Självständigt arbete på avancerad nivå

Independent degree project – second cycle

Master of Science in Computer Engineering:
Computer Engineering

3D matrix-based visualization system of association rules

Biying Wang

Mittuniversitetet
MID SWEDEN UNIVERSITY

MID SWEDEN UNIVERSITY

The Department of Information and Communication Systems

Examiner: Tingting Zhang, tingting.zhang@miun.se

Supervisor: Mehrzad Lavassani, mehrzad.lavassani@miun.se

Author: Biying Wang, biwa1500@student.miun.se

Degree programme: Master of Science in Computer Engineering, 120 credits

Main field of study: Computer Engineering

Semester, year: Spring, 2016

Abstract

With the growing number of association rules, it becomes more and more difficult for users to explore interesting rules due to its nature complexity. Studies base on human perception and intuition show that graphical representation could be a better illustration of how to handle data by using the capabilities of the human visual system to seek information. The 3D matrix-based approach visualization system of association rules called 3DMVS was implemented in present study. The main visual representation employed the extended matrix-based approach with rule-to-items mapping to general transaction data set. A novel method merging rules and assigning weight is proposed to generate new rules to reduce the dimension of the association rules, which will help users to find more important items in the new rule. Additionally, several interactions such as sorting, filtering, zoom and rotation, facilitate decision makers to explore the rules they are interested in various aspects. Finally, various evaluation techniques have been employed to assess the system from a logical reasoning point of view.

Keywords: 3D, matrix-based approach, visualization system, association rules

Table of Content

1	Introduction	1
1.1	Motivation and problem statement	1
1.2	Overall aim	2
1.3	Research question.....	2
1.4	Research hypothesis.....	4
1.5	Method	4
1.6	Scope	5
1.7	Contribution.....	6
1.8	Outline	7
2	Theory.....	8
2.1	Association rule	8
2.1.1	Support.....	8
2.1.2	Confidence	9
2.2	Association rule visualization	10
2.2.1	Table-based view	10
2.2.2	Graph-based view.....	12
2.2.3	Parallel coordinates	14
2.2.4	Scatter plot	16
2.2.5	Matrix-based view	17
2.3	Reducing the rules.....	19
3	Method	22
3.1	Data Set	22
3.2	Overview	22
3.2.1	Extending matrix-based view in 3D	23

3.2.2	Merging rules and assigning weight	24
3.3	Interactivity	26
4.	Implementation of the visual system	27
4.1	Introduction of the system	27
4.2	Mapping	28
4.3	Selection and presentation	29
4.4	Interactivity	30
4.4.1	Sorting and filtering	31
4.4.2	Detail overview	34
4.4.3	Zoom and rotation	35
5	Evaluation and discussion	37
5.1	Usability testing	37
5.1.1	Completeness	37
5.1.2	Spatial organization	38
5.1.3	Codification of information	38
5.1.4	State transition	39
5.2	Interaction testing	39
5.2.1	Orientation and help	39
5.2.2	Navigation and querying	40
5.2.3	Data set reduction	40
5.2	Discussion of the visual ability	41
5.2.1	Comparing the 3D matrix visualization with other methods	41
5.2.2	Comparison of original visual graph with weighted one	43
6	Conclusion and future work	44
6.1	Ethical considerations	45
	References	46

1 Introduction

1.1 Motivation and problem statement

Data mining is a process where a variety of data analysis tools are used to discover patterns and relationships in data that could be used to make valid predictions [1]. One of the most important and well-known techniques of data mining is association rule mining. This technique mainly aims to look for interesting correlations, frequent patterns, associations, information structures among sets of items or objects in large transaction databases and/or other data repositories. Moreover, the analysis of association rules has contributed to various areas, including market basket analysis, merchandise stocking, insurance fraud investigation and climate prediction.

In spite of its great success in demonstrating the relationship between items, it is not easy for users to find the rules they are most interested in from a large number of extracted rules. Since mining association rules often result in a large amount of rules all of which are listed in one file, the researchers are given the task to go through the results to discover which are of interest. Increasing the threshold of minimum support or minimum confidence is a good way to reduce the numbers of the rules. However, there is a risk this will decrease the possibility of finding meaningful rules that do not have very high support or confidence. By setting low thresholds in practice, users have to seek for interesting ones in a corpus consisting of hundreds or even thousands of resultant rules, which could be difficult and time-consuming.

Based on human perception and intuition, graphical representation could be more illustrated to seek information by using the capabilities of the human visual system[2]. Such visual representations have been developed since the 1990s, allowing users to evaluate and interpret what rules from a large number of discovered rules. Lots of visualization techniques have been employed to provide a graphical representation of the results. Recently, related studies demonstrated support for the analyses

of the association rules, however, it should be noted that illustrating association rules graphically is not easy when applying visualization methods. The main reason for this problem is that the multiple relational nature of association rules is difficult to show in an explicit way especially when there are a large quantity of rules or when the rules have many interrelated items. In addition, the interestingness measures such as support and confidence which are important aspects of association rules, should be represented along with the relations. Thus, increasing the complexity of visualizing a large number of rule.

Although several methods have been applied to visualize the association rules, many of these methods have failed to manage the large number of rules. As a result, a large number of data is often displayed in a single image without further details about the rules, which makes it difficult for knowledge managers to evaluate and interpret the rules. What makes it difficult is the screen clutter and lack of suitable design to display a large number of rules and relations.

1.2 Overall aim

This study attempts to implement a 3D visualization system of association rules based on a matrix-based representation which gives users an overview of whole sets of rules and relations intuitively. Generating new rules with the new measure weight of an item is proposed to reduce the number of rules. Basic functions of interactivity are also provided. The purpose of this system is to provide a convenient and understandable tool to facilitate for data managers to explore interesting rules.

1.3 Research question

Since the main factors of association rules need to be understandable and interpretative in the graphical representation, various visual techniques have been studied. One of the popular approaches is the matrix-based method, which usually maps the item-to-item relations between items and shows the measures by

color or shape. However, considering most of the resultant rules have multiple antecedents in reality, item-to-item mapping involves a clutter problem.

Instead of item-to-item mapping, this 3D visualization system employed the rule-to-item mapping method representing the relations between items. Could rule-to-item mapping be a way to tackle the clutter problem? By increasing one dimension, could the association rules become more comprehensible for users through 3D matrix-based representation?

