Archives of Advances
in Biosciences

Original Article:

Error Detection in Patients' Pharmaceutical Data: Application of Ontology-Based Text Miner

Hamid Moghaddasi[1*], Reza Rabiei[1], Sara Shadmani[1]

1. Department of Health Information Technology and Management, Faculty of Paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Cite this article as: Moghaddasi H, Rabiei R, Shadmani S. Error Detection in Patients' Pharmaceutical Data: Application of Ontology-Based Text Miner. Archives of Advances in Biosciences. 2022; 13:E37489. https://doi.org/10.22037/aab.v13i.37489

https://journals.sbmu.ac.ir/aab/article/view/37489

Article info:
Received: 27 Jan 2022
Accepted: 21 Feb 2022
Published: 10 Jun 2022

*** Corresponding author:**

Hamid Moghaddasi, PhD.

Address: Department of Health Information Technology and Management, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

E-mail: Moghaddasi@sbmu.ac.ir

Abstract

Introduction: Medication errors in patients' medical records can influence the healthcare quality and cause risks for them. It is, therefore, crucial to apply appropriate procedures to reduce these errors. This study sought to develop a software for detecting medication errors through qualitative analysis of patients' medical records.

Materials and Methods: The software was developed using object-oriented analysis and Java. The text was first pre-analyzed using a framework known as Stanford Core NLP. In the next stage, the text was turned into a semi-structured passage to be connected to Dr Onontology using Apache Jena framework. The name and dosage of available drugs were then extracted in the physician order forms and the patient progress notes. The areas of mismatch were identified through comparing the data obtained from these two forms.

Results: Software assessment was conducted in two stages. In the first stage, the capability of the software in proper recognition of medicine's name was measured, as100 completed forms containing physician order forms with a total number of 1014 drugs were used for text mining and error detection. After running the analysis in the error detection software, 93% of the drugs were properly recognized. In the next stage, comparisons were made between the physician order forms and the patient progress notes to find possible mismatches. Out of 1000 recorded drugs in the analyzed forms, the software was able to properly detect mismatches in 91.8% of the cases. The medication data available in i2b2 were used for conducting the assessment.

Conclusion: Given that medical records are of paramount importance and their human analysis is a complicated and time-consuming process, deployment of a text miner with the capability of quality analysis could facilitate error detection efficiently and effectively.

Keywords: Medical records, Medication error detection software, Qualitative analysis, Text miner

1. Introduction

Data stored in medical records are of critical importance [1, 2] since they can be used to ensure that benefits of patients and health care providers are realized and data could meet treatment, legal, and managerial expectations. Patients' prescription data constitute an important proportion of medical records [3]. These data are usually entered in the physician orders form and the patient progress notes [4]. The importance of medication data has to do with their role in detecting possible errors in prescribing or taking medicine, which significantly contribute to patients' reduced safety or even death. As long as regulations permit, data quality analysis is conducted

to identify and address deficiencies and errors in patients' medical records. This process guarantees data quality, including medication data, and makes it possible to prosecute the health care team, especially physicians and nurses, in medical auditing [5].

The process of quality analysis comprises a quantitative and a qualitative phase. In the quantitative phase, the presence or absence of data is investigated. In the qualitative phase, however, attempts are made to assess the adequacy and validity of data in medical records, which entail clarity, timeliness, relevance, definition, and data visualization [6, 7]. Carrying out qualitative analysis of medical record data is not easy for human beings and has some complications including:

Discrepancy in the nature of data

Data related to various treatment steps are different in nature. For example, the nature of laboratory data is different from that of pharmaceutical data. Given such discrepancies, different checklists are developed to qualitatively analyze data related to each of the treatment steps [6, 7, 8].

The time-consuming nature of qualitative data analysis

The process of qualitative data analysis is time-consuming. As a result, in many centers, random sampling is used to study a limited number of medical records [9].

