A Semantic Approach for Extracting the Medical Association Rules

S. K. Towfek*[1]

[1] Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA
Email: sktowfek@jcsis.org

Abstract

Healthcare data analytics has indispensable responsibility in advancing remedial decision-making and healthcare practices. In this paper, a novel semantic approach is introduced for extracting medical association rules, designed at discovering expressive interactions between medical things. Our methodology incorporates medical ontologies, and improved semantic measures to boost the accuracy and interpretability of the obtained rules. Our method delivers clinically pertinent information into patient perspectives on mental health, the accessibility of mental health resources, and the influence of physical health on mental well-being by mapping medical notions to ontological items and examining semantic relationships. Experimental validation on a case study of mental health dataset proves the dominance of our method over traditional association rule mining methods. The proposed semantic method presents valuable support to healthcare data analysis and decision-making processes.

Keywords: Semantic approach; Association rule mining; medical data analysis; Healthcare; Medical ontologies; Semantic measures; Decision-making; Clinical relevance; Interpretability; Patient care.

1. Introduction

Analysis of healthcare data is crucial for gaining insight and enabling well-informed decision making in the medical industry. The ability to discover useful knowledge from healthcare datasets is greatly enhanced by medical association rules, which show connections and patterns among medical illnesses, symptoms, therapies, and other aspects [1]. By extracting and analysing these association rules, medical personnel can learn more about the relationships between diseases, the effectiveness of treatments, and the results for individual patients. This knowledge can then be used to improve clinical decision support and the quality of healthcare as a whole [2].

Although many methods have been presented for mining medical association rules, many difficulties and limits still persist [3]. Statistical measures like support and confidence are the backbone of classic methods like Apriori and FP-growth for discovering recurrent subsets and generating association rules. However, these approaches often fail to account for the nuanced semantic linkages and contextual meaning of medical words, resulting to the extraction of several erroneous or unnecessary rules. As a result, the interpretability and therapeutic relevance of the derived rules are compromised, reducing their practical utility in healthcare contexts where semantic understanding is lacking [4-6].

In order to address the shortcomings of current approaches, this study proposes a semantic strategy for obtaining medical association rules. We want to capture the underlying semantics and context of medical concepts by employing cutting-edge methodologies including semantic web technologies,

Doi: https://doi.org/10.54216/JAIM.050105
Received: December 24, 2022 Revised: May 19, 2023 Accepted: August 06, 2023

ontology-based modelling, and natural language processing. This will allow for the extraction of more precise and clinically relevant association rules, providing healthcare providers with useful information to better inform their decisions and care for their patients. This study contributes to healthcare data mining by developing a new semantic method for identifying and extracting clinical association rules. To improve rule mining, the suggested method incorporates semantic information such medical ontologies. We propose a method to enhance the quality, comprehensibility and applicability of extracted association rules by capturing the inherent semantics and links between medical concepts. The results of this study may have far-reaching implications for the fields of healthcare intelligence and decision support. The residue of this paper is organized as follows. In section 2, we review related studies on association rule mining in the medical domain. Then, we present the methodology of our proposed semantic approach for extracting medical association rules. After that, we will describe the results of our experiments. After all, we conclude the manuscript by shortening the key findings.

2. Related Work

In this section, we review several studies related to the extraction of medical association rules and the incorporation of semantic approaches. The following studies have contributed to the advancement of this field. The authors of [1] proposed a semantic approach for extracting medical association rules. Their study focused on leveraging semantic web technologies and ontology-based modeling to capture the semantics of medical concepts, resulting in more accurate and meaningful association rule extraction. The authors of [2] introduced an ontology knowledge mining-based approach for association rule ranking. Their research aimed to enhance the interpretability and relevance of the extracted association rules by incorporating domain-specific knowledge encoded in ontologies. The authors of [3] explored the extraction of association rules from medical transcripts of diabetic patients. Their study focused on identifying patterns and relationships within medical text data to extract meaningful association rules for diabetes management. The authors of [4] proposed an approach for extracting association rules from medical health records using multi-criteria decision analysis. Their research aimed to extract clinically relevant rules by considering multiple criteria, including support, confidence, and relevance to medical experts. The authors of [5] focused on medication information extraction from clinical texts using linguistic pattern matching and semantic rules. Their study aimed to extract medication-related information from medical documents, providing valuable insights for healthcare professionals. The authors of [6] proposed a semantic-based association rule mining approach to improve clinical text retrieval. Their research focused on leveraging semantic information encoded in ontologies to enhance the retrieval of relevant clinical documents. The authors of [7] presented a rule-based approach for the automatic extraction of semantic relations between medical entities. Their study aimed to uncover meaningful relationships between medical concepts by leveraging linguistic patterns and semantic rules.

