International Journal of Computational and Experimental Science and ENgineering (IJCESEN)

Vol. 9-No.2 (2023) pp. 141-149
http://dergipark.org.tr/en/pub/ijcesen

ISSN: 2149-9144

Copyright © IJCESEN

Research Article

MARCMV: Mining Multi-View Association Rules from Clustered Multi-Views

Shadi AL SHEHABI[1]*, Meltem Y. IMAMOGLU[2]

[1]University of Turkish Aeronautical Association, Faculty of Engineering, Departmant of Computer Engineering, 06790, Ankara-Turkey
* **Corresponding Author** : **Email:** salshehabi@thk.edu.tr - **ORCID:** 0000-0003-0545-9104

[2]University of Turkish Aeronautical Association, Faculty of Engineering, Departmant of Computer Engineering, 06790, Ankara- Turkey
Email: meltemyi@gmail.com - **ORCID:** 0000-0002-8574-4097

Article Info:

DOI: 10.22399/ijcesen.1292987
Received : 05 May 2023
Accepted : 12 June 2023

Keywords

Multi-view
Clustering
MultiSOM
MARCMV approach
Multi-view association rule

Abstract:

Data mining involves examining vast quantities of data to uncover valuable insights that can be utilized for making informed decisions and driving business objectives. The study focuses on the task of finding relationships between features belonging to two different views using multi-view model, and proposes a novel approach called MARCMV. This approach extracts multi-view association rules from different views of the same data set using multi-clustering neural model. The study finds that MARCMV outperforms conventional symbolic methods in terms of association rule quality and running time.

1. Introduction

The importance of automated knowledge extraction techniques has increased with the amount of data that is now readily available. Extraction of implicit, undiscovered, and possibly practical knowledge from data is the goal of knowledge discovery. Data mining, is also known as Knowledge Discovering in Databases (KDD) [1], is a process that analyzes huge amounts of structured or unstructured information to create compact and practical summaries of information. Knowledge extraction from databases frequently uses symbolic techniques like Apriori [2], FP-growth [3], and Charm [4], although these techniques have limitations, including the creation of redundant association rules. As a result, choosing the regulations is a difficult issue. Unsupervised learning models (clustering models), which are adept in separating and gathering data from one another, have thus been suggested as alternatives to traditional approaches for extracting useful knowledge. However, when used with high-dimensional data description spaces, clustering models can also result in imprecise knowledge [5]. A multi-view approach that splits the description space of data into several subspaces, a subspace is known as a view, has been proposed to address this problem [6,7,8]. Knowledge extraction can be more precisely performed than with the global approach since each view can be clustered and represented by a clustering model [9]. Using the MultiSOM paradigm, the clustered views can communicate with one another while maintaining the connections between the subspaces [10]. In contrast to traditional symbolic methods, the unique strategy proposed in this study for mining association rules based on the communication of the clustered viewpoints produces fewer but more valuable association rules. Finding connections between various viewpoints, or subspaces, that represents the dataset in a space of several dimensions (multi-dimensional space), is the goal. A clustering model that allows things to be grouped together based on similarities is required to create a view. This approach is crucial when working with large databases because it allows us to accurately extract knowledge from clusters by using the presenter of the data rather than scanning the entire database. In order to solve the issue of producing straightforward association rules based on the communication of clustered views between characteristics from various subspaces, this work provides a method for extracting numerical association rules between views. Overall, this research suggests a way for obtaining association rules from views using clustering models and shows that it produces more relevant association rules with fewer rules than traditional symbolic methods.

2. Related work

Association rule mining has been the subject of a variety of studies in the knowledge discovery in databases (KDD) sector. Symbolic techniques like Apriori, FP-growth, and Charm have been extensively applied to this. However, these methods have some limitations, including the generation of a huge number of redundant rules, making the rule selection process complex. Moreover, they cannot directly extract association rules between different views. Al Shehabi and Lamirel describe a method to extract straightforward association rules [9] using clustering models, where an association rule consists one feature in the preceding part and one feature in the following part. The space of high-dimensional data descriptions is first divided into a variety of subspaces, each of which is referred to as a view. A clustering model is then used to group each view. Unsupervised Neural Network Models are then utilized to preserve the relationship between the subspaces (views) by communicating between the clustered views. When compared to conventional symbolic methods, the suggested approach produces fewer but more beneficial simple association rules.
A technique to discover non-redundant association rules that explain the relationships between two perspectives of the same dataset was put out by Leeuwen and Galbrun [11] in 2015. They introduced translation tables that connect the two viewpoints and offer lossless translation between them. These translation tables are made up of bidirectional and unidirectional rules. They introduced three TRANSLATOR algorithms and a score that relies on the principle of Minimum Description Length. The evaluation of the method using real-world data demonstrates that it can recognize the two-view structure that is present in the data with a limited number of relations. The technique, however, is unable to distinguish between the significant and robust associations and the weak ones. It should be noted that there are very few works on this specific problem of extracting association rules between views using clustering models. Therefore, our proposed approach aims to fill this gap and provide a novel and effective solution