The visualization of association rules are facing a challenge in terms of an increasing number of association rules which are becoming more difficult to be manage on a single screen. There is a method that cloud be used to improve the representative ability of the visual tools, but it shows little improvement because it struggles to handle the various features of the rules and depict all rules in one graph at the same time. Reducing the number of discovered rules could be an another alternative, but the accuracy and interesting rules could be lost as a consequence. A new method to generate new rules with weight was proposed in present study to decrease the number of rules. Does the new method reduce the number of rules? Does this method lose any accuracy and does it retain the degree of representative ability?

Although graphical representation could facilitate exploration of interesting rules for the users, the most important step is supporting user interactivity with the visualization system. Sorting and filtering of the main measures of the association rule is the basis of exploring the rules with higher support or confidence. A real-time interactivity is highlighting selected items and displaying the details in a visual system at the same time. Are the users' interactivity appropriate for exploration? Does this mechanism cope with the display of details without losing the overview?

1.4 Research hypothesis

The 3D visualization system employed uses the rule-to-items mapping method instead of item-to-item, and only focuses on the association rules with unique consequence. The rule-to-item mapping method mainly solves the clutter problem when the rules have multiple antecedents. In addition, the graphical representation should show the relations between the items including important measures, which is useful for exploring interesting rules.

A new measure of weight was put forward to reduce the number of association rules. Merging the rules with the same consequence to one and assigning weight to each item could maintain the completeness of the original sets of rules and reduce the number of discovered rules. However, the worst case scenario is when all the resultant rules have different consequences. Since a merging method does not drop any association rules and provides a novel way to explore the interesting items, this kind of visualization could have the same representative ability as the original data sets.

For a good visualization system, it is necessary to investigate interactivity. The sorting and filtering function of the association rules were offered to facilitate for users to find useful rules with higher support and confidence. Highlighting selected item and displaying the details in a visual system at the same time could also make the whole sets of rules more clear to users, which makes it easier for users to understand the basic information of a specific rule. In addition, weight filters could easily select items that are more important for the consequence of the rule.

1.5 Method

For decades, visualizing association rules have been studied and many researchers have proposed different methods to obtain an overview of a large number of rules. Many methods have been studied to visualize the association rules, such as table-based view, graph-based view, parallel coordinates, scatter plot and matrix-based view.

Comparing all visualization methods of association rules, this study proposes a modified matrix-based method by taking advantage of 3D view, which can handle a large number of rules and interpret rules directly. Instead of using 2D matrix to display item-to-item association rules, the modified matrix-based view depicts the rule-to-item relation, which can maintain an overview of the resultant rules as well as the relation between each item of the rules.

Considering the large number of rules, this study proposes a merging method to reduce the dimension of the graph. By assigning the weight to each item users will get an overview of all important items, which enables comparing the items within the same generated rules as well as the consequences of the different rules.

In order to provide a more interactive system to users, 3DMVS offers users functions like filtering, zooming and sorting of support, confidence and weight. These functions could help users explore the discovered rules to find interesting rules. The details of the rule selected will show in text with a click of the mouse.

1.6 Scope

Data visualization is an important branch of data mining, which has two different areas of application. Model visualization is the process of using visual techniques to make the discovered knowledge understandable and interpretable by humans while exploratory data analysis aims to identify interesting and previously unknown patterns [1].

Association rule visualization is employed to offer users an easier way to understand and interpret the relation between rules. This study only focuses on visualizing the resultant rules while ignoring the process of association rules generation.

Although many studies take dynamic association rules into account [3] [4] [5], this study only focuses on statistic association rules, where each rule has specific interestingness measures. In order to adapt the new merging method, the association rules used in this study consist of multiple antecedents with unique consequences rather than only considering Boolean attributes in the antecedent and consequences of the rules [2]. Since the new method merging original association rules is based on unique consequences, 3DMVS mainly pays attention to transaction data sets where most of the resultant rules have the same consequence. From usability and interaction aspects, the evaluation of the 3DMVS were discussed. Moreover, 3DMVS only supports those files with the xls filename extension.

1.7 Contribution

3DMVS could be a way to visualize the association rules with multiple antecedents. 3DMVS not only depicts the relations between items but also the information of each item. Displaying the matrix-based view in 3D makes the visual graph more vivid and clear for investigation.

The new method of merging the original rules with a new assigned measure weight reduces the number of the discovered rules, which simplifies the problem of the visualization system. Visualizing the merged association rules with weights provides an overview of all important items of the rules, making it easier for users to find a larger number of interesting items.

Interactivity is a necessity for an exploration system. 3DMVS provides basic sorting and filtering of interestingness measures including support, confidence and weight. It also provides another real-time interactivity: when a user wants to learn more about a specific rule, a mouse-click is enough to show details about the selected rule, which keeps a balance between displaying the overview and showing the details.

1.8 Outline

The remainder of this paper is organized as follows:

Chapter 2 defines fundamental and important concepts of association rules and visualizations, compares previous work on visualizations of large number of rules. In Chapter 3, the design and fundamental features of 3DMVS are described. This is followed by an explanation of the implementation and how this system can be useful for finding interesting rules. It also discusses the impact on the visualization of a large number of rules in Chapter 4. In Chapter 5, the evaluation of such a system is discussed. Finally, the study is concluded with a summary of the primary contributions of this work and an outline of future work in Chapter 6.

2 Theory

This section briefly introduces work related to 3DMVS. The concepts of association rules are introduced first, followed by several well-known techniques used for visualization of the association rules. Next, diversity methods used to reduce the number of association rules are compared. Finally, an approach is proposed.

2.1 Association rule

Let $I=\{i_1, i_2, \dots, i_m\}$ be a set of m distinct attributes, T be the transaction that contains a set of items such that $T \subseteq I$, and D is a database with different transaction records T_s . An association rule is an implication in the form of $X \rightarrow Y$ where $X, Y \subset I$ are sets of items called item sets, and $X \cap Y = \emptyset$. X is called antecedent or left-hand-side (LHS) while Y is called consequent or right-hand-side (RHS), the rule means X implies Y [4][6].

There are two measures of association rules, support (s) and confidence (c). As mentioned before, an association rule is applied to discover the interesting frequent patterns from a large data set for users. This means the possibility of occurrence should be greater than or equal to the given threshold as well as the two thresholds that are minimal support and minimal confidence.

