A Review of Existing Applications and Techniques for Narrative Text Analysis in Electronic Medical Records

Alexandra Pomares-Quimbaya
Pontificia Universidad Javeriana, Colombia

Rafael A. Gonzalez
Pontificia Universidad Javeriana, Colombia

Santiago Quintero
Pontificia Universidad Javeriana, Colombia

Oscar Mauricio Muñoz
Pontificia Universidad Javeriana, Colombia

Wilson Ricardo Bohórquez
Pontificia Universidad Javeriana, Colombia

Olga Milena García
Pontificia Universidad Javeriana, Colombia

Dario Londoño
Pontificia Universidad Javeriana, Colombia

INTRODUCTION

Electronic Medical Records (EMR) contain information that is typically stored in structured (coded) attributes as well as unstructured narrative data. These narrative data or free text includes: discharge notes, progress notes, radiology reports, pathology reports, nursing notes, and general clinical notes. These texts contain valuable patient information, but are often underused and only looked at for a particular patient, remaining hidden for population studies, clinical studies or administrative analysis. As a consequence, finding and using evidence contained in medical records remains fragmented due to the fact that narrative text is cumbersome to analyze.

To face this challenge, different research projects have applied natural language processing techniques and/or text mining techniques to improve the quantity and quality of the information extracted from EMR aimed at supporting clinical research and administration. From the content point of view, many such projects are focused on the analysis of a specific disease, while others try to improve the pharmacovigilance process. From the technical point of view, most of them deal with the extraction of patterns from narrative texts, while others propose methods for de-identification (anonymization). Another group emphasizes on solving co-reference and redundancy aspects. Lastly, others work on the generation of summaries from original texts.

The objective of this chapter is to present a thorough exploration of these projects and synthesizes this diversity into a coherent classification. To achieve this, the chapter is organized as follows: in Section 2 we present the background of the work and the method used to obtain the projects that were included in the study. Then, Section 3 presents different applications of natural language processing to improve the analysis of specific diseases or process, and Section 4 explores generic tools and methods that have been propose for the analysis of narrative medical information or narrative information in general. Finally, Section 5 explores future challenges and research perspectives, and Section 6 concludes the chapter proposing a classification of recent works around natural language processing and text mining applied on the analysis of narrative text contained in medical records.

BACKGROUND

The obstacles for using the extensive narrative data found within EMR in research projects, mainly due to their lack of structure and standardization, have motivated different types of works. This chapter presents projects that have demonstrated successful use of Natural Language Processing (NLP) and/or data mining techniques for the exploitation of EMR narrative data. These works can be classified into two broad groups: the first group uses NLP or data mining techniques in the context of a disease or a process; for instance, the analysis of a specific disease or within a pharmacovigilance process; this group is called **N***LP applications for medical analysis*. The second group, called *Generic NLP Methods and Tools*, comprises works that propose methods or techniques to improve the analysis of texts regardless of the context, including, generating summaries, de-identifying narrative texts and solving redundancy aspects.

This chapter surveys recent work in NLP and text mining over medical records. The period of the analysis ranges from 2008 to the beginning of 2014. Even though there are previous works on this subject, we decided to restrict the dates considering the recent advances on NLP and text mining the last years.

Papers were identified using Web of Science database[1], and specifically the results obtained from the following query: TS=(EHR or Electronic Health Record or Medical Health Record) and TS=(text mining or natural language processing or information retrieval) and TS= (text-free or free-text or free text or narrative text or text or medical notes or nursery notes)) <i> AND </i>LANGUAGE: (English). From the obtained list of paper we selected interesting publications by analyzing the titles and their abstracts.

Besides the works identified though this query, we identified other previous works that recognizes the importance of applying text mining techniques to biomedical literature in order to recognize entities, their relationships, create summaries and enhance question-answer systems from scientific publications, like the work presented by (Zweigenbaum, Demner-Fushman, Yu & Cohen, 2007), and using ontologies to facilitate associations between genomic data and information from biomedical texts, as is the case of the paper presented by (Tiffin, 2005). However, we decided not to include them in this study because they were more focused on the analysis of literature and not the analysis of medical records, which is the focus of this chapter.

NLP APPLICATIONS FOR MEDICAL ANALYSIS

In health domain NLP applications have been used for the analysis of a specific disease with successful results. The first group of solutions is focused on the application and/or evaluation of NLP processes or data mining techniques to improve the knowledge of a disease or healthcare process.

Related Content

Refinement of Hypothesis Testing in Conjugation Tables of r(c Size on the Example of Testing New Forms of Treatment
Lidiya Filippovna Taenvatand Mikhail Mikhailovith Taenvat (2022). *International Journal of Health Systems and Translational Medicine (pp. 1-13).*
www.irma-international.org/article/refinement-of-hypothesis-testing-in-conjugation-tables-of-rc-size-on-the-example-of-testing-new-forms-of-treatment/306691

IoMT in Smart Cities Scaling Challenges, Smart Healthcare Applications, Data Integrity
Tarun Kumar Vashishth, Vikas Sharma, Kewal Krishan Sharma, Bhupendra Kumar, Sachin Chaudharyand Rajneesh Panwar (2024). *Lightweight Digital Trust Architectures in the Internet of Medical Things (IoMT) (pp. 281-309).*
www.irma-international.org/chapter/iomt-in-smart-cities-scaling-challenges-smart-healthcare-applications-data-integrity/347590

Revolutionizing Diabetic Retinopathy Diagnostics and Therapy Through Artificial Intelligence: A Smart Vision Initiative
R. S. M. Lakshmi Patibandla, B. Tarakeswara Raoand M. Ramakrishna Murty (2024). *Transformative Approaches to Patient Literacy and Healthcare Innovation (pp. 136-155).*
www.irma-international.org/chapter/revolutionizing-diabetic-retinopathy-diagnostics-and-therapy-through-artificial-intelligence/342824

Topical Use of Plant Extract-Based Oil Blend in Relieving the Symptoms of Primary Dysmenorrhea: An Independent Clinical Study
Amul S. Bahl (2021). *International Journal of Health Systems and Translational Medicine (pp. 47-61).*
www.irma-international.org/article/topical-use-of-plant-extract-based-oil-blend-in-relieving-the-symptoms-of-primary-dysmenorrhea/270953

The Urine Drug Screen in the Emergency Department: Overuse, technical pitfalls and a call for informed consent.
 (2022). *International Journal of Health Systems and Translational Medicine (pp. 0-0).*
www.irma-international.org/article//282680