3. Theory Background

3.1. Multi-view notion

The principle of the multi-view building involves dividing the original space of dataset into various subspaces, which correspond to different subsets of features [12]. The union of these multi-views constitutes the original space of the dataset. The view subsets may overlap and can be associated with different feature subsets that correspond to specific dataset subfields. As long as they can be described using description vectors, the view model is adaptable enough to handle dataset descriptions from various media. The communication of the clustered views described below takes advantage of the inter-view model described above. It maintains an overall perspective of the interaction between the data while resolving the low-quality issue with a global clustering strategy.

3.2. Self-Organizing Map (SOM) clustering model

The SOM network architecture is founded on the hypothesis that representing data features is done by a self-organizing map (Fig. 1), this map is set up in a geometric grid [13].

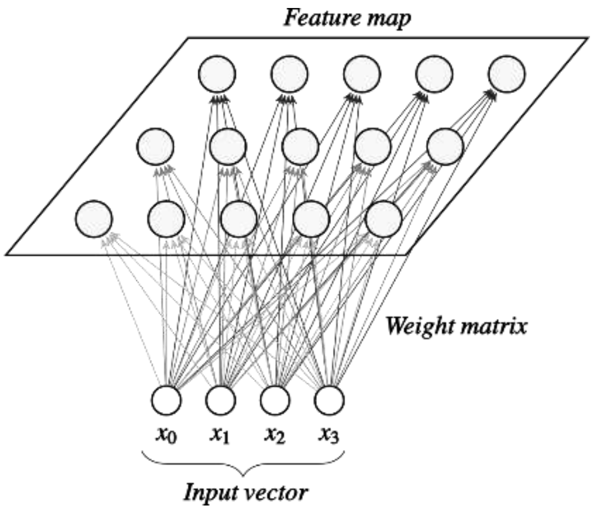


***Figure 1**. SOM Architecture*

This grid represents a two-dimensional area occupied by neurons as a high-dimensional data space. The two main steps of the SOM algorithm are picking a winner on the map and changing the weights of the chosen winner's neighbours. The training data is assigned to the clusters (neurons) of the map once the Self-Organizing Map algorithm is finished. The creation a Self-Organizing Map is not done directly and requires a number of learning processes, each map will be evaluated and compared with some other created maps to identify a trustworthy or ideal one.

3.3. Communication of Clustered Multi-view

The views are clustered using MultiSOM neural clustering model, which is an extension of the original SOM algorithm [10]. MultiSOM improves the quality of the views needed for data analysis by clustering each view using a single SOM map. This approach helps to minimize noise that may be present when only one space of data is clustered. Moreover, the overall analysis of the views is maintained by establishing connections between the clustered SOM maps. The superiority of MultiSOM's multi-views analysis, as compared to SOM, is evident in its ability to achieve accurate mining results. The clustered SOM maps are linked in such a way that it enables the identification of significant relationships between different topics belonging to various subspaces [14]. This connection process is carried out in three steps (fig. 2). The initial placement of the original activity is onto the data-related source cluster. Then, a two-stage transmission takes place from the source cluster to the target cluster. The first stage of transmission is done by activating a cluster in the source map then it is transmitted to the cluster in the target map through the their shared data.

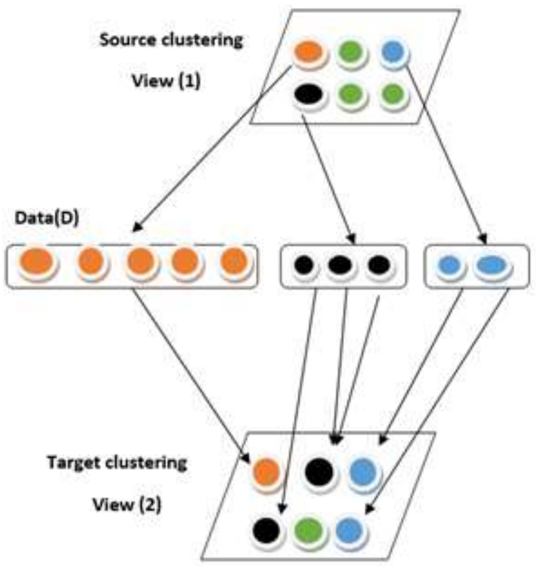


Figure 2. *MultiSOM Architecture*

3.4. Association Rule Mining

Consider a dataset D, consider the features of D are represented by P. Suppose the function f : P → D that associates a set of features p with the set of data objects that, at least, contain all the features of p (i.e., $f(p) = D_p$; where D_p is a set of objects that contains p). Suppose the function g : D → P that associates a set of objects with the set of features that are shared to all the data objects of d (i.e., g(d) = Pd; where Pd is a set of features that are shared to all the data objects of d).