Human errors

Due to the complexity of qualitative data analysis process and the importance of information in medical records, carrying out qualitative data analysis requires maximum precision. Thus, errors may occur if humans conduct qualitative data analysis [10].

The presence of such complications indicates the importance of using facilitating tools to improve the process of qualitative data analysis. A group of tools that can be used for this purpose are text miners [11].

Text mining of medical records is typically comprised of two steps. In the first step, the entity of expected words and expressions are properly recognized and extracted from the text [12]. This process, which is known as nominal entity recognition, is conducted in light of five techniques, namely rule-base, machine learning, dictionary-based, ontology-based, and combined technique [13, 14, 15, 16]. In the next step, the extracted entities are analyzed [13]. At the end of this step, the results of text mining emerge.

A small number of studies have used text mining to analyze medical records. For example, a research was conducted in University of Texas based on the UIMA framework and the RxNorm [17] naming system. The study led to the development of a medication information extraction system– namely Medex-UIMA, which has the capability to identify the name, dosage, and drug use period in healthcare texts [18].

In another study, text mining was utilized to establish a framework to extract pharmaceutical concepts and their interrelations in patients' medical records. This study was carried out using UMLS database [19].

Given the importance of conducting qualitative data analysis in medical records and the significant role of medication data error recognition in improving patients' safety [20], the current study sought to use text mining techniques to develop a software to detect medication data errors in medical records. The designed software was subsequently used to compare data in the physician order forms and the patient progress notes with the aim of identifying errors in medicine documentation as soon as they are committed.

2. Materials and Methods

Materials

In this study, attempts were made to develop a software which could be used to analyze and compare medication forms in medical records with the aim of recognizing and reducing errors. To this end, first the software requirements were identified. In the next stage, based on the identified requirements and by the use of object-oriented analysis, the software was analyzed precisely and three-layer architecture was designed for it. Each layer not only has its particular responsibilities and features, but also can directly interact with the next layer above or below. The three layers entail data access layer, business layer, and presentation layer. The responsibilities of each layer are described below.

Data Access layer:

This layer is responsible for accessing data outside the software. Using the Apache Jena framework and Sparql programming language, it interacts with the ontology used in the software and conducts necessary searches. The ontology, which is named DrOn and is based on RXNORM, contains medication information such as brand name, generic name, and equivalent names for drugs. The search output is sent to the higher level for processing.

Business layer

This layer includes the major logic of the software and processes the data sent from the lower layer. In this layer, StanfordCoreNLP framework is used for some text processing. After text processing, the results of text mining of available data in the two forms are compared and the final output is sent to the upper layer to be displayed to the user.

Presentation layer

This layer encompasses all parts of the software that are visible to the user and is responsible for establishing relationship between the user and the software logic.

The software was developed using Java.

Software assessment

Software assessment was carried out in two stages. In the first stage, the software capability in proper recognition of the names of medicines in the text was gauged. To do so, 100 completed forms which contained physician orders with a total number of 1014 drugs were used for text mining and error detection. In the next stage, comparisons were made to find similarities between the physician order forms and the patient progress notes.

3. Results

At the end of the process of developing the software which aimed at detecting errors in medication data of medical records, attempts were made to assess the software. To do so, 100 completed forms which contained prescriptions with a total number of 1014 drugs were used for text mining and error detection. After running the analysis in the error detection software, 93% of the drugs were properly recognized. Overall, 76.7% were correctly recognized, while 23.3% were erroneously identified. In the next stage, comparisons were made to find similarities between the physician order form and the progress note. In this stage, 100 pairs of the physician order forms and the patient progress notes were compared.

In each comparison, the medication data recorded in the physician order form were compared with those recorded in the patient progress note. Out of the 1014 available drugs, 91.8% of medication mismatches (which indicate errors in recording medication data in the two forms) were properly detected. However, the software failed to recognize 8.2% of the mismatches in the two forms. It should be noted that the medication data available in i2b2 were used for conducting the assessment.

