Poster: Medical Concept Extraction using Smartphone and Natural Language Processing Techniques

Asim Abbas
Kyung Hee University
Yongin-si, Gyeonggi-do, Republic of Korea
asimabbasturi@oslab.khu.ac.kr

Muhammad Zaki Ansaar
Kyung Hee University
Yongin-si, Gyeonggi-do, Republic of Korea
zakiansaar@oslab.khu.ac.kr

Sungyoung Lee
Kyung Hee University
Yongin-si, Gyeonggi-do, Republic of Korea
sylee@oslab.khu.ac.kr

ABSTRACT

Over the past few decade, smartphone technology has played a vital role in the healthcare domain. In this paper, we proposed a methodology to allow end users to automatically process the medical text image using a camera or enter text manually. Then the system output allows to extract medical concepts, its semantic type and entity type from medical text apply Natural Processing Language techniques from UMLS medical dictionary. The medical text can be a health report, a clinical case, or other kinds of a text containing medically related information. The aim of this kind of methodology of the mobile application is to contribute in the area of natural language processing (NLP), intelligent system and to increase the medical students and bioinformatics researchers interest to quickly accessing information about medical data.

CCS CONCEPTS

• **Information systems** → *Mobile information processing systems*;

KEYWORDS

Information Extraction; UMLS; NLP; Intelligent System

1 INTRODUCTION

The smartphone application is gaining popularity rapidly in the healthcare domain to provide potential and quick information. Mostly data present in the smartphone is text-based, some natural language processing techniques are used to extract useful information from text. Many applications have been developed and published to the smartphone application market related to healthcare information extraction such as M-Health [1] and search system for medicine information [5]. The goal of this kind of applications is to provide a piece of quick medical information related to medicine

prescription, diagnosed diseases, and name of medicine to the medical staff and students. In this paper, we proposed a methodology to automatically extract medical information from medical text image using a smartphone camera or users manually text enter. UMLS medical dictionary is used, follow exact matching approach to find medical information. Some image processing techniques are used to extract the text from the image and then applied natural language processing (NLP) techniques to process the unstructured text data to structure format.

According to the literature review, a researchers are developing new tools, methodology and algorithm to extract medical concepts from an unstructured medical text. A smartphone application has been developed for medical students. The aims of this application to provide content and features that related to medical education and provide a modern education-learning environment to the students [2]. A dialog based medicine information application has been developed. In this application user question-answering about medicines subscription for a specific disease like influenza using information extraction and NLP techniques [5]. An educational smartphone application has been developed for a patient with coronary artery disease (CAD). This application performs as an educational learning instrument for CAD patient and to assess the patient level of satisfaction [3].

2 PROPOSED METHODOLOGY

The proposed methodology based on two modules sees Fig.1. The first module-1 is responsible for extracting text from the medical text image and the second module-2 is responsible for applied natural language processing techniques on unstructured text to extract medical concepts from the UMLS dictionary. Each module has discussed below.

2.1 Text extraction form text image (Module-1)

2.1.1 Extracting text from Image. The user is allowed to take a picture of a medical text image. The image then sent for pre-processing. The processed image goes through the Optical Character Recognition (OCR) program. This latter puts the extracted text in a text file.

2.1.2 Process User manually entered text. The user is also allowed to enter text manually into the text file of the application. User can obtain result quickly by entered text manually instead of text image. In Fig.1 module-2, techniques will directly be applied on a user manually entered text to convert text into a structured format from unstructured format.
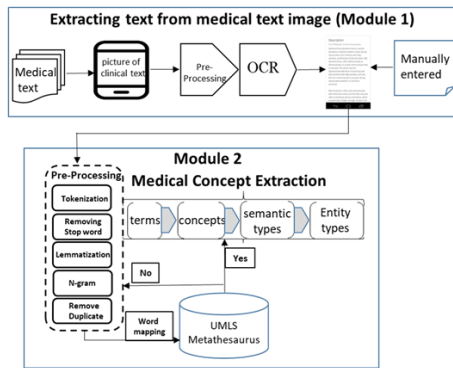
Figure 1: Medical Concept Extraction using smart phone

2.2 Medical Concept extraction(Module-2)

We looked at 2018AB version of UMLS metathesaurus containing over 1 million concepts and exactly matched medical terms approach have followed. UMLS Terminology Services (UTS) provide restful web services to extract medical concepts from UMLS metathesaurus. The processes performed in this module have discussed below.

