Chapter 16

Natural Language Processing and Machine Learning Techniques Help Achieve a Better Medical Practice

Oana Frunza
University of Ottawa, Canada

Diana Inkpen
University of Ottawa, Canada

ABSTRACT

This book chapter presents several natural language processing (NLP) and machine learning (ML) techniques that can help achieve a better medical practice by means of extracting relevant medical information from the wealth of textual data. The chapter describes three major tasks: building intelligent tools that can help in the clinical decision making, tools that can automatically identify relevant medical information from the life-science literature, and tools that can extract semantic relations between medical concepts. Besides introducing and describing these tasks, methodological settings accompanied by representative results obtained on real-life data sets are presented.

INTRODUCTION

Since its early beginnings, artificial intelligence research directed its interests towards the medical field. Expert systems are among the first applications that merged the two fields. Since the early 70’s, when expert systems like MYCIN\(^1\) were built, a lot has changed. Living in an information explosion era, a big emphasis is currently made on automatic ways to extract relevant medical information form huge amounts of various data types (data mining). Textual data is one of the most common ways of representing medical information and, with the recent trend of using digital medical records, this type of data becomes a valuable resource for both the medical and the computational linguistics field.

DOI: 10.4018/978-1-4666-1803-9.ch016
Achieving a superior medical practice most often correlates with having medical practitioners use automatic tools. This chapter presents work done on three major tasks in which the computerized world can help: assisted clinical decision making for obesity-related diseases, identification of relevant medical information for building systematic reviews on medical topics, and identification of important relations between medical concepts like disease, treatments, and tests.

The first task that we address, assisted clinical decision-making, is motivated by the fact that disease identification and profiling represents a challenging task in the medical domain. A valuable source of information for understanding the way medical conditions appear, progresses, and get cured is represented by clinical narratives. Besides scientific discoveries that get published in technical journals, new information that can be extracted from medical records becomes accessible to the research community, facilitating progress both in medicine and in computational linguistics. The fact that reliable and relevant information can help solve these paramount tasks motivated us to use natural language processing techniques combined with machine learning tools to solve these tasks.

Since obesity has become a major medical problem nowadays, we decided to focus our work on building clinical decision support systems capable to automatically identify obesity-related diseases in clinical data. The benefits of having such tools are in assisting with the clinical decision process (the believed rate of diagnostic error is between 15-20%), in identifying groups of patients (e.g., patients that belong to a specific type of disease, high-risk patients, patients suitable to participate in a specific clinical trial, etc.), and in extracting information that can be useful for disease prevention and therapy.

The second research issue that we address in this current work is the identification of relevant medical information from huge amounts of data. Identifying relevant information from the life-sciences literature can help in identifying the best medical evidence for a certain medical problem and, more importantly, can help in the process of building systematic reviews — summaries of research discoveries which represent fundamental tools for decision making in an evidence-based medicine practice. From a computerized point of view, this task entails a text-classification approach that automatically identifies relevant data to a specific medical problem. We focused our work on identifying relevant information that answers the issue of efficiency of medical care for elderly people.

The third task which we address in this current work is focused on identifying relations between medical concepts in technical and clinical data. Three types of concepts: diseases, treatments, and medical tests are included in the list of eight semantic relations that we try to automatically identify. The outcome of this task can help in the development of medical ontologies, question-answering systems on medical problems, in the creation of clinical trials, in stratifying patients by disease susceptibility and response to therapy, in reducing the size, duration, and cost of clinical trials, and ultimately in leading to the development of new treatments, diagnostics, and prevention therapies.

**BACKGROUND**

From the wealth of relevant research that has been done in the biomedical domain, we are going to present only representative work focused on the tasks that we address. Research on clinical data is less represented in the literature, due to lack of access to data.

The research endeavors that we followed in order to build a clinical decision support system that can assist in the diagnosis of obesity-related diseases builds on previous research that has been done on this topic. Some of the representative work on this topic uses lexical and domain knowledge...
16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/chapter/natural-language-processing-machine-learning/67262?camid=4v1


www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

Behavior Change through ICT Use: Experiences from Relatively Healthy Populations
Marieke W. Verheijden (2007). Web Mobile-Based Applications for Healthcare Management (pp. 118-141).
www.igi-global.com/chapter/behavior-change-through-ict-use/31154?camid=4v1a

A Next Generation of IT: A Health Record Example with Terminology Binding
www.igi-global.com/article/a-next-generation-of-it/211976?camid=4v1a

The Arts as Public Service: A Critical Analysis Exploring the Relationship between the Arts and Wellbeing
www.igi-global.com/article/the-arts-as-public-service/101341?camid=4v1a

Identifying the Essential Design Requirements for Usable E-Health Communities in Mobile Devices
www.igi-global.com/chapter/identifying-essential-design-requirements-usable/73853?camid=4v1a