Information Extraction from Clinical Practice Guidelines: A Step Towards Guidelines Adherence

Musarrat Hussain and Sungyoung Lee*

Department of Computer Science & Engineering, Kyung Hee University, South Korea
{musarrat.hussain, sylee}@oslab.khu.ac.kr

Abstract. Clinical Practice Guidelines (CPGs) are an essential resource for standardization and dissemination of medical knowledge. Adherence to these guidelines at the point of care or by the Clinical Decision Support System (CDSS) can greatly enhance the healthcare quality and reduce practice variations. However, CPG adherence is greatly impeded due to the variety of information held by these lengthy and difficult to parse text documents. In this research, we propose a mechanism for extracting meaningful information from CPGs, by transforming it into a structured format and training machine learning models including Naïve Bayes, Generalized Linear Model, Deep Learning, Decision Tree, Random Forest, and Ensemble Learner on that structured formatted data. Application of our proposed technique with the aforementioned models on Rhinosinusitis and Hypertension guidelines achieved an accuracy of 82.10%, 74.40%, 66.70%, 66.79%, 74.40%, and 83.94% respectively. Our proposed solution is not only able to reduce the processing time of CPGs but is equally beneficial to be used as a preprocessing step for other applications utilizing CPGs.

Keywords: Information Extraction, Clinical Practice Guidelines Processing, Guidelines Adherence, Text Processing.

1 Introduction

Clinical Practice Guidelines (CPGs) are "systematically developed statements to assist practitioners and patient decision about appropriate healthcare for specific circumstances" [1]. It has an indispensable role in disseminating medical knowledge, enhancing healthcare quality, reducing cost, and decreasing practice variations. CPGs aims to help healthcare providers and patients to make the best decision about treatment for a particular condition, by picking most suitable strategies in a specific clinical situation [2]. It can either be used by healthcare providers or can be transformed to machine interpretable format to be part of the Clinical Decision Support System (CDSS) to support clinicians at the point of care.

Despite the valuable goal and importance, the adherence rate of CPGs varies between 20% to 100% depending upon clinical scenario and the nature of the CPG [3]. The main hurdle in the adherence of CPGs is the current format (unstructured document) of the CPG and clinician/healthcare provider unawareness about CPGs. Most of the healthcare providers are unaware about the existence of CPG and they face difficulties in understanding on directing them toward a specific CPG [4].

There are many other obstructs in adherence to CPGs related to clinicians, patients behaviors and CPGs itself [4]. Besides other obstacles, one of the major hurdles in adherence to CPG is the format of the CPGs. Most of the CPGs are published online in medical journals having an unstructured format. CPG contents based on proximity to technical solution can be categorized into two parts: background information and disease specific information. Background information includes anecdotes and thoughts of the author. While more concrete information, relating to causes, consequences, and actions form the disease-specific information. Due to the wide range of variation in the nature of CPGs, it is very important to understand each CPG before transforming it into a machine interpretable format. It requires a lot of time to locate scenario/disease specific information at a limited time during real practice. Therefore, clinicians avoid utilizing CPGs during real practice. Also, it creates difficulties during the conversion of CPGs to computer interpretable format. The adherence rate can be increased by finding a mechanism that can extract relevant information from the CPG and filter out irrelevant information.

The primary goal of this research is to find and extract relevant information also called recommendation statements from CPGs and filter out background information. To achieve this goal, we transform hypertension [5] and Rhinosinusitis [6] CPG text to structured format (word vector) and train machine learning models including Naïve Bayes, Generalized Linear Model, Deep Learning, Decision Tree, Random Forest, and Ensemble Learner. The trained models achieved accuracy of 82.10%, 74.40%, 66.70%, 66.79%, 74.40%, and 83.94% respectively on extracting recommendation statements. This technique has twofold advantages. It can extract disease specific information for a clinician in real time at the point of care. It can also be used as a preprocessing step for CPGs transformation to computer interpretable format, which can increase CPG adherence, improve the healthcare quality and can eventually reduce healthcare cost.

2 Related Work

The history of CPGs started in late 1970 by the National Institute of Health Consensus Development Program. The objective of the program was to improve healthcare quality by identifying and adopting best practices [7].