2.2.1 Pre-Processing. Data pre-processing will apply on text received from module-1 see Fig.1 to convert the unstructured format of text into structured. Data pre-processing step includes tokenization, removing stop words, lemmatization, n-gram and remove duplicate words.

2.2.2 Finding Terms. After getting words from pre-processing, exact word matching approach has followed to the UMLS see Fig1. If word match to the UMLS, it will be categorized as a term else word will be excluded.

2.2.3 Concept Extraction. A concept is the meaning of medical terms and each meaning contain different names. UMLS contains over one million concepts [4]. In the proposed methodology, we are extracting a list of concept for each term. Each term contains one or more than one concept.

2.2.4 Semantic Type Extraction. Another knowledge source of UMLS is Semantic Network, contains 135 semantic types (STY's) and 54 semantic relationships for classifying and categorizing all concepts represented in the UMLS metathesaurus [1]. Each concept contains at least one or greater than one semantic types, few concepts assigned five semantic types.

2.2.5 Entity Type Extraction. An entity type presents the parent relation for the concept. It presents the concept in a more standard and obvious way as compared to the semantic type. It is possible that each concept has multiple semantic types but have only a single entity type.

3 RESULTS AND DISCUSSION

In Fig-2, we performed statistical analysis and test to measure the accuracy of text extraction from a text image. The accuracy was measured on ten medical narrative text images by the percentage of recognized words in the text out of the total number of words in
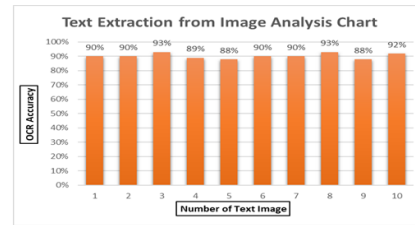

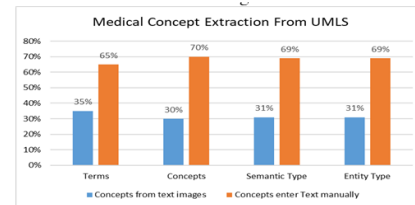Figure 2: Statistical Analysis chart for Text Extraction

Figure 3: Statistical Analysis of Medical Concept Extraction

the text. We got a 90% average accuracy for text extraction, which means 90% of words were recognized by the OCR techniques. In Fig-3, represents medical concept extracted from text image using a camera and manually entered text. We supposed 10 clinical scenarios for test and evaluation purpose. Result obtained for manually entered text was better than text image. Because many concepts have not been identified and missed using a camera and OCR Techniques.

4 CONCLUSION

In this paper, we have proposed a methodology to extract medical concept from the text image or the user manually entered text using a smartphone, applied some NLP techniques to extract medical concepts, from UMLS Metathesaurus. We obtained results for text extraction from medical text image using OCR technique. User manually enters text give a good result as compare to text image for extracting medical concepts.

ACKNOWLEDGMENTS

REFERENCES

[1] Fernando Aparicio Galisteo, Gaya López, María Cruz, Manuel De Buenaga Rodríguez, Diego Gachet Páez, et al. 2015. M-health mobile app usability tested with engineering students. (2015).
[2] Min Jung Cho, Jae Lan Sim, and Seon Young Hwang. 2014. Development of smartphone educational application for patients with coronary artery disease. *Healthcare informatics research* 20, 2 (2014), 117–124.
[3] Helena Galhardas, Vasco D Mendes, and Luısa Coheur. 2014. Medicine. Ask: a Natural Language Search System for Medicine Information. (2014).
[4] Online Learning. 2019. Online Learning. https://www.nlm.nih.gov/research/umls/new_users/online_learning/SEM_001.html
[5] Maximilian Sandholzer, Tobias Deutsch, Thomas Frese, and Alfred Winter. 2015. Predictors of students' self-reported adoption of a smartphone application for medical education in general practice. *BMC medical education* 15, 1 (2015), 91.