4. Discussion

The importance of medication data has to do with their use in recognizing errors committed in the process of prescribing and consuming drugs. These errors could strongly reduce patients' safety and may even result in their death. Text miners – one of which was developed in the current study – can be used for qualitative analysis during the documentation process to detect errors as soon as they occur and to prevent subsequent problems. This error detection process not only helps patients but also is a great aid to health care providers and health information management personnel.

Despite its importance, there are not a lot of studies on this issue. The most relevant studies are mainly conducted on drug availability detection. For example, a study entitled "Extracting and standardizing medication information in clinical text- the medEx-UIMA system" was carried out in 2014. The results indicated that the F value (which is an indicator of measurement accuracy) for data extraction was around 98.5% and 97.5%. Moreover, the F value for mapping the extracted drugs into RxNorm was about 85.4% and 88.1% [18]. Another study entitled "A medication extraction framework for electronic medical records" was conducted in 2012. The results indicated that the F value ranged from 800 to 900 [19].

The software developed in this study was used to analyze two medical forms – namely the physician order forms and the patient progress notes. It can be regarded as a primary step to develop an advanced software to thoroughly analyze patients' medical records. On the other hand, the accuracy of the developed software is not significantly different from the available ones in terms of identifying drug availability.

There are some important issues about the developed software that must be taken into consideration:

1. This software is only capable of recognizing medicines available in DrOn ontology. Every drug that is not available in this ontology will be discarded as extra token.
2. In each round of analysis, the software is able to compare and analyze a pair of forms.
3. Violation of grammatical rules in the texts can reduce the precision of analysis conducted by the software.

The presence of words or expressions that are partially similar to drugs' names in the text can reduce

the precision of analysis conducted by the software.

5. Conclusion

Given that medical records are of paramount importance and their human analysis is a complicated and time consuming process, deployment of a text miner with the capability of quality analysis could facilitate error detection efficiently and effectively.

As mentioned before, the results of analysis indicated that 23% of the items recognized as medicine by the software were not drugs at all. This can be attributed to the incompatibility of software requirements and the used ontology. Thus, the development of an ontology that can fulfill the software requirements will significantly improve the recognition precision of the software. Also, violating grammatical rules can reduce the accuracy of recognizing drugs. Adding machine leaning algorithms can measurably minimize this problem.

Ethical Considerations

Compliance with ethical guidelines

There were no ethical considerations to be considered in this research.

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Author's contributions

Hamid Moghaddasi proposed the idea and topic; formulated research problem; reviewed the literature; helped with the methodology and software development. He is Professor of Health Information Management & Medical Informatics. Email Address: Moghaddasi@sbmu.ac.ir.

Reza Rabiee has an advisory role in the thesis.

Sara Shadmani collected data; reviewed the literature; and developed the software under supervision of Dr. Hamid Moghaddasi.

Conflict of interest

The Authors declare that there is no conflict of interest.

Acknowledgments

We would like to extend our appreciation and thanks to university research center which made this study possible.

References

[1] WHO. Manila: WHO Regional Office for the Western Pacific [Internet] 2003. Available from: http://apps.who.int/iris/handle/10665/206974

[2] Gilligan L. HES Standards and recommended Practices for Health Care Records Management. Transmission physics and consequences for materials selection, manufacturing, and applications. J Eur Ceram Soc. 2014. p.10-12. https://www.hse.ie/ eng/about/who/qid/quality-and-patient-safety-documents/v3.pdf

[3] Uzuner O, Solti I, Cadag E. Extracting medication extraction from clinical text. J Am Med Inform Assoc. 2010; 17(5):514-8. [DOI:10.1136/jamia.2010.003947] [PMID] [PMCID]

[4] Huffman E. Medical record management. 9th edition. United States: Physicians Record Company;1990. https://books.google.com/books?id=tmxrAAAAMAAJ&q=Huffman+E.+Medical+record+management&dq=Huffman+E.+Medical+record+management&hl=en&sa=X&ved=2ahUKEwjaiPiE-df4AhWERPEDHYO3CowQ6AF6BAgJEAI

[5] Karp D, Huerta JM, Dobbs CA, Dukes DL, Kenady K. Medical record Documentation for patient safety and physician defensibility. California: Medical insurance Exchange of California; 2008. file:///C:/Users/mahlaa/Downloads/medical-record-documentation-for-patient-safety%20(1).pdf

[6] Moghaddasi H. Information Quality in Health Care. 2th edition.Tehran: Vajehpardaz Publish company.2012.

