NLP Automation to Read Radiological Reports to Detect the Stage of Cancer Among Lung Cancer Patients

Er. Khushbu Gupta
Section of Clinical Epidemilogy and Biostatistics Faculty of Medicine, Ramathibodi Hospital Bangkok, Thailand
`khusbu.gup@student.mahidol.edu`

Ratchainant Thammasudjarit, Ph.D.
Section of Clinical Epidemilogy and Biostatistics Faculty of Medicine, Ramathibodi Hospital Bangkok, Thailand
`ratchainant.tha@mahidol.edu`

Prof.Ammarin Thakkinstian, Ph.D.
Section of Clinical Epidemilogy and Biostatistics Faculty of Medicine, Ramathibodi Hospital Bangkok, Thailand
`ammarin.tha@mahidol.edu`

Abstract

A common challenge in the healthcare industry today is physicians have access to massive amounts of healthcare data but have little time and no appropriate tools. For instance, the risk prediction model generated by logistic regression could predict the probability of diseases occurrence and thus prioritizing patients' waiting list for further investigations. However, many medical reports available in current clinical practice system are not yet ready for analysis using either statistics or machine learning as they are in unstructured text format. The complexity of medical information makes the annotation or validation of data very challenging and thus acts as a bottleneck to apply machine learning techniques in medical data. This study was therefore conducted to create such annotations automatically where the computer can read radiological reports for oncologists and mark the staging of lung cancer. This staging information was obtained using the rule-based method implemented using the standards of Tumor Node Metastasis (TNM) staging along with deep learning technology called Long Short Term Memory (LSTM) to extract clinical information from the Computed Tomography (CT) text report. The empirical experiment shows promising results with the accuracy up to 85%.

1 Introduction

It is not a secret that the world is facing lack of health care professionals. According to a study in 27 countries, there is an extreme shortage of oncologists, with a ratio of 1 oncologist per more than 1000 incident cancers. (Roxanne Nelson 2018) The number of radiologists serving per million populations is no better.(Nakajima et al. 2008) With such circumstances, health care industry is in the absolute need of innovative automations. However, the health care issues are very sensitive and require doctor's personal touch in every phase. Furthermore, a number of clinical researches could be conducted that could lead to many improvements of health care industry. Unfortunately, many medical reports available in clinical settings are in unstructured format and not yet ready for statistical analysis. (Sun et al. 2018)

While Machine learning (ML) algorithms are powerful to contribute in such circumstances, ML itself has no utility unless it is fed with the annotated datasets. (Young et al. 2018), (Goodfellow, Bengio and Courville 2016) The health practitioners do not have sufficient time to annotate the data to implement ML algorithms on it. We therefore conducted a study that could generate such labels using the standard guidelines of the literature to classify a report into the exact stage of lung cancer according to TNM system. (Detterbeck 2018) These labels however require verification from the domain experts.

For this research, we could obtain only a small quantity of verified reports. We used deep learning model as LSTM to train other similar data. (Zhang et al. 2018) (Bengio, Simard and Frasconi 1994) The empirical experiment showed promising results, although we are yet to get full verifications from the radiologists.

2 Methods

Data Acquisition was done under the considerations of Ramathibodi Ethical Committee. The Department of Radiology, Ramathibodi Hospital, Mahidol University provided us with 2000 CT-scan reports of lung cancer patients during the time of 2011 to 2018. Patients were eligible if they met the following criteria: Adult patient's aged 18 years or older who were diagnosed as early stage lung cancer and had CT axial size of slide 2-3mm. Patients were excluded if their CT reports were incomplete. Although, we could obtain 2000 CT scan reports, the data did not have TNM labelling. We therefore used 23 reports annotated by single annotator who is a medical doctor for this preliminary experiment.

While exploring the datasets, we found 6 unique labels of TNM staging which contained three stages of Tumor (T1, T2, T3), three stages of Node (N0, N1, N2) and two stages of Metastasis (M0, M1). Also, we found that the growth of unique words converges at the eighth report up to 436 words. Such convergence reflects the nature of word distribution in the context specific reports. Therefore, in the absence of sufficient labelled reports we implemented the automatic annotation using TNM guidelines.

The proposed system took raw text as input and returned with the staging information of lung cancer as the primary outcome. The process of the proposed engine was as follows: **Raw Text (Radiological report), Tokenization, Stop word removal, Clinical Information Extraction, Structured Data, Verification with manual rules of TNM staging, Get Staging Labels.**

Given a CT text report, the first process is the data preprocessing to extract the raw text into sentences and then words respectively. Then we remove stop words. The next process is the clinical information extraction where we apply manually crafted rules on the availability of the keywords. For example, "Increased size of two ground-glass nodules at apicoposterior segment of the LUL (SE 3; IM 37), and superior segment of the LLL (SE 3; IM 58), now measuring about 1 cm (previously size 0.4 cm), and 1.1 cm (previously size 0.7 cm)." First we look for the keyword "nodule". If it is found, we look for other predefined keywords such as "cm, mm", "LUL" etc. We thus obtain structured data as shown in the table below:

Table 1. Example of structured data

Nodule size	Location of Nodule	Number of nodule	Nature of nodule	Change in the nodule
1 cm	apicoposterior segment of the LUL	1	Ground-glass	Increased by 0.6 cm
1.1 cm	superior segment of the LLL	1	Ground-glass	Increased by 0.4 cm

Once we obtained such structured data, we applied rules based on the TNM staging guidelines. For instance, The TNM guidelines state that if the size of the tumor is more than 1 cm and less than 2 cm- it is classified under T1b. There are similar rules for N and M, from which we obtain TNM labelling.

