Chapter 2
Classical Association Rule Mining Algorithms for Medical Data

Humara Yaqub Thakur
http://orcid.org/0000-0003-3815-9124
King Khalid University, Saudi Arabia

Malik Misbah
King Khalid University, Saudi Arabia

Mahjabeena Najar
King Khalid University, Saudi Arabia

Shaista Qadir
King Khalid University, Saudi Arabia

ABSTRACT

The explosive growth of medical data in recent years ranging from electronic health records to clinical trial outcomes has created a pressing need for advanced data analysis techniques that can reveal hidden disease patterns, improve diagnostic processes, and strengthen evidence-based medical practices. Among these techniques, association rule mining (ARM) has become a key tool for uncovering meaningful relationships in complex, high-dimensional healthcare datasets. This chapter aims to connect computational approaches with clinical value by demonstrating how ARM can transform raw healthcare data into actionable insights for doctors, researchers, and healthcare administrators. As medical decision-making increasingly relies on

DOI: 10.4018/979-8-3373-6691-3.ch002

interconnected datasets including patient histories, symptoms, treatments, test results, and outcomes, ARM algorithms provide a powerful way to identify relationships that may not be immediately obvious to human experts..

INDTRODUCTION

The continuing development of medical informatics is keeping pace with the exponential increase in the volume, variety, and complexity of health data. Each day, modern health systems manage millions of electronic health records, diagnostic images, genomic sequences, and transactional data that, if harnessed and utilized effectively, can offer clinicians and researchers invaluable, life-saving information. However, deriving actionable insights is not straightforward. The medical domain presents unique obstacles, such as heterogeneous clinical variables, missing data, and highly non-linear relationships between patient characteristics and outcomes (Bellazzi & Zupan, 2008). These issues create an urgent need for transparent and interpretable analytical methods that can traverse dense medical datasets and surface novel, actionable insights (Agrawal & Srikant, 1994).

Association Rule Mining (ARM) has gained prominence because of its ability to facilitate the identification of intricate, previously unknown connections within datasets. In contrast to supervised learning methods that predict outcomes, ARM searches for co-occurrence patterns and generates profiles of symptoms, diagnoses, treatments, and demographic combinations that tend to cluster together, determining the strength and clinical significance of these relationships (Agrawal, Imieliński, & Swami, 1993). In medical contexts, such findings can both corroborate existing scientific theories and generate new hypotheses to enhance decision-support systems and improve clinical outcomes (Patel et al., 2009).

Many of the classical ARM algorithms, such as Apriori, Eclat, and FP-Growth, have been instrumental in advancing biomedical research and clinical practice.

Like many recent advancements in association algorithms, Apriori is also famous for being among the first algorithms ever developed. It uses the 'level-wise' approach in mining, becoming the first algorithm for mining frequent itemsets, which was termed ‘Apriori’ because it uses the principle that 'a superset can be frequent only if all of its subsets are frequent.' Eclat and FP-Growth give parametric speeds of processing: Eclat uses vertical data and intersections for quick mining of large itemsets. FP-Growth uses a compact tree structure model for large itemsets and scans data to reduce overhead. (Agrawal & Srikant, 1994).

Prior research on association rule mining in healthcare has mainly looked at using individual ARM algorithms for specific medical issues, such as disease co-occurrence analysis, adverse drug reaction detection, and chronic disease risk