

Research on text analysis and its application in electronic medical record data analysis

Phonesavat Sitthiphong

Abstract: This study explores the application of text analysis methodologies to electronic health records (EHRs), aiming to extract clinically meaningful insights from unstructured narrative data. While structured data in EHRs is readily quantifiable, a vast reservoir of patient information resides in free-text clinical notes, discharge summaries, and diagnostic reports. This research develops and evaluates natural language processing (NLP) techniques, including named entity recognition, topic modeling, and sentiment analysis, to systematically process Vietnamese medical texts. The primary objectives are to identify disease patterns, extract medication and symptom relationships, and detect adverse event signals that are often missed in coded data. A corpus of de-identified EHRs from a major hospital was used for model training and validation. Results demonstrate that the proposed text analysis framework significantly improves the accuracy of information retrieval and supports predictive modeling for patient outcomes. The findings underscore the potential of text mining to enhance clinical decision support, streamline medical coding, and facilitate large-scale epidemiological research, ultimately contributing to more efficient and data-driven healthcare systems.

Keywords: Text analysis, Electronic health records, Data mining, Natural language processing