2.1.1 Support

Support of the association rule $X \rightarrow Y$ is defined as the percentage of records that contain $X \cap Y$ to the total number of records in the database. The count does not consider the quantities of the items but the records covered by all items presented in the rule. Hence, support is also called coverage. Support is calculated by the following equation:

$$\text{Support}(X \rightarrow Y) = \frac{\text{recordscontains}(X \cap Y)}{\text{total recordsof transaction in } D} \quad (1)$$

Based on the above definition, the support of an item is a statistical value calculated from the records of the database. If it assumes that the support of an item is 1%, it means only 1 percent of the records in the transaction dataset contains this item. Obviously, the items with low support are not attractive enough for most customers. The retailer will not take interest in such items, but rather the items with high support. Before the process of mining association rules, algorithms such as Apriori [7] and FP-growth [8] require users to specify the minimum support as a threshold, which means that the users are only interested in certain association rules with support greater than or equal to the given threshold. Nevertheless, sometimes association rules generated from the transaction dataset are still important, but the support is lower than the given threshold. For example, some products are expensive and they are not purchased so often. As a result, the supports of association rules including these products are lower than threshold. But association rules of this type are as significant as other frequently brought as items to the retailer.

2.1.2 Confidence

The confidence of an association rule is defined as the percentage of the records containing all the items in the rule compared to the number of records only containing the antecedents. If the percentage is greater than the threshold of confidence, rules will be generated as a result. The following is the formula of confidence:

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \cap Y)}{\text{Support}(X)} \quad (2)$$

With respect to the transaction dataset, the confidence of the rule $X \rightarrow Y$ is the proportion of transactions that contain X as well as Y ; and the rule's specified minimum confidence is predefined by users

In addition, confidence can be interpreted as an estimate of the conditional probability, the possibility of finding the RHS of the rule in transactions under the condition that these transactions also contain the LHS [9] [10].

2.2 Association rule visualization

Mining association rules always results in a large number of rules [11] [12], including support and confidence, which should be organized in such a way making them comprehensible to users. By taking advantage of human perception and intuition, the visualization of the association rules could be more comprehensive for executives and other knowledge workers to dramatically improve their ability to grasp information found in the data.

The visualization of association rules needs to make the relationship between the antecedents and consequence more easily accessible to users. In general, users lack in knowledge of data mining. The visualization also needs to clearly depict the two main measures support and confidence at the same time.

The visualization of association rules has been studied for more than forty years [13]. Several techniques have been developed to visualize association rules; they can be categorized into five different groups: table-based, graphs, parallel coordinates, scatter plots and matrix.

2.2.1 Table-based view

Tables have been used as a tool to store data since the 1970s [14], but it is also an alternative way for data visualization. The table-based view is the most common and traditional way of representing association rules.

Based on the definition of table-based visualization, each variable of the association rule can be represented in one column, including rule IDs, items in the LHS and RHS, support and the confidence measures.

In this illustrative technique, each row represents a specific association rule. An example of a table-based visualization of the association rules is shown in Table 2.1. It is useful for users to consider the features and relations of the specific rules, even with querying and sorting mechanisms. Due to its simplicity, the table-

based view has remained popular in visualizing association rules [3] [15]. However, with a growing number of rules, it is not easy for users to look for interesting items, rules, and relations within the table because of its textural representation.

Table.2.1 Sample of table-based visualization

ID	LHS	RHS	support	confidence
1	{CityStay=nonurban,Cyber_Victim=No}	{Cyberbullying=No}	0.332	0.931
2	{TrainingUsageInternet=No,FeelSafeUseInternet=Enough,Cyber_Victim=No}	{Cyberbullying=No}	0.257	0.897
3	{UsageInternetPersonalCapabilities=b,FeelSafeUseInternet=Enough,TheftPersonalData=No,Cyber_Victim=No}	{Cyberbullying=No}	0.252	0.895
4	{FeelSafeUseInternet=Enough,TheftPersonalData=No,Cyber_Victim=No,descr_YearsStartUseInternet=(5,10]}	{Cyberbullying=No}	0.292	0.894
5	{CityBirth=nonurban,Cyber_Victim=No}	{Cyberbullying=No}	0.287	0.892
6	{CityBirth=urban,FeelSafeUseInternet=Enough,TheftPersonalData=No,Cyber_Victim=No}	{Cyberbullying=No}	0.287	0.892
7	{TrainingUsageInternet=No,FeelSafeUseInternet=Enough,Cyber_Victim=No}	{TheftPersonalData=No}	0.272	0.948
8	{UsageInternetPersonalCapabilities=b,FeelSafeUseInternet=Enough,TheftPersonalData=No,Cyberbullying=No}	{Cyber_Victim=No}	0.252	1
9	{CityStay=urban,CityBirth=urban,TheftPersonalData=No,Cyberbullying=No}	{Cyber_Victim=No}	0.272	1
10	{CityBirth=urban,FeelSafeUseInternet=Enough,TheftPersonalData=No,Cyberbullying=No}	{Cyber_Victim=No}	0.287	1

2.2.2 Graph-based view

Another widely used technique for the visualization of association rules is the graph-based view [15] [16]. The visualization of association rules in the form of graphs, vertices in the graph typically represent the items or item-sets, and edges indicate the relation between the LHS and the RHS. Interesting measure support is usually added to the graph by increasing the area of a vertex, while the confidence is depicted by color or width of the arrows displaying the edges. For example, in Figure 2.2, one resulting association rule is that the items represented by green 6 and green 8 imply the items represented by red 1 and red 9. The example rule has a darker and wider edge, which indicates the higher support and confidence of the rule. But how much higher is difficult to tell.

Although this kind of visualization represents association rules in a more vivid and concise manner than table-based view, the graph-based visualization has a cluttering problem and is difficult to interpret when the number of items and association rules increase [17]. In addition, since there are a large number of rules and items, the most important disadvantage of a graph-based view is finding a specific item or rule in the graph. Thus, the graph-based view is only available for small sets of rules.

Figure 2.2 shows a subset of the association rules in the graph-based view, connecting the antecedents and consequences through the blue vertex. The edges in the graph encode the support (darkness of lines) and confidence (width of lines) of the rules. While in Figure 2.3, around 10,000 association rules are displayed, which will make it more difficult for the data manager to explore its information to any greater extent.