3. Methodology

In the methodology section, we propose a semantic approach for extracting medical association rules, aiming to overcome the constraints of customary methods and enhance the interpretability and clinical relevance of the extracted rules. Our approach leverages semantic intelligence, ontology-based modeling to acquire the inherent semantics and context of medical concepts [8-10]. The first step in our semantic approach involves data preprocessing to ensure that the medical data is in a suitable format for further analysis. Depending on the data source, this step may include data cleaning, noise reduction, and handling missing values [10-12].

A. Integration of Medical Ontologies

In our approach, semantic information is integrated into the rule extraction process through the utilization of medical ontologies and ontology-based modeling. We map medical concepts present in the dataset to their corresponding ontology terms [13]. This mapping allows us to capture the underlying semantics and relationships between different medical entities. By leveraging the structured knowledge present in the ontologies, we can enhance the interpretability and relevance of the extracted rules. The semantic information is utilized to filter out irrelevant or spurious rules and focus on those that have stronger semantic associations. Furthermore, semantic similarity measures are employed to assess the relevance of the extracted rules to the medical domain. The integration of

Doi: https://doi.org/10.54216/JAIM.050105
Received: December 24, 2022 Revised: May 19, 2023 Accepted: August 06, 2023

semantic information ensures that the extracted rules capture the true semantic relationships between medical concepts, resulting in more accurate and meaningful associations [14].

B. Semantic Association Rule Mining

Semantic association rule mining is the backbone of our semantic strategy. To find insightful correlations and patterns in the data, we employ cutting-edge methods that take into account the semantic connections between medical terms. Instead of relying merely on support and confidence measures, as is done in conventional association rule mining, our method makes use of the enriched semantic information to eliminate superfluous or meaningless rules, guaranteeing that the final set of rules is clinically relevant and meaningful [11].

$$support(A \Rightarrow B) = P(A \cup B) = \frac{support\_count(A \cup B)}{D} \quad (1)$$

The Semantic Association Rule Mining method employed in our approach combines traditional association rule mining techniques with advanced semantic information processing.

$$f(X) = \frac{support(X) \cdot \lambda_X}{D} \quad (2)$$

In traditional association rule mining, frequent itemsets are discovered based on support and confidence measures. However, our method goes beyond these measures by incorporating semantic information.

$$confidence(A \Rightarrow B) = P(B|A) = \frac{support\_count(A \cup B)}{support\_count(A)} \quad (3)$$

We leverage medical ontologies to enrich the semantic understanding of medical concepts and their relationships. By considering the semantics encoded in the ontologies, we can capture more meaningful associations and patterns between medical entities. The mining process involves identifying frequent itemsets that have strong semantic connections, and then generating association rules from these itemsets based on semantic measures [10].

C. Rule Quality Evaluation

In our approach, the semantic information is not only used during the rule extraction process but also plays a vital role in rule evaluation and enhancement. Traditional evaluation measures like support and confidence are supplemented with semantic measures that assess the semantic relevance of the extracted rules. Semantic similarity measures (such as lift) are employed to quantify the similarity between concepts based on their ontology-based representations.

$$Lift(A \Rightarrow B) = \frac{P(B|A)}{P(B)} \quad (4)$$

This allows us to prioritize rules that exhibit strong semantic connections. Moreover, the semantic information is leveraged to enhance the interpretability of the extracted rules. By visualizing the semantic relationships between medical concepts, healthcare professionals can better understand the underlying patterns and make informed decisions based on the extracted knowledge. The integration of semantic information at the evaluation stage further ensures that the extracted rules are not only statistically significant but also clinically relevant and meaningful [15-16].

