Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200151

End-to-End Approach for Structuring Radiology Reports

Kento SUGIMOTO[a,1], Toshihiro TAKEDA[a], Shoya WADA[a], Asuka YAMAHATA[a], Shozo KONISHI[a], Shiro MANABE[a] and Yasushi MATSUMURA[a]
[a] *Department of Medical Informatics, Osaka University Graduate School of Medicine*

Abstract. Radiology reports include various types of clinical information that are used for patient care. Reports are also expected to have secondary uses (e.g., clinical research and the development of decision support systems). For secondary use, it is necessary to extract information from the report and organize it in a structured format. Our goal is to build an application to transform radiology reports written in a free-text form into a structured format. To this end, we propose an end-to-end method that consists of three elements. First, we built a neural network model to extract clinical information from the reports. We experimented on a dataset of chest X-ray reports. Second, we transformed the extracted information into a structured format. Finally, we built a tool that enabled the transformation of terms in reports to standard forms. Through our end-to-end method, we could obtain a structured radiology dataset that was easy to access for secondary use.

Keywords. Natural Language Processing, Radiology Report, Information Extraction

1. Introduction

Radiology reports are created by radiologists to communicate with referring clinicians, and play an important role in patient care. Radiology reports are usually written in free-text format. It has been reported that the ambiguity of terminology or style in free-text can reduce the clarity of the report, causing inaccurate communication [1]. European Society of Radiology has taken the initiative in developing structured reporting to improve the quality of radiology reports [2]. They mention structured reporting also has the potential to facilitate clinical research and the development of radiological applications.

There are a lot of studies on information extraction from radiology reports [3]. Some studies examined Japanese radiology reports [4,5]. In recent years, extraction method using a deep learning has drawn much attention [6]. However, most studies have only focused on the extraction process and did not refer to the structuring process.

In this study, we propose an end-to-end approach for structuring radiology reports. Our aim is utilizing massive amounts of unstructured reports as training data for a lesion detection system [7]. Specifically, we take three steps: *extraction, structuring* and *normalization*. In the first step, we build a recurrent neural network-based model for entity recognition. In the second step, we organize the extracted information for storage in a database. The data is stored in a tabular database for accessibility. It is well-known that

[1] Corresponding Author, Osaka University, 1–1 Yamadaoka Suita, Osaka, Japan;
E-mail: sugimoto.kento@hp-info.med.osaka-u.ac.jp

radiology reports contain heterogeneous writing styles, including non-standard terminology and abbreviations [8]. In the third step, we transform non-standard terms in the report to a standard form.

2. Material and methods

2.1. Radiology dataset

In this study, we used chest X-ray reports from 2000 to 2017 that were stored in the radiology information system at Osaka University Hospital. The dataset consisted of 319,130 reports, most of which were written in Japanese. This study was approved by the institutional review board of Osaka University Hospital (Permission number: 17166).

2.2. Entity recognition

2.2.1. Gold Standard

We randomly sampled 5,000 reports for a gold standard dataset. A report was segmented into a single sentence and word tokenization was implemented using MeCab [9].

One non-medical expert was responsible for the annotation process. Two medical experts (a clinician and a radiological technologist) reviewed the results.

A previous study defined 5 semantic classes (*Anatomy, Anatomy Modifier, Observation, Observation Modifier, and Uncertainty*) [10]. Another defined 4 semantic classes (*Clinical Finding, Body Location, Descriptor and Medical Device*) [6]. With reference to previous studies, we defined 3 semantic classes: *Clinical Finding (CF), Body Location (BL),* and *Body Location Modifier (BLM)*. CF includes terms related to observation and abnormalities in the report. BL includes terms related to the anatomical area. BLM encompasses terms that modify the BL. To differentiate the certainty of the Clinical Findings, they were split into three labels: *Clinical Finding Positive (CFP), Clinical Finding Suspicious (CFS), Clinical Finding Negative (CFN)*. Figure 1 shows an example of annotation.

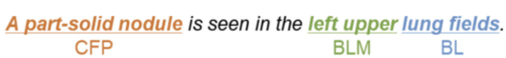

Figure 1. annotation example

2.2.2. Neural Network Architecture

Our deep learning model is based on the architecture described by Ma and Hovy [11]. The encoder part of this model is composed of the Bidirectional Long Short Term Memory (BiLSTM) [12] layers. The character representation of each word computed by convolutional neural networks (CNNs) is concatenated with the word representation. Finally, the output of BiLSTM is fed to the CRF [13] layers to decode the label sequence.

2.3. Structuring

Entities from the report need to be transformed to tabular format records. We created simple rules using position information of the entities. In the common word order of radiology reports written in Japanese, BL appears after BLM, before CF. Based on this assumption, pairs of BLM and BL, BL and CF were created (Figure 2).

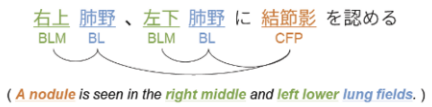

Figure 2. an example of pairs in the report

When transforming, a pair of BLM term and BL term was concatenated as one term, and a CF was divided into CF and Certainty. Multiple records were created from one sentence when BL and CF had a one-to-many (many-to-one) relationship (Figure 3).

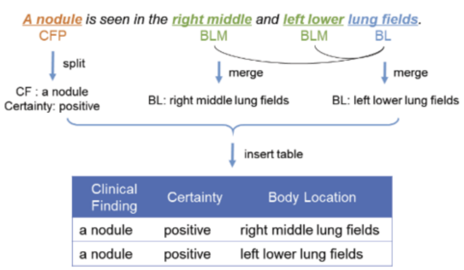

Figure 3. an example of creating records from the extracted entities.