Definition 1 (Support count (sup)): the term 'support count' of a set of features (p) refers to the number of data objects that contain the set of features.
Definition 2 (Support (s)): the term 'support' for a feature set (p) refers to the proportion of data items that contain that particular set of features.
Definition 3 (Frequent set of features): it is defined as a set of features whose support count is greater than or equal to a threshold called minsup.
Definition 4 (Closed set): A set of features (p) is a closed set if gof(p) = p [15].
Definition 5 (Association rule): it takes the form X→Y, where both X and Y are sets of features [16], and can be interpreted as an implication statement.
Definition 6 (Confidence (conf)): The confidence value of an association rule X → Y is determined by calculating the ratio of the support of the combined sets X and Y, to the support of set X. This can be expressed as conf(X→Y) = sup(X, Y) / sup(X) [16].
Definition 7 (Multi-view association rule): it is an association rule X → Y that the antecedent X is a set of features of a view and the consequent Y is a set of features of another view.
Definition 8 (Recall of a feature in a cluster): It measures to what extent a feature is associated with a single cluster [9].
Definition 9 (Precision of a feature in a cluster): It assesses the degree to which a feature is linked to the entirety of the data contained within a cluster [9].

4. Proposed Model

We propose an approach for mining multi-view association rules extracted between two clustered, known as MARCMV. In order to build a hybrid method for extracting useful knowledge, our strategy uses two criteria, recall (definition 8) and precision (definition 9), and combines a closed sets (definition 4) (using a symbolic method) with a cluster of a clustering method. Our method's algorithms each stand for a specific class of multi-view association rules, enabling deliberate rule selection. Importantly, these categories are not dependent on the confidence and support measures traditionally used to identify important rules. Under the first category (Fig. 3), when recall and precision for pertinent attributes are both 1, we extract the most important rules. Only attributes that are shared by all the data in the cluster and are present in the same cluster in both views, are included in these rules.

```
Algorithm 1: Category 1
Input: Two-view dataset D, first view clustering C, second view clustering C'
Output: set of association rules
    For each c ∈C
        Find P_c^*
        For each c' ∈C'
            if c ∩ c' ≠ ϕ
                Find P_c'^*
                % create A_cp^c the set of peculiar features from c such that their recall and
                % precision are 1
                A_cp^c = {a | a ∈P_c^* , precision(a) = 1, recall(a) = 1}
                % create B_cp^c' the set of peculiar features from c' such that their recall and
                % precision are 1
                B_cp^c' = {b | b ∈P_c'^* , precision(b) = 1, recall(b) = 1}
                % association rules extraction
                if |A_cp^c| ≠ 0 and |B_cp^c'| ≠ 0
                    ExtensionTest (A_cp^c, B_cp^c')
                End if
            End if
        End for
    End for
Function ExtensionTest (A, B)
  if Extent(A) ⊂ Extent(B) then A → B
    else if Extent(B) ⊂ Extent(A) then B → A
      else if Extent(A) ≡ Extent(B) then A ↔ B
        else if conf (A → B) > conf (B → A) then A → B
          else if conf (A → B) < conf (B → A) then B → A
            else A ↔ B
```

Figure 3. *First category of multi-view association rules*

```
Algorithm 2: Category 2
Input: Two-view dataset D, first view clustering C, second view clustering C'
Output: set of association rules
    For each c ∈C
        Find P_c^*
        For each c' ∈C'
            if c ∩ c' ≠ ϕ
                Find P_c'^*
                % create A_p^c the set of peculiar features from c such that their precision are 1
                A_p^c = {a | a ∈P_c^* , precision(a) = 1, recall(a) ≠ 1}
                % create B_p^c' the set of peculiar features from c' such that their precision are 1
                B_p^c' = {b | b ∈P_c'^* , precision(b) = 1, recall(b) ≠ 1}
                % association rules extraction
                if |A_p^c| ≠ 0 and |B_p^c'| ≠ 0
                    ExtensionTest (A_p^c, B_p^c')
                End if
            End if
        End for
    End for
```

Figure 4. *Second category of multi-view association rules*

With a precision of 1, the second category (Fig. 4) concentrates on finding relationships between all of the data in the clusters in both views. Each view should have these properties in at least two clusters. In the third category (Fig. 5), our objective is to find correlations between exclusive features that are seen only in clusters in both viewpoints. The exclusive features do not appear in multiple clusters, and our method guarantees a recall of 1. The fourth category (Fig. 6) consists of the following three actions in order:

- Pick out a portion of the dataset that appears in a cluster of the first clustered view and a cluster in the second clustered view.
- Recognize two sets of features, one from each view, that are common to the previously found portion of the data.

```
Algorithm 3 Category 3
Input: Two-view dataset D, first view clustering C, second view clustering C'
Output: set of association rules
    For each c ∈C
        Find P_c^*
        For each c' ∈C'
            if c ∩ c' ≠ ϕ
                Find P_c'^*
                % create A_p^c the set of peculiar features from c such that their recall are 1
                A_r^c = {a | a ∈P_c^* , precision(a) ≠ 1, recall(a) = 1}
                % create B_r^c' the set of peculiar features from c' such that their recall are 1
                B_p^c' = {b | b ∈P_c'^* , precision(b) ≠ 1, recall(b) = 1}
                % association rules extraction
                if |A_r^c| ≠ 0 and |B_r^c'| ≠ 0
                    ExtensionTest (A_r^c, B_r^c')
                End if
            End if
        End for
    End for
```

Figure 5. *Third category of multi-view association rules*

- Establish whether or not the two features subsets are closed sets of features. We extract correlations between the two subsets of attributes if they are closed.

To extract compact multi-view association rules, this category blends symbolic computation with numerical techniques.

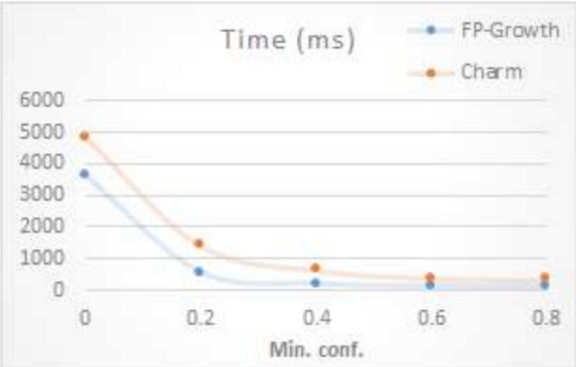
```
Algorithm 4 Category 4
Input: Two-view dataset D, first view clustering C, second view clustering C'
Output: set of association rules
    For each c ∈C
        For each c' ∈C'
            if c ∩ c' ≠ ϕ
                Find D_cc' the set of data d that are shared between c and c', D_cc' = c ∩ c'
                D_cc' = { d | d ∈c, d ∈c' }
                % create A_s the set of attributes from the first view which are shared with
                % all data of D_cc'
                A_s = {a | a ∈d, ∀d ∈D_cc', a ∈V_1}
                % create B_s the set of attributes from the second view which are shared
                % with all data of D_cc'
                B_s = {b | b ∈d, ∀d ∈D_cc', b ∈V_2}
                % association rules extraction
                if |A_s| ≠ 0 and |B_s| ≠ 0
                    if h (A_s)=A_s and h (B_s)=B_s
                        ExtensionTest (A_s, B_s)        % Closed itemsets
                    Else
                        ExtensionTest (A_s, B_s)        % non-closed itemsets
                End if
            End if
        End for
    End for
```

Figure 6. *Fourth category of multi-view association rules*

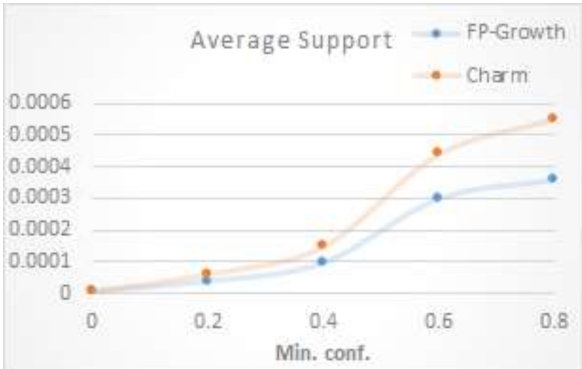
5. Experimental Results and Discussions

5.1. Mining multi-view association rules with symbolic techniques

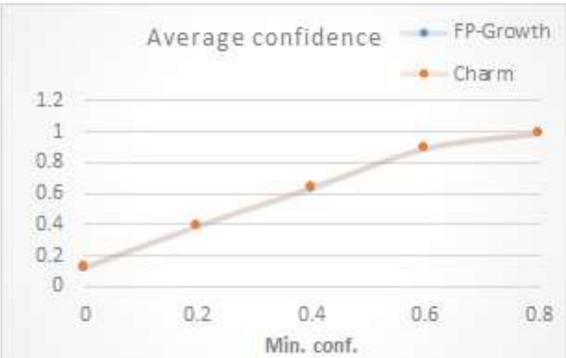
To assess our proposed approach, we will exhibit outcomes produced by two distinct algorithms for multi-view association rule mining on two datasets, namely Car and Tic-Tac-Toe, acquired from UCI Machine Learning Repository [17]. To conduct these tests, we must specify minimum support and confidence values. We calculated the minimum support values for all our tests as 1 divided by the number of rows in each dataset. We used minimum confidence values of 0.0, 0.2, 0.4, 0.6, and 0.8. Our analysis employed the Charm and FP-Growth algorithms, which are all contained in the SMPF tool, which is a data mining library available as open-source software [18].