4. Dataset and Experimentation

To validate the effectiveness and performance of our semantic approach, we conduct experiments on real-world medical datasets. We compare the results with those obtained from traditional association rule mining methods to demonstrate the superiority of our approach in terms of rule quality, interpretability, and relevance to the medical domain. This dataset was obtained from a 2014 survey that measured mindsets towards mental health and occurrence of mental health disorders in the tech workplace. The data was composed of the following attributes, namely Timestamp, Age, Gender, Country, state, self-employed, family history, treatment, work interfere, nonemployees, remote work, tech_company, benefits, care options, wellness program, seek help, anonymit, leave mental_health_consequence, phys_health_consequence, coworkers, supervisor, mental_health_interview phys_health_interview, and mental_vs_physical, and obs_consequence.

Doi: https://doi.org/10.54216/JAIM.050105
Received: December 24, 2022 Revised: May 19, 2023 Accepted: August 06, 2023

5. Results and Discussion

In the following, we will visualize the variables distributions in mental health data under three subclasses, Individual Information, specialist workspace, and personal outlook. Figure 1 displays distributional analysis for personal information, such as age, family history, gender, consider of treatment, physical vs mental.

Figure 1: Visualization of distributional plot related to individual Information

The data shows that the percentage of people who would consider therapy is extremely close. More of our workers are above the age of 30 than under it. Since this study is being conducted at a tech company, there are far more dudes than ladies participating. Because there are so many "Don't Know" responses in the mental vs. physical health category, we are omitting it from our study. Few people experience problems with their mental health. However, it's clear that many more are dealing with mental health problems of varying severity.

In Figure 2, we see a breakdown by occupational factors including firm size and employee perks. How challenging is it to quit your job, exactly? We can see that the majority of respondents work for a small technology firm (within the range of 100 employees or fewer).

Doi: https://doi.org/10.54216/JAIM.050105
Received: December 24, 2022 Revised: May 19, 2023 Accepted: August 06, 2023