M Yetisgen-Yildiz et al. [8] proposed a text processing pipeline that can identify and extract recommendation sentences from radiology reports using Natural Language Processing (NLP) and supervised text classification techniques. In the pipeline, the task performed in the sequence involves section segmentation, sentence segmentation, and sentence classification. The author used 800 anonymized radiology reports for training the machine learning model which was annotated by radiologist and internal medicine physician independently. The trained model was then used to classify the new input sentence into a recommendation or non-recommendation sentence. However, the nature of radiology reports is completely different from that of CPGs.

Radiology report has semi-structured nature while CPGs are completely unstructured documents. These reports are divided into different sections specified for one type of information. The recommendation sentences are written in *Impression Sections* of the report. And also all the reports follow the same structure, therefore, a model can easily be trained on these document.

A. Khalifa et al. [9] proposed a mechanism that detects cardiovascular risk factors in clinical notes of diabetes patients by using the existing NLP techniques and tools. The risk factors include high blood pressure, high cholesterol levels, obesity, and smoking status. They used existing tools: Apache UIMA Textractor and cTAKES for text preprocessing and the risk factors identification using a regular expression. They considered smoking status risk factor in the study. However, the mechanism is effectively identifying risk factors in the cardiovascular domain. Each domain has different risk factors and a generalized solution is required that can identify and extract the required information from all types of documents.

S. Priyanta et al. [10] did the comparative analysis of machine learning and rule-based models for sentence subjectivity classification. The rule was generated using opinion patterns. The experiment was performed on classifying Indonesian news sentences into two categories subjective and objective. The machine learning model used for the classification and comparison were Multinomial Support Vector Machine (SVM) and Nave based classifier (NBC). The experiment was performed on 2550 document, containing 46393 sentences. The experimental result showed that the rule-based classifier outperformed by achieving 80.36% accuracy as compared to SVM 74% and NBC 71%.

H. Hematialam et al. [11] proposed an automatic technique of finding and extracting recommendation statements in CPGs. The authors used a supervised machine learning model (Nave Bayes, J48, and Random Forest) that classify CPG sentences into three categories: NC (no condition), CA (condition-action), and CC (condition consequence). The domain expert annotated three type of guidelines (hypertension, chapter 4 of asthma, and rhinosinusitis) and the authors used these guidelines as a training set for training machine learning model. The authors used Part of Speech (POS) tags as a feature to make the model more domain independent. Each action-condition statement has a modifier, and the most used modifiers in the CPGs used by authors in their study were "If", "in", "for", "to", "which", and "when". The candidate statements were found by using regular expressions. The identified candidate statements were transformed/paraphrased to "if condition, then consequences" format for rules generation.

In our previous study [12] we used a semi-automatic technique to extract recommendation statements from CPG documents. We manually analyze the same annotated hypertension guideline [5] as used in this study and extract heuristic patterns that can recognize recommendation statements in the unseen CPGs. The identified patterns were able to extract recommendation statements with 85.54% accuracy. However, the limitation of that work is the manual effort required for pattern extraction. Also, the extracted heuristic patterns depend on the contents of CPGs which may vary in different CGPs format. Therefore, the heuristic patterns may not be able to perform well on all CPGs. To overcome the manual effort, in this research, we used an NLP pipeline that can extract recommendation statements with satisfactory performance.

3 Proposed Methodology

The primary focus of this research is to automatically and efficiently extract recommendation statements from a CPG and filter out background information using state-of-the-art machine learning algorithms. To achieve this goal, we devised an NLP pipeline as depicted in Fig. 1. The proposed pipeline accomplished the aforementioned goal in two major steps, it transforms a CPG into a structured format (Word Vector) and then trained an ensemble learning model that uses the base classifier including Naïve Base, Generalized Liner Model, Random Forest, Deep Learning, and Decision Tree on the generated structured document. The trained model is then used to classify unseen CPG statements to recommendation statement (RS) or non-recommendation statement (NRS). The detail of the process is given in the following sections.