Secondly, we also used LSTM network for the same task and compared the results. For which the process was as follows: **Define Network, Compile Network, Fit Network, Make Predictions, Evaluate Network.**

We used Keras library for this purpose. First, we defined LSTM network as a sequence of layers defining model as sequential. We tokenized the text using one hot encoding and also used pre-trained word embedding called glove containing 100 –dimensional embedding vectors for 400,000 words. For the embedding layer, we defined the maximum features as 500 and the embedding dimensionality as

100. The reason behind choosing 500 as number of possible tokens was the convergence of growth of unique words at eighth report with 436 words. Then we added LSTM layers followed by fully connected layer called dense for the prediction. While defining the model we used 32 dimensions of 'Relu activation function' and use 6 for the 'sigmoid' activation function as we have 6 unique labels. Additionally, we froze the embedding layer while loading the glove matrix into the Embedding layer so that it didn't get updated during the training. Then we compiled the network which transformed the layers into series of matrix. Here we defined parameters such as the optimization algorithm as 'rmsprop' and the loss function as 'binary-crossentropy'. We then fitted the data to adapt the weights on a training dataset. The network was then trained with back propagation where the data was partitioned in groups of input output patterns called batches in each epoch. We split 70% of data for the training purpose and 30% for the validation phase. The next process was to make predictions and evaluate the trained model. The result thus obtained was then validated with the actual labels.

3   Evaluation

To conduct this experiment, although we were able to obtain larger number of reports (2000), we could only use 23 reports annotated by a medical doctor. Hence we implemented an automatic rule-based annotation system using TNM guidelines. However, the accuracy of the system thus built was not that great. Therefore, we used LSTM whose accuracy was much better than the heuristic approach. Following are the results of comparison between the two methods approached in the research.

Based on the heuristic approach, accuracy of the system was achieved up to 59% for T-staging, 36% for N-staging and 41% for M-staging. While performing error analysis, it was found that the 41% of error in T-staging was due to the length of tumour which was not mentioned in the report. According to the TNM guidelines, when the length of the tumour is not mentioned the invasion of tumour from one portion within the lung or outside is taken into consideration. Meanwhile, heuristic approach could detect the location of the nodule, but it was not perfect due to the inability in negation detection in most cases and invasion detection was not great either. Similarly, for N and M-staging results were highly dependent on the location and invasion of nodules and lymph nodes which was pointed out to be the weak point in the heuristic approach. Whereas while implementing LSTM, the system could bring up to 85% of accuracy for the collective TNM staging.

We are aware that the conclusions made on this small dataset may not be reliable. At the same time getting annotations from the domain expert for the larger dataset is itself a challenge. Therefore, we are in the process of getting TNM annotations from the cancer registry department of the same hospital. We can then reuse this experiment on the larger datasets and verify our results from them.

4   Conclusion

The automated engine developed as clinical information extraction tool to classify lung cancer staging is useful in clinical settings. This work is focused in lung cancer which is one of the top causes of death worldwide. This initial prototype has shown promising results with the accuracy up to 85%. There is tremendous information hidden in other Electronic Health Records. This work can be replicated in other domains like brain, heart etcetera. Automatic detection of lung cancer staging can streamline the process of diagnosis and help in designing the treatment regarding prognosis and add up to help the limited workforce of oncologists.

References

Bengio, Y., P. Simard & P. Frasconi (1994) Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks,* 5**,** 157-166.

Cheng, L. T. E., J. Zheng, G. K. Savova & B. J. Erickson (2010) Discerning Tumor Status from Unstructured MRI Reports—Completeness of Information in Existing Reports and Utility of Automated Natural Language Processing. *Journal of Digital Imaging: the official journal of the Society for Computer Applications in Radiology,* 23**,** 119-132.

Detterbeck, F. C. (2018) The eighth edition TNM stage classification for lung cancer: What does it mean on main street? *The Journal of Thoracic and Cardiovascular Surgery,* 155**,** 356-359.

Goodfellow, I., Y. Bengio & A. Courville (2016) Deep Learning| The MIT Press. *Cambridge, Massachusetts*.

Graves, A., S. Fernández & J. Schmidhuber. 2005. Bidirectional LSTM networks for improved phoneme classification and recognition. In *International Conference on Artificial Neural Networks*, 799-804. Springer.

Nakajima, Y., K. Yamada, K. Imamura & K. Kobayashi (2008) Radiologist supply and workload: international comparison--Working Group of Japanese College of Radiology. *Radiat Med,* 26**,** 455-65.

Roxanne Nelson, R., BSN (2018) Profound Lack' of Oncologist's in Many Parts of the World.

Saleem, S., Y. Sabri & A. Saeed. 2009. Radiology Education in the Faculty of Medicine at Cairo University (Kasr Al-Ainy Hospital). In *Radiology Education*, 283-294. Springer.

Sun, W., Z. Cai, Y. Li, F. Liu, S. Fang & G. Wang (2018) Data Processing and Text Mining Technologies on Electronic Medical Records: A Review. *Journal of Healthcare Engineering,* 2018**,** 9.

Yang, W., J. H. Williams, P. F. Hogan, S. S. Bruinooge, G. I. Rodriguez, M. P. Kosty, D. F. Bajorin, A. Hanley, A. Muchow, N. McMillan & M. Goldstein (2014) Projected supply of and demand for oncologists and radiation oncologists through 2025: an aging, better-insured population will result in shortage. *J Oncol Pract,* 10**,** 39-45.

Young, T., D. Hazarika, S. Poria & E. Cambria (2018) Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine,* 13**,** 55-75.

Zhang, Y., D. Y. Ding, T. Qian, C. D. Manning & C. P. Langlotz (2018) Learning to summarize radiology findings. *arXiv preprint arXiv:1809.04698*.