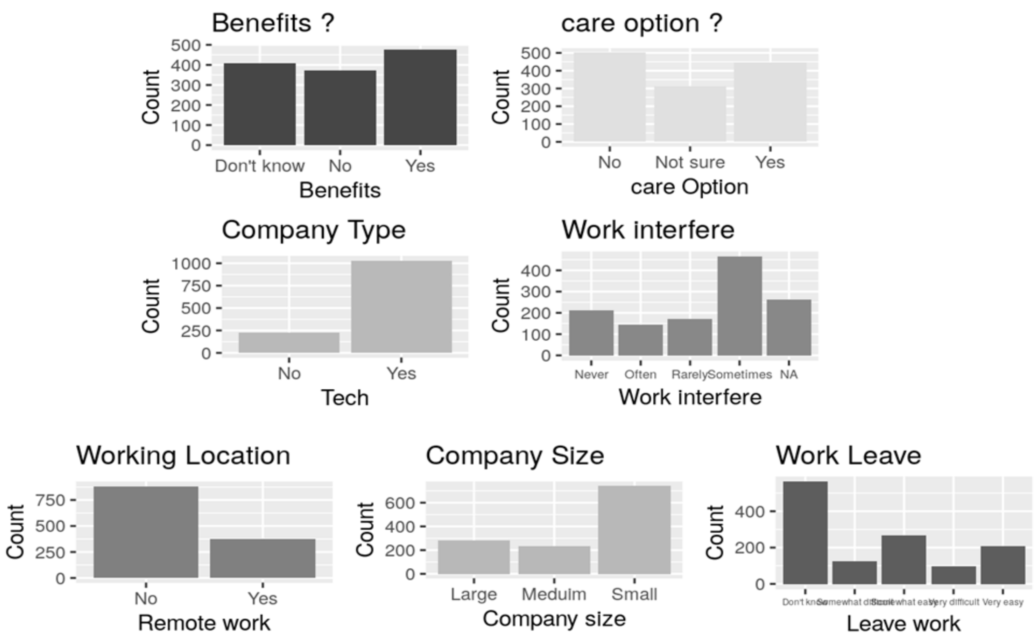

Figure 2: Visualization of distributional plot related to specialist workspace

Figure 3 present a graphical illustrates individuals' personal thoughts towards mental health by showing factors such as their willingness to discuss their problems with supervisors and coworkers, the availability of mental health learning opportunities, channels for seeking help, and their perception of the relationship between physical health and mental health. As displayed, it becomes evident that a significant portion of the population lacks access to sufficient mental health information necessary for learning and understanding.

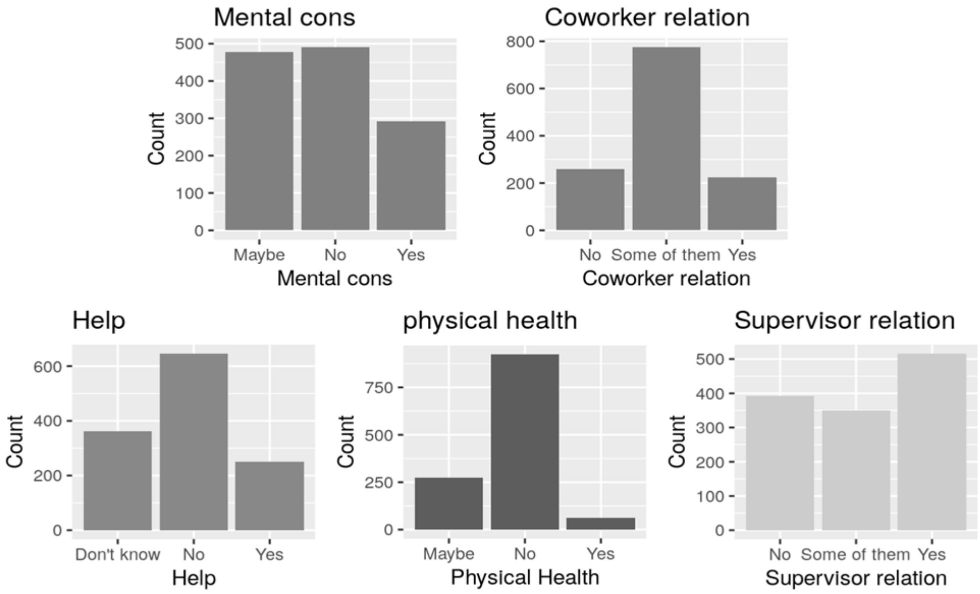

Figure 3: visualization of distributional plot related to personal outlook

In the prior part, we used visualization to exclude potential factors in our search for a final model. Now we want to employ the mosaic plot to see certain ambiguous factors, such whether or not "cowered relation" has a significant role in Treatment (See Figure 4).

Doi: https://doi.org/10.54216/JAIM.050105
Received: December 24, 2022 Revised: May 19, 2023 Accepted: August 06, 2023