(a)

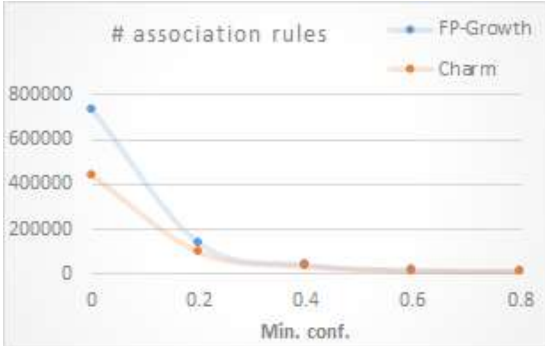


(b)



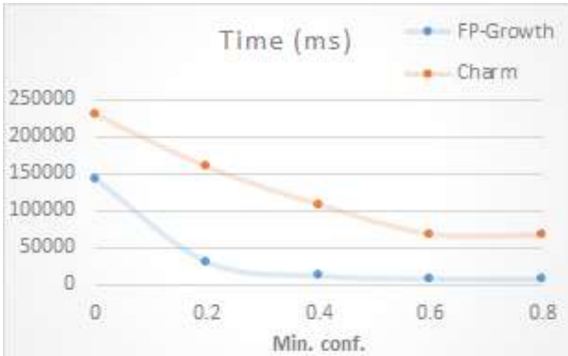
(c)

(d)

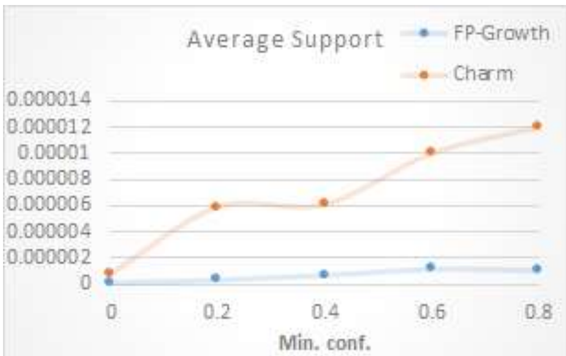


(e)

Figure 7. *Comparison of symbolic methods on the Car dataset*

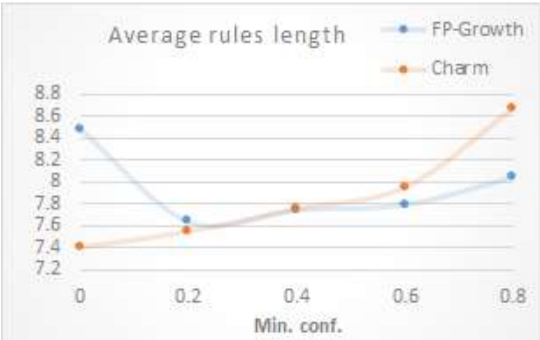


(a)

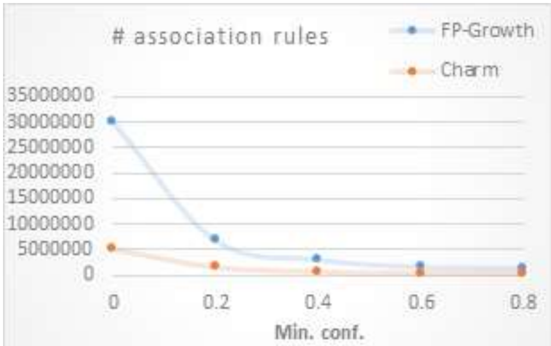


(b)

(c)



(d)



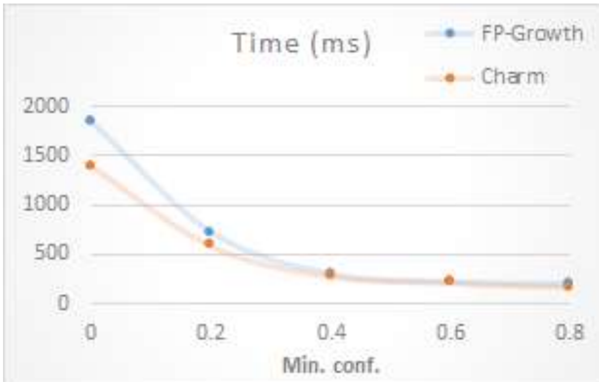
(e)