Fig 2.2 Example of the graph-based view of association rules (5 rules) [2]

Fig 2.3 Example of the graph-based view of association rules (9,785 rules)
[17]

2.2.3 Parallel coordinates

Parallel coordinates [18] [19] [20] [21] as a visualization technique is another way of representing individual high-dimension data in two dimension [15]. In parallel coordinates, each dimension of the data is displayed separately as a parallel coordinate line (x-axis) and each item is represented as a multiple segment line that intersects the parallel coordinates.

In the parallel coordinate view, each item in the dataset provides a parallel coordinate which means the number of coordinates is the same as the sum of the items in the LHS and RHS of the resulting rules, and the association rules are drawn by the connection of related items within the parallel coordinate structure [22]. As shown in Figure 2.3, the top line on the top depicts a rule(soda, popcorn) $\rightarrow$ (salty snack), where the thickness represents the confidence.

A number of visualization methods for association rule have been developed based on this technique [22] [23]. Visualizing association rules based on parallel coordinates [22], the items were arranged by item group in order to display the items belonging to the same group together. To avoid the cluttering problem, the items within the same group were listed in descending according to the support. In [23], an association rule is visualized by connecting items in the rule, one item on each parallel coordinate with continuous polynomial curves. Each coordinate can also be used to visualize an item taxonomy tree.

However, it is obvious that such methods have overlapping lines when there is a large number of rules, which also creates ambiguity when multiple association rules are included in the same item.

Fig 2.4 Example of visualization of association rules by parallel coordinates
(10 rules) [24]

Fig 2.5 Example of visualization of association rules by parallel coordinates (774 rules) [17]

2.2.4 Scatter plot

A straightforward visualization technique for association rules is scatter plot, which is graphical display of sets of data in Cartesian coordinate system. One coordinate represents the horizontal distance and the other represents the vertical distance (dependent variable) of a data point from the coordinate axis [25].

Instead of displaying the items, the scatter plot shows the dispersion of two important measures of the association rules with the x-axis representing the support and y-axis depicting the confidence [26].

Figure 2.6 introduces a special version of a scatter plot called a two-key plot. Here, support and confidence are used for the x-axes and y-axes and the color of the points is used to indicate "order," i.e., the number of items contained in the rule.

Fig 2.6 Example of scatter plot for the visualization of association rules [24].

The advantage of the scatter plot technique is that it works well for very large sets of association rules and it is easy to implement zooming into the plot. However, there is not enough space to display the labels of the rules, which is time-consuming when exploring specific rules out of a large number of rules.

2.2.5 Matrix-based view

The matrix-based view is a good solution for the visual representation of the high dimensional nature of association rules. Within the matrix, rows display the RHS items and columns depict the LHS items, which means techniques of this type organize the antecedent and consequent item-sets on the x-axes and y-axes respectively [5] [27] [28] [29]. Support and confidence are shown in different colors and shapes at the intersections between the LHS items and RHS items. If no rule is available for an antecedent/consequent combination the intersection area is blank. As shown in Fig 2.7, light green is used for selected LHS and the dark green is for unselected LHS. While the color red is used for the RHS. The grey-scale colors are used to code different values of support.

Although this kind of technique provides an effective overview of association rules, there are drawbacks. Since the matrix becomes equally large as the number of the association rule when the items increase. The multiple antecedents or consequences will have conflict in the same intersection, it becomes difficult to decode [30]. Furthermore, it is not easy to find the relationship between the items even when the data-set is small.

Fig 2.7 The matrix view shows the maximum support of the association rules (encoded as dark grayscale) at the intersection of the LHS (left) and RHS (top) items [2]

Ong et al. [5], presented another method of mapping the two measures support and confidence to the rows and columns to overcome scalability issues. Each cell in this model depicts a group of rules with same measures. Nevertheless, the issues such as identifying specific relations and decoding the visual display are still existing.

Another improved matrix-based visualization view of association rules is the use of 3D bars at the intersection [5] [24]. An example is shown in Fig 2.8, the drawbacks are that the number of rows/columns depends on the number of unique consequents/antecedents in the set of rules and the size of the colored squares or the 3D bars become very small and hard to see due to the multiple antecedents of the discovered rules.

Fig 2.8 An example of the matrix-based visualization with 3D bars [31]

Compared to the matrix-based visualization with bar chart in 3D for text mining [32], a novel method was proposed that represents a topic by columns and each item belonging to the topic in the finding rules is displayed on rows in this column. Different colors are employed to distinguish the LHS items and the RHS items. Support and confidence is depicted by the bar chart at the back of the matrix. It not only offers users a more precise portrayal of the relation between LHS items and RHS items, but it is also easy for the users to understand and interpret. However, when the items and the rules increase by a lot, this method is still not scalable.

2.3 Reducing the rules

However, all of visualization techniques have a common weakness. The methods do not work well with a large number of rules because of the screen clutter problem. One way to solve this is to reduce the number of rules discovered, which can help tackle this problem on a large scale.

Reducing redundant resulting rules was proposed as a good method to decrease the number of rules, which is done in two steps, first the rules that have the similar meaning are identified and then eliminated them [33]. Cover methods processed using clustering and selecting participants within clusters have been studied by many researchers, such as cover structure [35] [36], rule clustering[34], instance cover [34] [35] [37].

For cover structure method, the [36] formed cluster of rules based on the structure distance of antecedent. A cover structure is to select the most representative rule from each cluster. While [35] define a structure cover as a subset of the original set of rules. Such rule cover contains the most general rules of the original rule set.

The rule clustering was processed by selecting highly ranked association rules one by one and forming clusters of objects covered by each rule until all the objects in the data set were covered [34].

In the instance cover in [37], the author employed hierarchical clustering to partition the initial rule set into thematically coherent subsets, which was able to group and summarize large sets of association rules according to the items contained in each rule.

However, when the discovered rules have been reduced by the methods above, there will be still a large number of rules considering the extreme dispersion of the rules. Taking all interesting measures into account, pre-processing the rules by selecting high ranking rules (less than 50% of the total amount) before running the cover algorithm, shows that a small number of rules can produce a similar result as whole association rule set [38].