2.4. Normalization

Each term needs to be normalized automatically. We created a concept table from scratch since there is no comprehensive radiology vocabulary in Japan. This table has 2 columns; *mention* and *concept*. Values of mention column are terms that occurred in the radiology reports, and they were collected from the outputs of our entity extraction model (As we described in 2.2.2). Value of concept column is the standard form of its mention column.

3. Results

3.1. Entity recognition

Five thousand reports were divided into 4,000 for training, 500 for development, and 500 for validation. Hyper-parameters were tuned by using a development dataset. We used pre-trained word vectors to help the learning process. Pre-trained word vectors were obtained using Contiguous Bag of Words (CBOW) [14] from 317,130 unlabeled radiology reports. We used the RMSProp algorithm [15] with a batch size of 32. A 0.5 dropout rate was applied to avoid overfitting.

Our experimental results are shown in Table 1. Our model achieves 0.938 in the F1-score, which shows that the model can accurately extract target entities.

Table 1. Performance metrics of each semantic class

Semantic classes	Precision	Recall	F1-score	No. of entities
Body Location Modifier	0.953	0.968	0.960	4,024
Body Location	0.952	0.959	0.955	7,205
Clinical Finding Positive	0.904	0.920	0.912	7,574
Clinical Finding Suspicious	0.874	0.902	0.888	1,672
Clinical Finding Negative	0.992	0.932	0.961	2,818
Total	0.937	0.941	0.938	23,297

3.2. Structuring

We manually created 453 records from 200 reports. We evaluated the accuracy of the structuring process assuming that all entities were correctly extracted. Our structuring processer was able to correctly transform 436 records (Accuracy is 0.96).

3.3. Normalization

We created a concept table for normalization. This table was created by some medical experts (clinicians and radiological technologists). We created 21 concept terms for BL by dividing chest X-ray images. And, the concept for CF was decided by picking up frequently occurring words that were extracted as CF entity. We then combined similar concepts into one concept term. Finally, we defined 121 concept terms.

4. Discussion

In Table 1, we show the performance metrics of each semantic class. The F1-score of semantic classes exceeded 0.9, excluding CFS. The lower F1-score in CFS was related to the number of entities in the training dataset. Table 1 shows the number of entities in the CFS group was lower in comparison to other semantic classes.

As we described in 3.2, the simple rule-based method is applicable for structuring reports. This result may be due to the high similarity of the writing style of the reports.

The present study was associated with some limitations. First, some important information could not be covered in our semantic classes. For example, the size and shape of the lesions were not included, even though they are valuable for clinical research. Second, we only used an in-house dataset, which means that the performance of our method might differ if a dataset from a different hospital was used. Third, although our concept terms of BL and CF were 21 and 121, respectively, the granularity is not general and may not be appropriate for some secondary uses.

5. Conclusion

In this paper, we built an end-to-end method for transforming unstructured radiology reports into a structured format. We combined machine learning for entity recognition with a rule-based method for structuring and normalization to pursue a better approach. While our method leaves room for improvement of versatility, the present study shows the way to facilitate secondary use of radiology reports.

Acknowledgments

This work was supported by Council for Science, Technology and Innovation (CSTI), cross-ministerial Strategic Innovation Promotion Program (SIP), "Innovative AI Hospital System" (Funding Agency: National Institute of Biomedical Innovation, Health and Nutrition (NIBIOHN)).

References

[1] J.M.L. Bosmans, J.J. Weyler, P.M. Parizel. Structure and content of radiology reports, a quantitative and qualitative study in eight medical centers. *European Journal of Radiology*. **72** (2009),354–8.
[2] European Society of Radiology (ESR). ESR paper on structured reporting in radiology. *Insights Imaging*. **9** (2018),1–7.
[3] E. Pons, L.M.M. Braun, M.G.M. Hunink, J.A. Kors. Natural Language Processing in Radiology: A Systematic Review. *Radiology*. **279** (2016),329–43.
[4] T. Imai, Y. Onogi. Extracting numeric measurements and temporal coordinates from Japanese radiological reports. *Proceedings of SPIE - The International Society for Optical Engineering*. (2004),268–76.
[5] T. Imai, E. Aramaki, M. Kajino, K. Miyo, Y. Onogi, K. Ohe. Finding malignant findings from radiological reports using medical attributes and syntactic information. *Stud Health Technol Inform*. **129** (2007),540–4.
[6] S. Cornegruta, R. Bakewell, S. Withey, G. Montana. Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. *Proceedings ofthe Seventh International Workshop on Health Text Mining and Information Analysis (LOUHI)*. (2016),17–27.
[7] C. Wang, A. Elazab, J. Wu, Q. Hu. Lung nodule classification using deep feature fusion in chest radiography. *Computerized Medical Imaging and Graphics*. **57** (2017),10–8.
[8] B. Reiner, E. Siegel. Radiology reporting: Returning to our image-centric roots. *American Journal of Roentgenology*. **187** (2006),1151–5.
[9] T. Kudo. MeCab. http://taku910.github.io/mecab/.
[10] S. Hassanpour, C.P. Langlotz. Information extraction from multi-institutional radiology reports. *Artif Intell Med*. **66** (2016),29–39.
[11] X. Ma, E. Hovy. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. (2016),1064–74.
[12] A. Graves, J. Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*. **18** (2005),602–10.
[13] J. Lafferty, A. Mccallum, F.C.N. Pereira, F. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *Proceeding of ICML*. (2001).
[14] T. Mikolov, K. Chen, G. Corrado, J. Dean. Efficient Estimation of Word Representations in Vector Space. *ICLR*. (2013).
[15] Y.N. Dauphin, H. de Vries, Y. Bengio. Equilibrated adaptive learning rates for non-convex optimization. *Proceeding NIPS'15 Proceedings of the 28th International Conference on Neural Information Processing Systems*. (2015),1504–12.