Figure 8. *Comparison of symbolic methods on the Tic-Tac-Toe dataset*

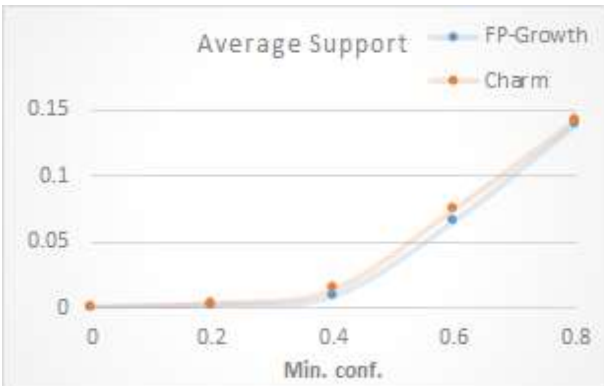
The FP-Growth method consistently outperforms the Charm approach, as seen in Figs. 7 and 8, with only a minor variation in average rule length for the two datasets. The average support values varied significantly in several experiments despite the average confidence differences between the two algorithms remaining equal due to the different numbers of association rules produced by each approach. Compared to FP-Growth method, which mines more redundant association rules and ineffective ones, the Charm method produces a comparatively fewer number of association rules, demonstrating its superior accuracy.
In our evaluation process, the second step, for extracting multi-view association rules, involves filtering the association rules obtained from the tests performed in the previous step. This is done by ensuring that the antecedent (left-hand side) of the rule contains features from the first view (source) and the consequent (right-hand side) contains features from the second view (target).
To do this, we created a straightforward application that evaluates all rules from each test via comparison and validation against a predefined condition. Depending on the dataset, the condition changes. For the car dataset, the requirement is that all attribute numbers on the first view's left side should fall between 1 and 15, while those on the second view's right side should do so between 16 and 25. We also take this condition's reverse order into account. The feature numbers of the first view's left side for the Tic-Tac-Toe dataset should be between 1 and 15, while those on the second view's right side should do so between 16 and 39. We also take this condition's reverse order into account.
Applying this filtering condition to each test from Figs. 7 and 8 leads in a significant decrease in the number of rules, as shown in Figs. 9 and Fig.10. As an illustration, the FP-Growth algorithm produced 732252 rules with a minimal confidence of zero, however after filtering, only 36322 rules were left.

5.2. Multi-View Association Rules Mining using MARCMV approach

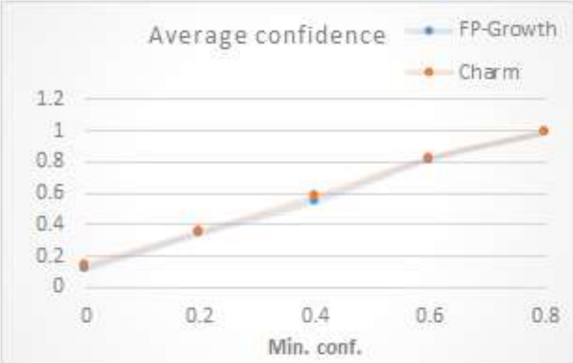
To extract multi-view association rules using our proposed approach, it is necessary to cluster each of



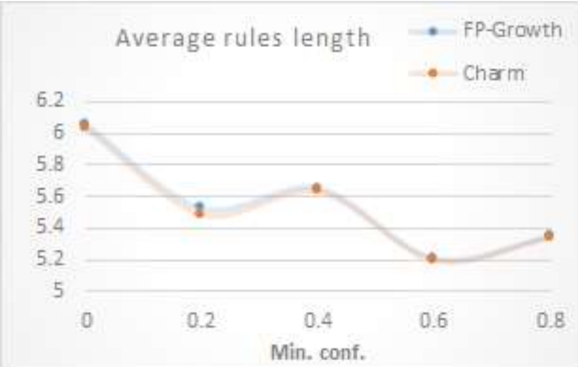
(a)



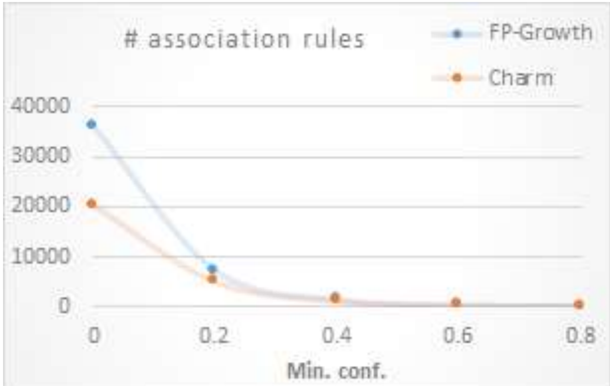
(b)



(c)

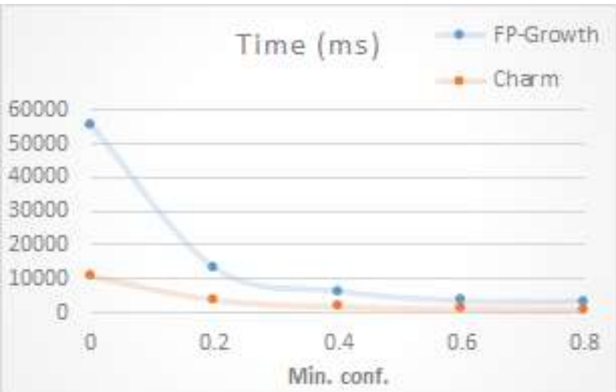


(d)