However, discarding rules using above methods may risk losing rules that could be of interest to users. Since the threshold of support and confidence was pre-set by data owners, increasing the threshold of support and confidence could easily and simply

reduce the number of rules. But considering the worst case scenario where decentralization of all items, the support could be set to a lower value, which means that even the discarded rules could be valuable for users. A new method merging is proposed in this study without excluding any discovered rules.

3 Method

In this section, a brief introduction to the data set is introduced, followed by the methods employed by 3DMVS.

3.1 Data Set

The association rule data set used in this study has 19 attributes and 134 rules. The objective of these resulting association rules refer to a student questionnaire on the internet usage, cyberbullying and cyber-victimization.

3.2 Overview

In Chapter 2 several popular techniques were introduced. The table-based view is the simplest one but only useful for a small number of association rules and specific rules. Graph-based visualization techniques are more intuitive but only can deal with small data sets, because of the cluttering. The parallel coordinates view is also intuitive but the overlapping problem cannot be avoided when there is a large number of rules. Another alternative method to visually represent association rules is the matrix-based view, which can provide an effective overview of the rules but it is hard to understand the relations between items and there is also an overlapping problem.

In other words, the most common problems are displaying a large number of association rules on a single screen without cluttering, clearly showing the relation between the antecedents and the consequents and making the visualization easier for users to understand.

Hence, 3DMSV employs an extended matrix-based approach with rule-to-items mapping depicting association rules in 3D graphical representation with various interactions. There are four areas of display: the Visual Representation Area, the Interaction Area, the Rules Information Area, and the Information of Selected Rule Area. The Visual Representation Area not only provides an

overview of all association rules, but also varies following the different users interactions. The Interaction area provides different functions like sorting, filtering, zoom, rotation and real-time highlighting of selected cell for user to explore the interesting rules as shown in Fig 3.1. The Rules Information Area shows the basic information of the rule set, such as the number of items involved and the number of association rules etc. The Information of Selected Rule Area offers complete details of the visual rules of the selected cell, which provides users with clear information about the specific rule. When applying novel methods' merging and weighting, users can find more important items within the generated rule, and discover the more frequent consequence items of the new rules.

Fig 3.1 A screenshot of the 3D visualization system with three displaying areas and interaction area

3.2.1 Extending matrix-based view in 3D

Taking multiple antecedents into consideration, the 2D item-to-item matrix mapping was replaced by extending the rule-to-items mapping from visualizing topics in text mining [5] in order to visualize the general association rules in the transaction data set.

Item-to-item 2D matrix visualization often result in clutter problems if the rules have multiple antecedents. To tackle this problem, this study extends rule-to-items mapping in general transaction data set, where one column depicts one rule with different cells displaying the antecedents and consequents. Also, the two important measures support and confidence are depicted along with the rules. All main features of a rule are distinguished by different colors. An extended 3D matrix-based visualization of this type offers a more comprehensible and understandable overview of all the association rules.

3.2.2 Merging rules and assigning weight

With a growing number of data sets, the association rules mining always results in a large number of rules, which is a main challenge for current visualization tools. As mentioned in Chapter 2, some reduction methods like selecting part of discovered association rules and then visualizing them using clustering may increase the risk of losing interesting rules for users. Thus, a new merging method based on the former rules is proposed in this study.

The merging processes is shown as follows:

- (1) Scan and group all the association rules by consequent
- (2) Merge the rules in each group
- (3) Assign the weight of each item in the new rule

Since this study only focuses on visualizing directly from resultant association rules rather than the generation process of association rules, the novel merging method is based on the rules' known information, the relation between antecedents and consequence, support and confidence. Support is defined as the coverage of the item sets in the transaction data set, which could also be regarded as the frequency of item sets that appear in the transaction data set. Confidence is defined as the probability of the consequents appearing in the records at the same time as antecedents. In other words, it can be considered as the importance of the rule.

However, the support and the confidence are generated by the generating process of the association rules. If focus is on the resultant association rules only, these two measures are not of relevance to each other.

By taking the frequency of the different consequents and importance of the rules into consideration, a new measure assigned by the novel merging method, which can be called the frequent-confidence of the generated rules. Weight or frequent-confidence is the frequency with which one item appears in antecedents combined with the confidence of the rules, which can show how important this item is to the rules. Since the group generation only takes consequence into consideration and confidence is an important measure, the calculation of weight relies on the appearance of each item compared to its consequence and the confidence of the rules.

To clarify the process of the weight calculation, the representation of association rules are changed so that all items in the LHS are connected to each other with ' + '. Thus an association rule can be denoted as $\sum_{i=1}^n X_i \rightarrow Y$, where n is the total number of the items within the LHS. The processes of weight calculation is shown as follows:

- (1) Scan the association rules set, group the rules with the same RHS into the same group
- (2) For each group, repeat (3),(4)
- (3) Within the group, the weight of each item in the LHS is the sum of confidence of the LHS containing this item. This is denoted w_i
- (4) Within the group, the weight of the consequent is the sum of all the confidence values belonging to this group. This is denoted w_j
- (5) The new rule generated using weight is represented as

$$\sum_{i=1}^n w_i X_i \rightarrow w_j Y$$

Where w_i represents the importance of each item within the rules, w_i shows the coverage of different consequents of the generated rules.

Following the new rule merging method, weighted association rules of this type have reduced the dimension of the rules. Not only has the number of rules decreased, but also the main factors have replaced the support and confidence with new measure weight. The weight calculation depends on both confidence and appearance of different items, where appearance could be considered the frequency of an item to the rules with the same consequence. In other words, if merging rules with the method we propose, all antecedents have their own probability implying consequent, which means the relation between antecedents is 'or' not 'and'. With 'or' relations between antecedents having the same consequence, the new support and confidence would approximate to 100%. Displaying these two incomparable measures becomes redundant.

3.3 Interactivity

Exploring interesting association rules requires user interaction, which is of significance to the visualization system. The fundamental interactivity is to provide sorting and filtering to the original support and confidence as well as filter the new measure weight. Interesting association rules explorers can easily find rules with high support confidence or weight.

The basic information of the association rules displayed in the visual graph provides a more concrete overview of the rule sets. In addition, once the user clicks a cell in the Visual Representation Area of the 3DMVS, the selected cell is highlighted and the details are shown in the visual graph. Highlighting mechanisms of this type allow users to specify one rule out a large number of discovered rule sets. Displaying the details of the rules helps users to further identify and understand specific information regarding these rules. In addition, interactivity such as zooming in, zooming out, and rotation of three axes is also included.

