Chapter 8
Text Mining and Natural Language Processing for Health Informatics:
Recent Trends and the Way Forward

Anoop V. S.
Rajagiri College of Social Sciences (Autonomous), Kochi, Kerala, India

ABSTRACT

Health informatics deals with applying informatics to medicine and healthcare that aims to store, process, and retrieve large amounts of healthcare data to enable optimal collaboration between different stakeholders. This has several applications in the healthcare domain from extracting information from medical documents such as case reports and prescriptions to analyzing data from sensors available in wearable devices. Recent advancements in information and communication technologies fueled the need of devising intelligent technologies for analyzing such data – not only in various forms but also in large quantities. This has posed many challenges and opportunities to use techniques such as text mining, natural language processing (NLP), and deep learning to unearth the latent themes from the vast array of textual data. This chapter proposes some prominent works in health informatics that use text mining and NLP and also discusses some active research areas in these dimensions. This chapter will be useful to understand the recent advancements and future research dimensions.

INTRODUCTION

The recent developmental acceleration happens in the areas of intelligent computing, healthcare informatics is getting shifted to a new paradigm where advanced techniques such as deep learning and natural language processing are getting widely implemented in many areas. Since the major share of data accumulated and processed in healthcare are in the form of unstructured text, mining them to find latent patterns is an active research area with many open challenges. There are many innovative research reports in healthcare informatics for addressing these challenges, but there are still a lot of avenues where text

DOI: 10.4018/978-1-7998-7188-0.ch008

mining and natural language processing can be fully exploited for tackling the challenges. Compared with other areas of intelligent computing, health informatics is relatively new and attempts to use information technology capabilities to organize and analyze large quantities of unstructured text to unearth the patterns. Using computational techniques, health informatics attempts to leverage useful trends and information from humongous volumes of medical data that may support the healthcare decision-making process. This field has gotten significant attention in the very recent past due to the COVID-19 pandemic outbreak where millions of people got affected very adversely. The pandemic has affected the world unprepared and there were no clues among the healthcare workers and administrators on how to tackle the challenges posed by COVID-19. The manual way of fighting the pandemic with the activities such as case identification and contact tracing was proved inefficient in many situations and highlighted the need for using information technology-enabled approaches for dealing with the scenario in a better way. The government and public healthcare workers and officials are overwhelmed with the data - the most important tool for them to take decisions expedited and for planning, decision-making, and measuring effectiveness. This has again proved the need for automated techniques and tools to analyze this data effectively and the adoption of health informatics practices in better dealing with such situations.

Text mining and natural language processing attempts to analyze vast bodies of unstructured text documents and unearth the hidden patterns using processes such as information extraction, text classification, and topic modeling. Recent advancements in machine learning such as deep learning outperformed many state-of-the-art shallow machine learning approaches. From rule-based systems to machine learning to deep learning, this field has gone through many significant changes that improved the accuracy of many machine learning models. Healthcare domain generates large quantities of unstructured text in many forms such as patient generated data, social medical data, clinical narratives, discharge summaries, to name a few. So there exists a huge opportunity to analyze and unearth the patterns containing entities and relationships, that may find several applications in the healthcare industry. Health informatics which is a new but active research area that has got significant attention in the recent past which heavily makes use of text mining, natural language processing, and machine learning. There are several interesting research papers reported in the literature but there also exists many interesting research problems that need to be explored. In this connection, this chapter outlines some of the prominent state-of-the-art in health informatics and also future research trends and dimensions in this area.

The remainder of this chapter is organized as follows. Section 2 discusses some of the very recent and prominent works that were reported in the literature on healthcare informatics exploring the potentials of text mining and natural language processing techniques. Section 3 details the recent trends on adopting the intelligent computing paradigms to healthcare informatics and in Section 4, the future trends and further research dimensions are presented. Section 5 concludes this chapter.

STATE-OF-THE-ART IN HEALTH INFORMATICS USING NATURAL LANGUAGE PROCESSING AND TEXT MINING

This section discusses some of the prominent works in health informatics that uses techniques such as text mining, and natural language processing. The recently reported approaches which are closely related to the theme of this chapter have been presented for the readers to understand what is already in the literature. The applications of natural language processing for health-related text contents were discussed in detail by Dina Demner et. al. (Demner-Fushman et al., 2021) in a recent work reported