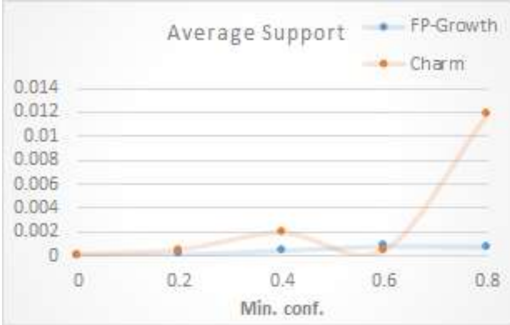


(e)

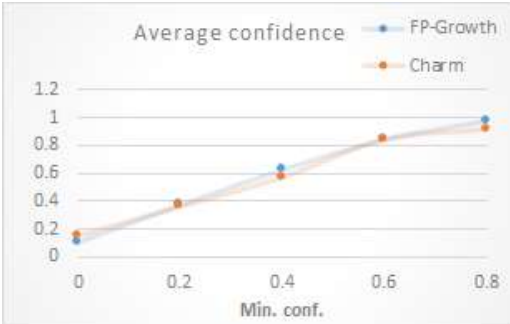
Figure 9. *Comparison of symbolic methods on the Car dataset after the filtering step*



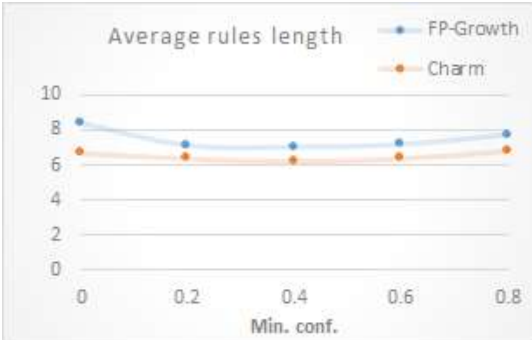
(a)



(b)



(c)



(d)

(e)

Figure 10. *Comparison of symbolic methods on the Tic-Tac-Toe dataset after the filtering step*

the datasets, Car and Tic-Tac-Toe, using MultiSOM with an optimal number of clusters. To determine this optimal number, we tested different numbers of clusters (2, 3, 4, 9, 16, and 25), with 20,000 iterations for each test. The clustering quality measures proposed in [19] were used to identify, for each dataset, the optimal number of clusters. Based on the

previously mentioned measures, three clusters are selected for each view of the Car dataset, also the same number of clusters as Car dataset are also selected for each view of the Tic-Tac-Toe dataset.
Multi-view association rules can then be extracted from the optimal clustering of each dataset using MARCMV.
A quick comparison of the outcomes in table 1 reveals that the quantity of association rules mined by the MARCMV approach is considerably less than what we discovered after using the Charm and FP-Growth algorithms on the experimental datasets. The averages estimated based on Figures 7 and 8 are shown for both algorithms in table 1. However, the times show the average duration of each algorithm as a whole, including the time for mining association rule and filtering them.
From experimental datasets, the number of multi-view association rules mined by MARCMV approach is considerably lower than the those extracted by symbolic algorithms. This shows that mining, from two views, association rules yields beneficial and easily understandable ones, while also excluding redundant and unhelpful rules. For example, only nine rules were produced when our method was applied to the Car dataset, as opposed to 5,446 and 9,164 rules generated by the Charm and FP-Growth algorithms, respectively. Similarly, for the Tic-Tac-Toe dataset, our approach generated only nineteen rules, as opposed to 31,804 and 121,069 rules generated by the Charm and FP-Growth algorithms, respectively.
In addition, we observed that the MARCMV approach was faster than the symbolic methods in terms of running time. For the Car dataset, the average time was 1.3 seconds, while for the Tic-Tac-Toe dataset, it was 1.2 seconds. In contrast, the average times for the symbolic methods were 2 seconds, 1.5 seconds, approximately 129 seconds, and 58 seconds, respectively.

Table 1. *Comparison results: (a) Car, (b) Tic-Tac-Toe.*

Algo.	Time	Support	Confidence	Rule Length	# Rules
Charm	2044.8	0.0472	0.577	5.53	5446
FP-Growth	1581.6	0.0432	0.562	5.55	9164
MARCMV	1362.9	0.1659	0.4593	2.25	9

(a)

Algo.	Time	Support	Confidence	Rule Length	# Rules
Charm	129552.6	0.003	0.572	6.49	31804
FP-Growth	57954.4	0.0004	0.587	7.46	121069
MARCMV	1258.8	0.1474	0.4588	2.15	19

(b)