4. Implementation of the visual system

This Chapter provides a brief introduction to 3DMVS, followed by the process of visualization [39], including mapping, selection, presentation, interactivity and human factors.

4.1 Introduction of the system

The main interface of 3DMVS was shown in Figure 4.1, where the Visual Representation Area is located to the left while the main functions can be found to the right. As mentioned in previous chapters, all of the functions are clearly displaying in our system. The remaining two areas provide further information from different aspects of the association rule sets and the selected rule. Figure 4.2 shows the weighted method part of the visualization system.

Fig 4.1 The interface of the 3D matrix-based visualization system, normal method

Fig 4.2 The interface of the 3D matrix-based visualization system, weighted method

4.2 Mapping

The first step of visualization process mapping was introduced in "Data and Information Visualization Methods, and Interactive Mechanisms: A Survey (2011) ", i.e. how to visualize information and transform the data representation into visual form. In 3DMVS, the input is a set of association rules stored in an Excel file, which should be represented in a fixed format. The data source of the Excel file should have two sheets, one listing all items included in the resulting association rules. The other sheet should provide the specific rules in the table-based view, where all items of the rules are separated into different columns with titles such as LHS1, LHS2,RHS, Support and Confidence.

To improve the efficiency of this visual system, 3DMVS employs a Data-table instead of an SQL server database as storage structure. Using an SQL server as storage structure was tested and turned out to be time-consuming in the first version of this system. The scanning time would increase as the number of the association rules increased. The Data-table method is offered by Visual Studio 2013, which uses the memory of the computer to

support formats like common table. By scanning the Excel file, the new mapped rules were generated in a format where each column represents one item and each row one rule. The values of support and confidence are considered one item, which displays at the end of the 3DMVS.

A simple example is shown in Figure 4.3. The different bars in the graph represent different factors of association rules. Each item label is displayed along the axis. Users can easily find the relations between items by mapping all the items with labels in one specific rule. This 3DMVS provides users with an overview of all the rules and offers an additional aspect which allows going into details of the specific rules.

Fig 4.3 An example of mapping rules

4.3 Selection and presentation

The second step of the visualization process is selection, which means selecting data from the whole resulting set directly depends on the aim to obtain visual graphics or pictorial representation [24]. Since there few resultant rules are available, selection does not have to be applied.

The visualization perspective presentation is the third step, i.e. how to efficiently manage and organize information in the limited space on the screen.

As shown in Figure 4.3, 3DMVS presents the rules with multiple antecedents and a unique consequent in a 3D matrix-based visual representation, where the blue bar depicts the antecedent and the red bar depicts the consequents. Support and confidence bars are displayed at the end of the matrix in different colors, and the value of these two measures are explained by the height of the bar.

The right part of the interface is the user exploration zone, where users can upload Excel files, manage the rules in the normal way, or the weighted way proposed by this study, and obtain details about the set of rules or about specific rules.

4.4 Interactivity

The next process is interactivity, which refers to provide functions used to organize, explore, and rearrange the visualization [39]. User-friendly interactivity is always considered of importance for a visualization system, since such interactivity enables users to better explore, understand and interpret association rules, which improves their exploration capabilities.

The 3DMVS provides various interactivity to simplify exploration, such as sorting and filtering, display of general information of the rule sets, highlight selected cells in the matrix and depict the detailed information of the rule at the same time, and offer zooming and rotation of the 3D display.

When clicking the submit button, visual representation is displayed immediately in the left visualization zone, and general information such as the number of items, number of rules, minimum support and minimum confidence is displayed in the text box at the same time. This is shown below in Figure 4.4

Fig 4.4 General information of the association rule sets

4.4.1 Sorting and filtering

For user exploration of interesting association rules, rules with high support and confidence seems to be more attractive. Sorting and filtering makes it easier for users to find interesting rules with high support and confidence. With the highest support and confidence displayed on the right side close to the items labels, 3DMVS brings a more explicit overview for users.

An example of sorting interesting measures support is shown in Figure 4.5. An example of sorting interesting measures confidence is shown in Figure 4.6.

Fig 4.5 Interestingness measure support sorting (dark green)

Fig 4.6 Interestingness measure confidence sorting (light green)

Filtering enables users to dynamically adjust the number of association rules to display, which means decreasing information quantity that need to displayed. Applying filtering to the interestingness measures of support or confidence helps users to select association rules with higher support or confidence. Since this purpose need some dynamic query values to manipulate are required, visual widgets can play an important role. The 3DMVS takes advantage of the track bar to ajust the threshold of support and confidence.

An example of filtering interesting measures support >0.5 is shown in Figure 4.7. An example of filtering interesting measures confidence >0.9 is shown in Figure 4.8. Figure 4.5 and Figure 4.6 display the whole rule set; filtered rules are easier to understand for users.

Fig 4.7 Interestingness measure support filtering (support >0.5)

Fig.4.8 Interestingness measure confidence filtering (confidence >0.9)

Applying the weighted method, filtering weight could help users to find items that are more important to the rules with the same consequence. The track bar enables quick adjustment of the query parameters and instantaneous view of filtered results in the visualization in real time. An example is shown in Figure 4.9.

Fig.4.9 the filtering of interestingness measure weight

4.4.2 Detail overview

3DMVS provides multiple views that combine a display of the overview and a detailed view. The overview gives a distribution for the whole association rule sets, and details are displayed in text. Highlighting mechanisms help users to specify one rule out of many the discovered rule sets without zooming. This type of interactivity makes the overview more accessible by exploring the rules they are interested in.

When users click the representation generated by the 3DMVS, all information about the association rules will be displayed in the text together with the visual representation, including antecedents, consequents, support and confidence. Since an association rule is a single entity, if only part of the association rules is displayed, it is not meaningful for explorers. If the cell clicked is empty, there is no more detailed information to display.

4.4.3 Zoom and rotation

The details in text-box enables users to obtain a part view of specific rules, while the visual interface zoom could facilitate for users to obtain the part visual representation as well. 3DMVS supports alternative interface where different details can be zoomed in and zoomed out take the user back to the overview. An example is shown in Figure 4.10.