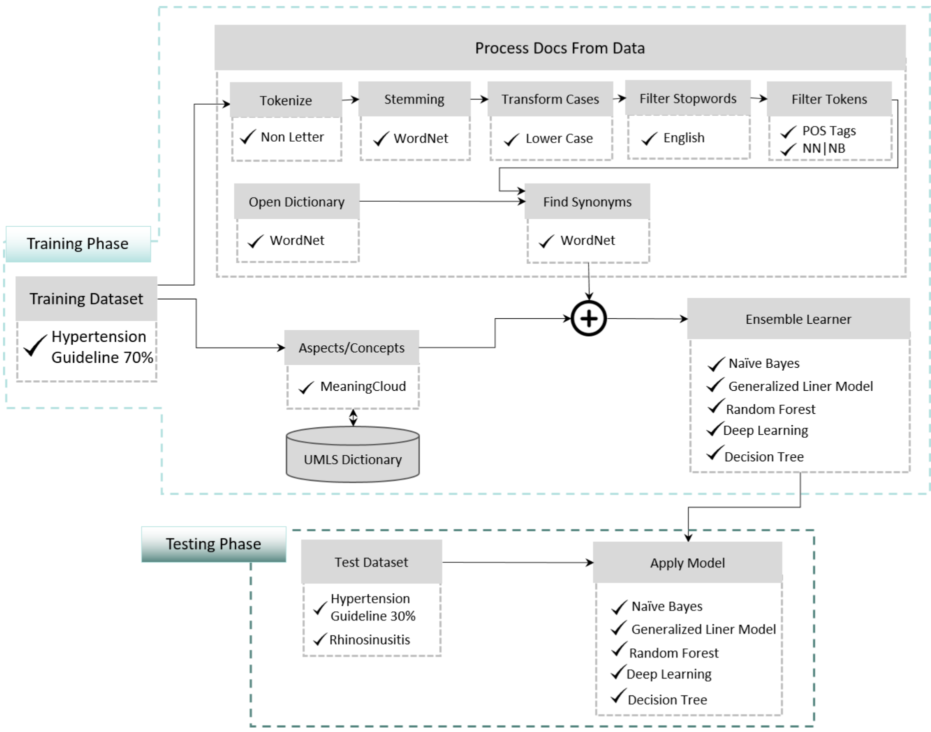

Fig. 1. Recommendation statements extraction pipeline.

3.1 Unstructured to Structured Conversion

In order to transform a given unstructured CPG to a structured format (word vector) multiple steps are designed in sequence.

Initially, the CPG sentences along with label are loaded to working space. We used the Term Frequency-Inverse Document Frequency (TF-IDF) scheme for the creation of word vector. The *Process Docs From Data* consist of others operators ranging from tokenization to synonyms identification. The token operator split the input text into tokens based on word spacing scheme. The words of each token are then transformed to its base format using WordNet stemming followed by *Transform Case* which converts all tokens to its lower case to maintain symmetry. Some of the word tokens despite maximum usage in the document may have limited impact known as stop words removed by *Filter Stopwords*. We applied the Part-of-Speech (POS) using PENN Tree Scheme, employed pattern (NN|NB) to filter the names and verbs used in the input text. As we notice that the CPG recommendation statements mostly consists of disease/medicine name and action on them. We also employed the word expansion mechanism to make the word vector more comprehensive for effective classification. For word expansions, we added synonyms component to the pipeline. We used WordNet dictionary for synonyms identification.

3.2 Aspects/Concepts Extraction

We used MeaningCloud services to find and extract aspects of the input text. We created the local copy of Unified Medical Language System (UMLS) dictionary at MeaningCloud and then used the APIs services for the aspects/concepts extraction based on the created dictionary. The aspects/concepts addition to the data increased the performance of basic classifier as well as ensemble learning base classification as discussed in the result section.

The final outcome of this process is a structured data (word vector) consists of all tokens of interest along with synonyms and their aspects/concepts. This document/structured data will be then used in the following section for training the machine learning algorithms.

3.3 Ensemble Learner

Ensemble learning combines and applies multiple models on the same instance of data to accurately predict the class label for this instance to reach the final conclusion. The algorithm considered in this study includes Naïve Bayes, Generalized Liner Model, Random Forest, Deep Learning, and Decision Tree. The majority voting technique was used to get the final decision. In this technique, the computed results of each algorithm are analyzed in order to determine the final class recommended by most of the algorithms.