[7] Moghaddasi H, Rahimi F. A systemic biologic model for healthcare data quality. HIM-interchange. 2016; 6(1):28-39. https://www.researchgate.net/publication/299598216_A_systemic_biologic_model_for_healthcare_data_quality

[8] Gerg M. The essential nature of healthcare databases in critical care medicine. Crit Care. 2008; 12:1-2. [DOI:10.1186/cc6993]

[9] Kushima M, Araki K, Suzuki M, Araki S, Nikama T. Text data mining of electronic Medical Record of the chronic hepatitis patient. Proc Int Multiconf Comp Sci Inf Technol. 2012; 1:1-5. http://www.iaeng.org/publication/IMECS2012/IMECS2012_pp569-573.pdf

[10] Khare R, An Y, Wolf S, Nyirjesy P, Liu L, Chou E. Understanding the EMR error control practices among Gynecologic Physicians. iConference 2013 Proceedings. 2013:289-301. [DOI:10.9776/13197]

[11] Chiaramello E, Pinciroli F, Bonalumi A, Caroli A, Tognola, G. Use of "off-the-shelf" information extraction algorithms in clinical informatics: a feasibility study of MetaMap annotation of Italian medical notes. J Biomed Inform. 2016; 63:22-32. [DOI:10.1016/j.jbi.2016.07.017]

[12] Chen HC, Fuller S, Friedman C, Hersh W. Medical Informatics: Knowledge Management and Data Mining in Biomedicine. New York: Springer; 2005. https://link.springer.com/book/10.1007/b135955

[13] Gurulingapa H. Mining the Medical and Patent Literature to Support Healthcare and Pharmacovigilance. (Ph.D. dissertation). Germany: University of Bonn; 2012. https://dblp.org/rec/phd/dnb/Gurulingappa12.html

[14] Pereira L, Rijo R, Silva C, Martinho R. Text mining applied to electronic medical records: a literature review. Int. J E-Health Med Commun. 2015; 6(3):1-18. [DOI:10.4018/IJEHMC.2015070101]

[15] Huang Z, Hu X. Disease named entity recognition by machine learning using semantic type of metathesaurus. International Int J Mach Learn Comput. 2013; 3(6):494-98. [DOI:10.7763/IJMLC.2013.V3.367]

[16] KolaliKhormuji M, Bazrafkan M. Persian named entity recognition based with local filters. Int J Comput Appl. 2014; 100(4):1-6. [DOI:10.5120/17510-8062]

[17] RxNorm. National Library of Medicine. [Internet]. Available from: http://www.nlm.nih.gov/research/umls/ rxnorm/

[18] Jiang M, Wu Y, Shah A, Priyanka P, Denny JC, Xu H. Extracting and standardizing medication information in clinical text- the medEx-UIMA system. AMIA Jt Summits Transl Sci Proc. 2014:37-42. [PMID] [PMCID]

[19] Bodnary A. A Medication extraction framework for electronic medical record. Massachusetts Institute Of Technology. 2013. https://www.researchgate.net/publication/279815871_A_medication_extraction_framework_for_electronic_health_records

[20] Doan S, Bastarache L, Klimkowski S, Denny J, Xu H. Integrating existing natural language processing tools for medication extraction from discharge summaries. J Am Med Inform Assoc. 2010; 17(5):528-31. [DOI:10.1136/jamia.2010.003855] [PMID] [PMCID]