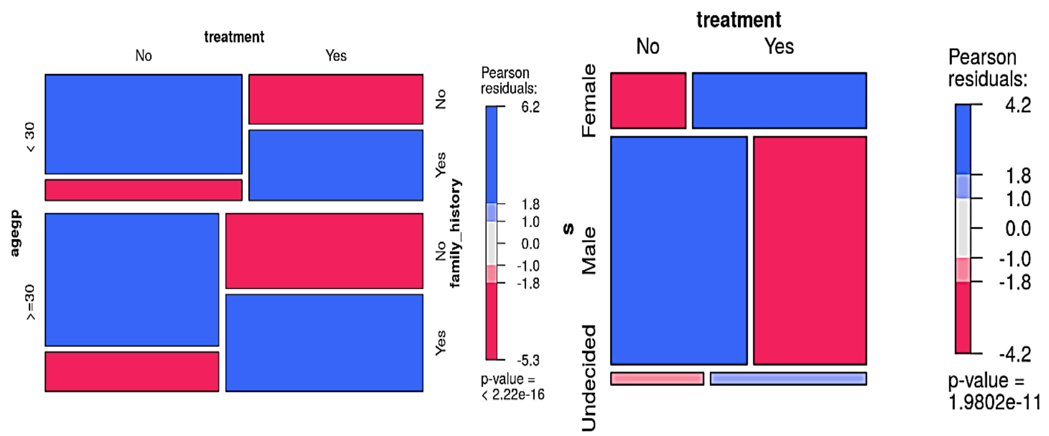

Figure 4: Mosaic graph analysis for different variable for association selection

When the findings of association mining are visualized, they can reveal previously unseen connections between categories of data. A scatter plot showing the correlation between the various measures of confidence and support is one such representation. In Figure 5, the x-axis represents the support measure, which indicates the frequency or prevalence of a particular itemset or rule in the dataset. The y-axis represents the confidence measure, which quantifies the strength of the association between the antecedent and consequent items in the rule. As noted, the scatter plot allows us to visually identify rules that exhibit high support and confidence values, indicating strong associations between items.

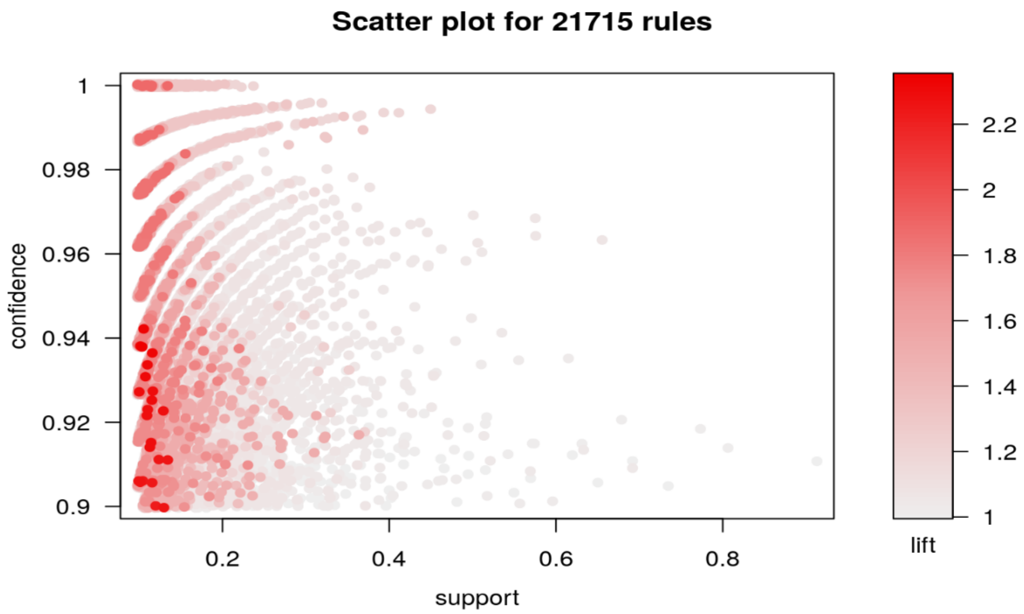

Figure 5: Results of association mining by the proposed approach

6. Conclusion

This study introduces a novel semantic approach for extracting medical association rules from healthcare data. By leveraging the integration of medical ontologies, and semantic association rule mining, we have addressed the limitations of traditional methods and achieved enhanced

Doi: https://doi.org/10.54216/JAIM.050105
Received: December 24, 2022 Revised: May 19, 2023 Accepted: August 06, 2023

interpretability and clinical relevance of the extracted rules. Through the application of our proposed approach on case study of mental health dataset, we demonstrated its superiority in providing more accurate and meaningful insights for healthcare professionals. The visualization of association rules as a scatter plot of confidence and support further facilitates the understanding of strong associations between medical concepts. Our approach enables better understanding of patients' attitudes towards mental health, identifies areas lacking sufficient mental health information, and highlights opportunities for improving mental health support.