4 Result and Discussion

We performed multiple experiments with different settings on annotated hypertension CPG [5] consists of 78 recommendation statements among total 278 statements. In this study, the CPG statements annotated as CA, CC, or A are considered as RS statements. The CPG was split into 70% and 30% for training and testing part. The training part of the CPG consist of total 195 statements including 58 recommendation statements. While the testing part consists of total 83 statements including 20 recommendation statements. The trained models were also validated on another CPG (Rhinosinusitis [6]) to authenticate the performance (in term of accuracy) of the models.

The experiment that outperformed among others achieved the best accuracy of 79.82% by Ensemble Learner algorithm as shown in Fig. 2. In this experiment, we used TF-IDF for word vector generation, Non Letters for tokenization, WordNet for stemming and English stopwords were filtered out. In the filter tokens component, we observed from multiple experiments that the NN and NB tokens have the maximum contribution in achieving the accurate result. Therefore, we filter out all other tokens. We find the synonyms of the remaining token using WordNet dictionary. However, in this experiment the aspects/concepts for input dataset were not included.

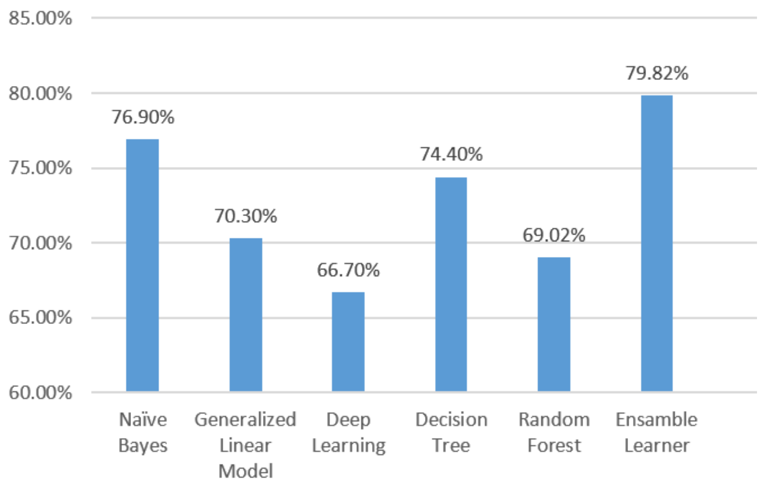

Fig. 2. Results achieved by each algorithm without aspects.

The experiment was repeated with the same setting as earlier but we also find and include aspects because clinical guidelines describe clinical scenarios and normally uses clinical terminology specific to a target disease. To extend the scope of the mechanism to be applicable on any CPG irrespective of the target disease we find the category (Aspect/Concepts) by utilizing UMLS medical dictionary. The final structured data generated is consists of word tokens, their synonyms and their aspect along with occurrence frequency.

We trained and tested machine learning models on aforementioned two CPGs. The models considered for the study includes Naïve Bayes, Generalized Liner Model, Deep Learning, Decision Tree, Random Forest, and Ensemble Learner as shown in Fig. 3. The models achieved 82.10%, 74.40%, 66.70%, 66.70%, 74.40%, and 83.94% accuracies respectively.

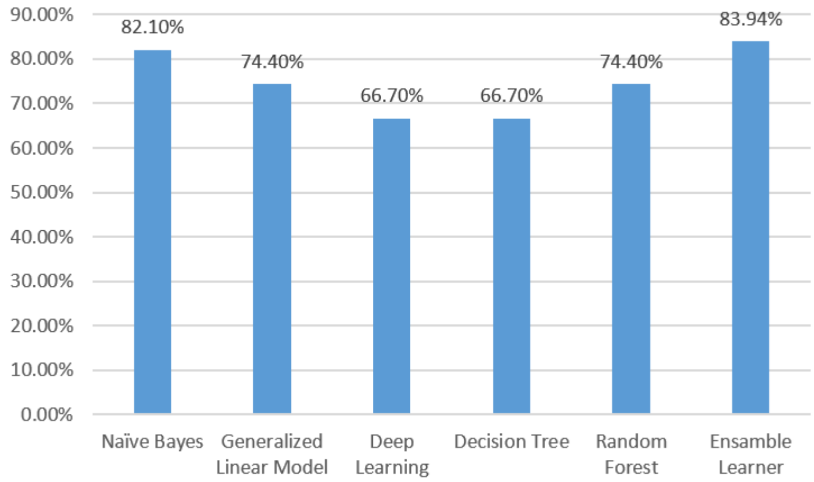

Fig. 3. Results achieved by each algorithm with aspects.

