Association Rule Mining of Emergency 911 Calls

Genevieve Mortensen, Mahsa Monshizadeh, Aditya Shahapure

October 2023

GitHub repository: project-gamorten-mmonshiz-ashahapu

Abstract

This proposal outlines a data mining approach to analyze the 911 emergency dataset utilizing association rule mining techniques. By exploring the rich information embedded within the dataset, the project aims to identify and interpret significant patterns and associations, providing insights into the underlying factors that influence the occurrence and nature of emergency calls. The study's outcomes are expected to facilitate a more data-driven and efficient emergency response system, enhancing resource allocation and improving the overall management of critical situations, thus contributing to the public safety and well-being of the community.

Keywords

data mining, 911 emergency, association rule mining, emergency calls, public safety

1 Introduction

Emergency services are a vital component of any community, serving as a critical lifeline during unforeseen crises. The abundance of data generated by these services, especially the 911 emergency call dataset, presents an invaluable opportunity to derive actionable insights for improving emergency response systems. The application of association rule mining to this dataset can offer a comprehensive understanding of the complex relationships between various parameters associated with emergency calls, including time, location, incident type, and other relevant factors. By uncovering hidden patterns and correlations, we can potentially enhance decision-making processes, streamline resource deployment, and ultimately, ensure more efficient and effective emergency services.

2 Method

The proposed method involves a comprehensive approach to analyzing the 911 emergency dataset using association rule mining techniques, primarily the Apriori algorithm. The process encompasses several key phases, including data preprocessing, the application of the Apriori algorithm, and the interpretation of association rules. Although this project was previously proposed to be implemented in R via GeeksforGeeks, this project will be implemented using the Python programming language and relevant data mining and visualization libraries. Specifically, visualizing interesting relationships through exploratory data analysis followed by association rule mining with relevant public news coverage to justify rules found will be conducted as a two-fold validation of results.

2.1 Dataset

The 911 emergency call dataset utilized is publicly available through the Emergency Dispatch Operation department for Montgomery County, PA. The data is provided through Kaggle as a single, flat CSV file containing 663,282 unique emergency calls. The dataset includes various attributes related to separate 911 calls which were responded to by emergency dispatch for a single county. Aspects related to the specific time and location of each emergency, descriptions of the emergency, and labels related to the specific service provided in response to the dispatched call is provided. The dataset specifically details location in several ways, including longitude, latitude, township, ZIP code, and general address. Because the categories within the dataset are distinct and diverse, multitudes of meaningful association rules can be extracted.

2.2 Data Preprocessing

The initial step involves thorough data preprocessing, encompassing data cleaning, transformation, and formatting. Missing values will be addressed through imputation or deletion, ensuring data completeness. Furthermore, numerical values will be converted to categorical for effective association rule mining. Specifically, the time stamp will be transformed into the day of the week, month, and time of day categories, enabling the identification of temporal patterns and correlations. Additionally, specific service dispatched, such as fire, EMS, or police, will be separated from the description of the emergency, as these combined values represent distinct attributes with defined categories.

2.3 Apriori Algorithm

Following data preprocessing, the Apriori algorithm will be applied to identify frequent itemsets within the dataset. The algorithm operates on the principle that if an itemset is frequent, then all of its subsets are frequent, and if an itemset is infrequent, all of its supersets are infrequent. An itemset is considered frequent if it satisfies a minimum support threshold. This algorithm employs a breadth-first search strategy to discover itemsets that satisfy this criterion and prune infrequent itemsets accordingly. Typically, frequent itemset candidates are generated by merging itemsets of length k with the frequent itemsets of length $k-1$ and pruning infrequent itemsets based on the minimum support threshold. However, the alternative method for candidate generation where two frequent itemsets $k-1$ are merged if their first $k-2$ items are identical is employed as it generates fewer candidates and is therefore more space efficient.

2.3.1 Formulas:

The support for an itemset I will be calculated using the formula

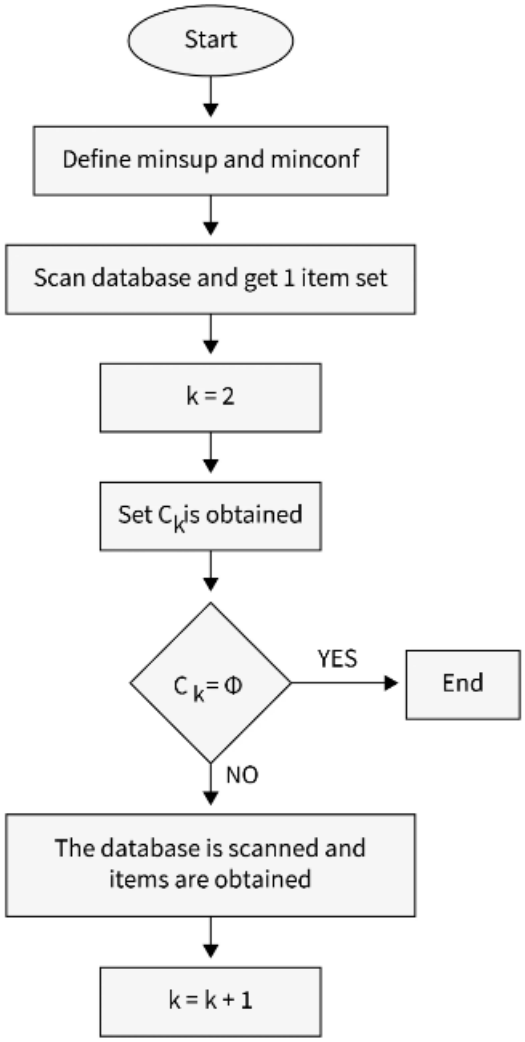
$$\text{support}(I) = \frac{\text{total number of transactions}}{\text{number of transactions containing } I}$$

Moreover, the confidence for an association rule $X \Rightarrow Y$ will be computed as:

$$\text{confidence}(X \Rightarrow Y) = \frac{\text{support}(X)}{\text{support}(X \cup Y)}$$

Furthermore, the lift for an association rule $X \Rightarrow Y$ will be determined using:

$$\text{lift}(X \Rightarrow Y) = \frac{\text{support}(Y)}{\text{confidence}(X \Rightarrow Y)}$$



2.4 Implementation in Python

The entire project will be implemented in Python, leveraging popular data mining libraries such as Pandas for data manipulation, and specialized libraries like mlxtend for the implementation of the Apriori algorithm. Python's flexibility and extensive range of libraries make it an ideal choice for handling complex data mining tasks, ensuring efficient and effective analysis of the 911 emergency dataset.

3 References

- https://github.com/tsengupta715/NLP-for-Ambulance-Calls
- https://www.geeksforgeeks.org/project-idea-analysis-emergency-911-calls-using-association-rule-mining/
- https://www.kaggle.com/datasets/mchirico/montcoalert
- https://www.cs.rit.edu/usr/local/pub/GraduateProjects/2165/gkm9983/Report.pdf

- https://data.cincinnati-oh.gov/Safety/Cincinnati-Fire-Incidents-CAD-including-EMS-ALS-BL/vnsz-a3wp/data
- https://www.scaler.com/topics/apriori-algorithm-in-data-mining/