Fig 4.10 A screen-shot of zooming

The rotation function supports three different axes, which could be a way to avoid the occlusion problem. The different aspects of the overview offers clearer and more accessible view for the explorer. An example is shown in Figure 4.11.

Fig 4.11 An example of rotation

5 Evaluation and discussion

3DMVS can help users explore the rules they are interested in, which did to produce tangible results. What makes 3DMVS a good visualization for association rules?

The quality of a visualization system lies in its ability to quickly and reliably transmit data from source to destination [40]. It is also mentioned in [Data and Information Visualization Methods, and Interactive Mechanisms: A Survey] [39] that visualization is the transformation of symbolic representation to geometric representation. While in [A Saliency-based Quality Metric for Visualization] [41], the quality of visualization images depends on factors such as domain-specific requirements, the user's needs and expectations, source data set and techniques used. However, in the case of a specific visualization system, we need to combine those evaluation techniques with an individual visualization system.

Using the results in [On Evaluating Information Visualization Techniques] [42], two sets of criteria are used for the evaluation, first the usability of visual representations is tested, followed by testing the interaction mechanisms.

5.1 Usability testing

The usability of a visualization system is introduced from four aspects, including completeness, spatial organization, information coding and state transition.

5.1.1 Completeness

Completeness is associated with the method of representing all the semantic contents of the data to be displayed. The data set used in this study is a set of association rules, consisting of five factors including antecedents, consequents, the relation between them, support and confidence. A good visual representation needs to display all factors clearly in the visualization.

According to the results, the five factors are represented in their entirety, each antecedent and the consequent of a rule, as well as interesting measures support and confidence. The four main factors are displayed in one column to connect the relationship between them. The basic information of the whole set of association rules includes 19 items and 134 rules, which was given in a text format. The relations between items are imposed by the visual representation using the rule-to-items method. However, with a growing number of association rule sets, the visual image could increase cognitive complexity, leaving users with a compact visual representation.

5.1.2 Spatial organization

Spatial organization was introduced in [On Evaluating Information Visualization Techniques] [42], which is related to the overall layout of a visual representation, which comprises an analysis of how easy it is to locate information elements on the display and be aware of the overall distribution of information elements in the representation [. 3DMVS displays association rules in 3D visual representation, where one column shows one rule using different colors to display the features of the rule, implying the relation between items within the rules. The Z-axis values are displayed by the bar chart which enables users a clear overall view of all items of the data set. A smaller value is denoted items within the rules in order to avoid occlusion problems, since support and confidence is displayed at the end of the visual representation.

5.1.3 Codification of information

Additional codification of information is another aspect 3DMVS uses to evaluate visualization techniques, which means not only to map data elements to visual elements, but also offer additional symbols or realistic characteristics to build additional representation. Since 3DMVS takes advantage of the ability of obtaining information from human perspective vision system. The opposite colors are easier to distinguish for human beings. The additional representation will be taken into consideration in future studies.

5.1.4 State transition

Another important aspect of information visualization techniques is the result of rebuilding the visual representation after a user action [42]. 3DMVS responds the user's movements by different state transitions that are supported by different data tables. Using different data tables to store the results of the queries, 3DMVS can maintain all state transitions according to interactivity, such as sorting or filtering with different interestingness measures. The time consumed by the technique to maintain all state transitions and changes in the spatial organization of the resulting image are also important factors that can affect the perception of information. Changing the resulting image in spatial organization requires two-layers of loops, which is unavoidable based on this matrix-based approach. A disadvantage is that if the number of the rules increases, the time will increase as well.

5.2 Interaction testing

To evaluate the interaction of a visualization system, various functions are needed to support user tasks in visualization applications. There are three categories for testing interaction mechanisms (orientation and help, navigation and querying, and data set reduction) [39].

5.2.1 Orientation and help

Help and user orientation features are defined by a representation of additional information, such as support for the user to control level of details, redo/undo user actions.

Orientation and help could help users become familiar with the visualization system and guide them to the right operation in this system. All interactions have a short instructions for each function supported by the 3DMVS, which could provide users to a better understanding of 3DMVS. When the mouse moves over the interface in 3DMVS, a description of the area will describe the functions of each area. If the users do not follow the exploration instructions a message box will pop up reminding them follow the instructions.

5.2.2 Navigation and querying

Following the evaluation process in [39], navigation and browsing techniques should be analyzed in terms of possibilities and how easy it is to a data element, change user point of view, manipulate geometric representations of data elements, and search and query for specific information. This part of the evaluation process should be tested by users. However, there is not enough time to evaluate our system using an online survey. Meanwhile, not all the users can provide objective evaluations.

Several features are proposed, such as clicking the mouse to obtain more details of the specific rule from the overview, zooming provides detail changes for the user, and rotation offers users different views of the 3D visual representation, which needs to be further evaluated by users.

5.2.3 Data set reduction

The last subset of criteria of the evaluation process mentioned in [Data and Information Visualization Methods, and Interactive Mechanisms: A Survey] [39] is related to the data set reduction features provided by the technique. 3DMVS provides filtering and a merging method allowing a reduction of the information shown at a certain point in time, which leads to a more rapid adjustment to the focus of interest.

Filtering of support and confidence enables data explorers to focus on the subset of the rules that have higher support and confidence. The new merging method in generating the new rules with weight reduces the dimension of the association rules, which also provides a novel way to view more important items of a specific rule.

5.2 Discussion of the visual ability

This section presents a more concrete discussion about the visual representation with the same set of association rules. This is followed by a comparison of the 3D matrix-based view and weighted merging rules visualization, to show both advantages and disadvantages.

5.2.1 Comparing the 3D matrix visualization with other methods

Figure 5.1 shows a scatter plot visual representation of the 134 rules. As mentioned in Chapter 2, users can only obtain an overview of the whole set of rules with the main measures support, confidence and lift. Discovering relations between items could be difficult for users; grayscale could contribute to interesting relations being difficult to find.

Fig 5.1 Scatter plot of the data set including 134 rules

Figure 5.2 illustrates the 134 rules in parallel coordinates. All factors of an association rule is displayed in a single image, where users can easily find the relations between items as well as the interestingness measures support and confidence. Thus if the number of the rules is low, parallel coordinates is a good option. However, users may struggle to explore interesting one since color and the width of the lines are unable to convey the differences between the rules.