5 Conclusion and Future Work

Clinical practice guidelines have the potential to overcome all deficiencies of healthcare. However, due to the nature and format of guidelines, it has lower adherence rate and faces difficulties in achieving this goal. Some of the deficiencies can be reduced by filtering out irrelevant information from CPG and provide disease-specific information at the point of care. This paper focuses on the transformation of the CPG to structured format (word vector) and uses machine learning models to filter out irrelevant information from CPG. This mechanism can provide two-fold benefits. First, it can be used for filtering out irrelevant information from guidelines. Which will increase the effectiveness of guidelines, improve healthcare quality, help in providing evidence-based practice, and reduce processing time for identifying disease-specific information. Secondly, it can be used as a preprocessing step for other text mining applications.

In the future, we are planning to extend the existing work to further improve accuracy. We are also working on a system that can generate a machine interpretable CPG which will play an imperative role in CPG adherence by integrating it into CDDS system. The current proposed work will serve as an essential part of the conversion process.

Acknowledgement

This research was supported by an Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No. 2017-0-00655). This work was supported by the Ministry of Science and ICT (MSIT), Korea, under the Information Technology Research Center (ITRC) support program

(IITP-2017-0-01629) supervised by the Institute for Information & communications Technology Promotion (IITP) and NRF- 2016K1A3A7A03951968

References

1. Lohr, K.N., Field, M.J.: Clinical practice guidelines: directions for a new program. National Academies Press (1990).
2. Wenzina, R., Kaiser, K.: Identifying condition-action sentences using a heuristic-based information extraction method. In: Process Support and Knowledge Representation in Health Care. pp. 26–38. Springer (2013).
3. Rello, J., Lorente, C., Bodí, M., Diaz, E., Ricart, M., Kollef, M.H.: Why do physicians not follow evidence-based guidelines for preventing ventilator-associated pneumonia?: a survey based on the opinions of an international panel of intensivists. Chest. 122, 656–661 (2002).
4. Kilsdonk, E., Peute, L.W., Riezebos, R.J., Kremer, L.C., Jaspers, M.W.: From an expert-driven paper guideline to a user-centred decision support system: A usability comparison study. Artif. Intell. Med. 59, 5–13 (2013).
5. James, P.A., Oparil, S., Carter, B.L., Cushman, W.C., Dennison-Himmelfarb, C., Handler, J., Lackland, D.T., LeFevre, M.L., MacKenzie, T.D., Ogedegbe, O.: 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). Jama. 311, 507–520 (2014).
6. Chow, A.W., Benninger, M.S., Brook, I., Brozek, J.L., Goldstein, E.J., Hicks, L.A., Pankey, G.A., Seleznick, M., Volturo, G., Wald, E.R.: IDSA clinical practice guideline for acute bacterial rhinosinusitis in children and adults. Clin. Infect. Dis. 54, e72–e112 (2012).
7. Jacobsen, P.B.: Clinical practice guidelines for the psychosocial care of cancer survivors. Cancer. 115, 4419–4429 (2009).
8. Yetisgen-Yildiz, M., Gunn, M.L., Xia, F., Payne, T.H.: A text processing pipeline to extract recommendations from radiology reports. J. Biomed. Inform. 46, 354–362 (2013).
9. Khalifa, A., Meystre, S.: Adapting existing natural language processing resources for cardiovascular risk factors identification in clinical notes. J. Biomed. Inform. 58, S128–S132 (2015).
10. Priyanta, S., Hartati, S., Harjoko, A., Wardoyo, R.: Comparison of Sentence Subjectivity Classification Methods in Indonesian News. Int. J. Comput. Sci. Inf. Secur. 14, 407 (2016).
11. Hematialam, H., Zadrozny, W.: Identifying Condition-Action Statements in Medical Guidelines Using Domain-Independent Features. ArXiv Prepr. ArXiv170604206. (2017).
12. Hussain, M., Hussain, J., Sadiq, M., Hassan, A.U., Lee, S.: Recommendation Statements Identification in Clinical Practice Guidelines Using Heuristic Patterns. In: 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). pp. 152–156. IEEE (2018).