of medical-named entities among radiology reports.

Fig. A2 The newly designed report list page for showing extracted medical entities. DESP, lesion_description; DISE, disease; SITE, lesion_site; SIZE, lesion_size; STAG, lesion_stage.

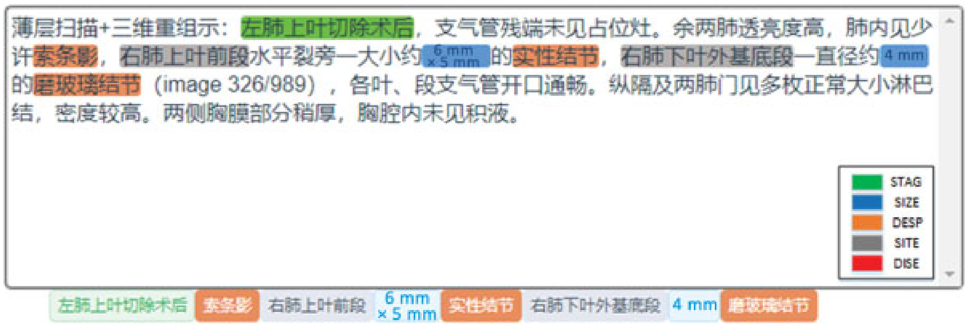

Fig. A3 The newly designed report view page for highlighting the medical entities. DESP, lesion_description; DISE, disease; SITE, lesion_site; SIZE, lesion_size; STAG, lesion_stage.