When generating association rules using symbolic methods, a minimum confidence value must be set to determine the importance of the association rule. All association rule mined with a confidence level lower than minimum confidence will be removed. However, from the experiments, we discovered that setting a low minimum confidence value has three major drawbacks. The first one is the time-consuming, particularly for huge datasets. Then, it mines a large number of redundant and unhelpful association rules, which makes it difficult to use these association rules in any decision-making process based on two views. The third one is consuming a lot of resources, making it infeasible to apply symbolic methods, especially when working with large datasets or datasets with a high number of features. These drawbacks are mitigated by giving additional temporary storage resources.
Working on clustered views of a dataset ensures that data with high similarities are grouped together while weak relationships among variables are eliminated, and all variables within each group are given equal importance. As a result, the MARCMV approach does not rely on minimum confidence values in generating association rules. Thus, it becomes more achievable to implement, regardless of the size of the dataset.

6. Conclusion and future work

Due to the swift expansion of data, there is a need to develop techniques for extracting useful information from large datasets. Symbolic methods, which are commonly used for knowledge extraction, have drawbacks in generating a large number of rules, including weak and redundant ones. In this paper, a novel approach called MARCMV was proposed to mine multi-view association rules via clustered datasets. The strategy focuses on detecting relations between several views of the data using MultiSOM clustering. This helps to group items with strong relationships and exclude weak and not useful association rules. The MARCMV approach is more feasible to implement regardless of dataset size compared to symbolic methods.
We plan in the future work to add more important and correalted multi-view association rules by adding more catgories to our approach.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Han, J., Pei, J., & Tong, H. (2022). Data mining: concepts and techniques. Morgan kaufmann.

[2] Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).

[3] Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data mining and knowledge discovery*, 8; 53-87.

[4] Zaki, M. J., & Hsiao, C. J. (2002, April). CHARM: An efficient algorithm for closed itemset mining. In Proceedings of the 2002 SIAM international conference on data mining (pp. 457-473). Society for Industrial and Applied Mathematics.

[5] Kanhere, S., Sahni, A., Stynes, P., & Pathak, P. (2021, January). Clustering Based Approach to Enhance Association Rule Mining. In *2021 28th Conference of Open Innovations Association (FRUCT)* (pp. 142-150). IEEE.

[6] Tang, C., Zheng, X., Liu, X., Zhang, W., Zhang, J., Xiong, J., & Wang, L. (2021). Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Transactions on Knowledge and Data Engineering*, *34*(10); 4705-4716.

[7] Zhao, J., Xie, X., Xu, X., & Sun, S. (2017). Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, *38;* 43-54.

[8] Li, Z., Tang, C., Zheng, X., Liu, X., Zhang, W., & Zhu, E. (2022). High-order correlation preserved incomplete multi-view subspace clustering. *IEEE Transactions on Image Processing*, *31;* 2067-2080.

[9] Lamirel, J. C., & Al Shehabi, S. (2005, September). Efficient Knowledge Extraction using Unsupervised Neural Network Models. In *5th Workshop On Self-Organizing Maps-WSOM 05*.

[10] Lamirel, J. C. (2002, May). MultiSOM: a multimap extension of the SOM model. Application to information discovery in an iconographic context. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290)* (Vol. 2, pp. 1790-1795). IEEE.

[11] M. van Leeuwen and E. Galbrun, “Association discovery in two-view data,” *IEEE Trans. Knowl. Data Eng.*, 27(12);3190–3202, 2015.

[12] Polanco, X., François, C., & Lamirel, J. C. (2001). Using artificial neural networks for mapping of scienceand technology: A multi-self-organizing-maps approach. *Scientometrics, 51;* 267-292.

[13] Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V., & Saarela, A. (2000). Self organization of a massive document collection. *IEEE transactions on neural networks*, *11*(3); 574-585.

[14] Lamirel, J. C., Shehabi, S., Francois, C., & Polanco, X. (2004). Using a compound approach based on elaborated neural network for Webometrics: an example issued from the EICSTES Project. *Scientometrics*, *61*(3); 427-441.

[15] Pasquier, N., Bastide, Y., Taouil, R., & Lakhal, L. (1999). Discovering frequent closed itemsets for association rules. In *Database Theory—ICDT’99: 7th International Conference Jerusalem, Israel, January 10–12, 1999 Proceedings 7* (pp. 398-416). Springer Berlin Heidelberg.

[16] Tan, P. N., & Kumar, V. (2000). Interestingness measures for association patterns: A perspective.

[17] Dua, D., & Graff, C. (2017). UCI machine learning repository.

[18] Fournier-Viger, P., Lin, J. C. W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., & Lam, H. T. (2016). The SPMF open-source data mining library version 2. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part III 16* (pp. 36-40). Springer International Publishing.

[19] Lamirel, J. C., Francois, C., Shehabi, S. A., & Hoffmann, M. (2004). New classification quality estimators for analysis of documentary information: application to patent analysis and web mapping. *Scientometrics*, *60*(3); 445-562.