Funding: “This research received no external funding”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

[1] Thamer M, El-Sappagh S, El-Shishtawy T, A semantic approach for extracting medical association rules. International Journal of Intelligent Engineering Systems, 13(3), 280-293, 2020.
[2] M Saber, Y Jitsumatsu, MTA Khan, A simple design to mitigate problems of conventional digital phase locked loop, Signal Processing: An international journal (SPIJ), 6(2), 65-77, 2012.
[3] M. Saber, Efficient phase recovery system, Indonesian Journal of Electrical Engineering and Computer Science (lJEECS), 5(1), 123-129, 2017.
[4] Mohamed Saber, A novel design and Implementation of FBMC transceiver for low power applications, Indonesian Journal of Electrical Engineering and Informatics (IJEEI), 8(1), 83-93, 2020.
[5] Ibrahim Abdelhameed, El-Sayed M. El-kenawy, Applications and datasets for superpixel techniques: A survey. Journal of Computer Science and Information Systems, 15(3),1-6, 2020.
[6] Taif Khalid Shakir, Ahmed N Al Masri, Single Valued Neutrosophic Set for Selection of Water Supply in Intelligent Farming. Journal of International Journal of Advances in Applied Computational Intelligence, 2 (2), 37-44, 2022.
[7] Alharbi AH et al., Diagnosis of Monkeypox Disease Using Transfer Learning and Binary Advanced Dipper Throated Optimization Algorithm. Biomimetics, 8(3),313, 2023.
[8] Irina V Pustokhina, Denis A Pustokhin, An Intelligent Neutrosophic Model for Evaluation Sustainable Housing Affordability. Journal of International Journal of Advances in Applied Computational Intelligence, 2 (2), 45-53, 2022.
[9] Blake C, Pratt W, Better rules, fewer features: a semantic approach to selecting features from text. In Proceedings 2001 IEEE International Conference on Data Mining, 59-66, 2001.
[10] Ben Abacha A, Zweigenbaum P, Automatic extraction of semantic relations between medical entities: a rule based approach. Journal of biomedical semantics, 2(5), 1-11, 2011.
[11] Shereen Zaki, Mahmoud M. Ibrahim, Mahmoud M. Ismail, Interval Valued Neutrosophic VIKOR Method for Assessment Green Suppliers in Supply Chain. Journal of International Journal of Advances in Applied Computational Intelligence, 2 (1), 15-22, 2022.
[12] Reem Atassi, Fuad Alhosban, Milan Dordevic, Computational Intelligence for Automatic Detection Cardiac Arrhythmia from ECG Signals: Taxonomy and Open Issues. Journal of International Journal of Advances in Applied Computational Intelligence, 2 (2), 16-26, 2022.
[13] M. M. E. Bahy, S. A. Ward, M. Badawi and R. Morsi, "Particle-initiated negative corona in co-axial cylindrical configuration. Annual Report Conference on Electrical Insulation and Dielectric Phenomena, Montreal, QC, Canada, 343-348, 2012.
[14] E. M. Shaalan, S. M. Ghania and S. A. Ward, Analysis of electric field inside HV substations using charge simulation method in three dimensional. Annual Report Conference on Electrical Insulation and Dielectic Phenomena, West Lafayette, IN, USA,1-5, 2010.
[15] Mohamed A. Abouelatta, et al. , Measurement and assessment of corona current density for HVDC bundle conductors by FDM integrated with full multigrid technique. Electric Power Systems Research, 199, 2021.
[16] Eid Marwa M, Fawaz Alassery, Abdelhameed Ibrahim, and Mohamed Saber, Metaheuristic optimization algorithm for signals classification of electroencephalography channels. Computers, Materials & Continua, 71(3), 4627-4641, 2022.

Doi: https://doi.org/10.54216/JAIM.050105
Received: December 24, 2022 Revised: May 19, 2023 Accepted: August 06, 2023