Fig 5.2 Parallel coordinates plot for 134 rules

3DMVS is an interactive system as opposed to visual images. 3DMVS not only provides user interactivity for further exploration, but also displays all factors of the association rules needed on a single screen. In such a 3D matrix-based view, the rules become more understandable and comprehensible.

5.2.2 Comparison of original visual graph with weighted one

Comparing Figure 4.1 with Figure 4.2, there are some differences.

The visual representation of all association rules mainly focuses on displaying all factors of a rule in a more understandable and comprehensible way to support user exploration. Such visualization depicts the relations between the items of a rule, as well as support and confidence, which helps user to explore interesting rules.

A method of weight is proposed to handle a large number of association rules based on assumptions, which could reduce the number rules. Moreover, the weight of the item cloud also be explained as the importance or contribution of one rule with specific consequence, which also employs the filtering to select more vital ones. The weight of consequence provide a comparison of the generated rules, which explains the attendances of a specific consequent in the data set of association rules.

6 Conclusion and future work

In this study, a visualization of association rules called 3DMVS was implemented. 3DMVS employs rule-to-items mapping instead of item-to-item mapping, and is able to multiple antecedents. The new measure weight was calculated during rules merging, which represents probability of an item implying the same consequence. Necessary interactivity such as sorting and filtering was also achieved in this system. Moreover, once a user clicks the a cell in the matrix, the details will be shown immediately in the visual representation.

3DMVS is able to visualize the association rules with multiple antecedents, not only depicts the relations between items but also providing information about each item. It is designed according to Shneiderman's visual information seeking metric "overview first, zoom and filter, then details on demand". Displaying the matrix-based overview in 3D makes the visual representation more vivid and clear for the exploration. The new method merging the original rules with assigning a new measure weight reduces the discovered rules reduces the problem with the visualization system. Visualizing the merged association rules with weight provides an overview of each item within the rules, making it easier for users to find more interesting items in the rules. The interactivity of sorting and filtering interestingness measures including support, confidence and weight, facilitates exploring and finding interesting rules. Another real-time interactivity consists of users being able to learn more about a specific rule by clicking the visual representation, which helps users to further identify and understand specific information about these rules. The results indicate that the design can easily handle hundreds of multiple antecedents and a large number of association rules in a 3D display including human interactions, without occlusion and no screen swapping.

An obvious trade-off in this system is that it only supports Excel files with the xls filename extension. The system is inefficient for a large number of rules; the system requires two loops to prepare

rules for visualization. The algorithm for mapping the real rules with the visual representation needs improvement to decrease time. The support for different kinds of files needs completion. Improving the efficiency of the mapping algorithm is also of significance.

6.1 Ethical considerations

The input xls file of the resultant association rules has been kept private during this study, not any actual content from the data source were published.

The xls file was only used to test 3DMVS. But the result of the system cloud contain sensitive or valuable information, the use of which should be limited.

References

- [1] Ali, A B M Shawkat, "Dynamic and Advanced Data Mining for Progressing Technological Development", 2009. P344. Line 28-30.
- [2] Y. Sekhavat and O. Hoeber, "Visualizing Association Rules Using Linked Matrix, Graph, and Detail Views", International Journal of Intelligence Science, Vol. 3 No. 1A, 2013, pp. 34-49.
- [3] K. Techapichetvanich and A. Datta, "VisAR: A New Technique for Visualizing Mined Association Rules", Proceedings of the International Conference on Advanced Data Mining and Applications, Wuhan, 22-24 July 2005, pp. 88-95.
- [4] Chan K C C, Au W H. Mining fuzzy association rules[C]// International Conference on Information and Knowledge Management. ACM, 1997:209-215.
- [5] H.H.Ong, K.L.Ong, W.K. Ng and E.P.Lim, "Crystal Clear: Active Visualization of Association Rules," Proceedings of the International Workshop on Active Mining, Maebashi City, 9 December 2002, pp. 123-132. [15]
- [6] Imielienskin T, Swami A, Agrawal R, "Mining association rules between set of items in large databases" [J]. Acm Sigmod Record, 1993, 22(2).
- [7] Agrawal, Rakesh; and Srikant, Ramakrishnan; "Fast algorithms for mining association rules in large databases", in Bocca, Jorge B.; Jarke, Matthias; and Zaniolo, Carlo; editors, Proceedings of the 20th International Conference on Very Large Data Bases (VLDB), Santiago, Chile, September 1994, pages 487-499
- [8] Han. "Mining Frequent Patterns Without Candidate Generation". Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. SIGMOD '00: 1–12. doi:10.1145/342009.335372
- [9] Hipp, J.; Güntzer, U.; Nakhaeizadeh, G. (2000). "Algorithms for association rule mining-a general survey and comparison". ACM SIGKDD Explorations Newsletter 2: 58. doi:10.1145/360402.360421

- [10] Zhao Q, Bhowmick S S, "Association Rule Mining: A Survey" [J], Nanyang Technological University, 2003, 3(3):157–169.
- [11] Srikant R, Agrawal R, "Mining Quantitative Association Rules in Large Relation Tables" [J], *Acm Sigmod Record*, 1999, 25(2):1-12.
- [12] Tung A, Lu H, Han J, et al, "Efficient mining of inter-transaction association rules[J], *IEEE Transactions on Knowledge & Data Engineering*, 2003, 15(1):43-56.
- [13] Fukuda T, Morimoto Y, Morishita S, et al, " Data mining using two-dimensional optimized association rules: scheme, algorithms, and visualization" [J], *Acm Sigmod Record*, 1996, 25(2):13-23.
- [14] Codd, E.F.(1970), "A Relational Model of Data for Large Shared Data Banks", *Communications of the ACM* 13 (6): 377–387, doi:10.1145/362384.362685
- [15] C. Romero, J. M. Luna, J. R. Romero and S. Ventura, "RM-Tool: A Framework for Discovering and Evaluating Association Rules", *Advances in Engineering Software*, Vol. 42, No. 8, 2011, pp. 566-576. doi:10.1016/j.advengsoft.2011.04.005
- [16] M. Hahsler, S. Chelluboina, K. Hornik and C. Buchta, "The Arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets", *Journal of Machine Learning Research*, Vol. 12, No. 1, 2011, pp. 2021-2025.
- [17] Bruzzese D, Buono P, "Combining visual techniques for Association Rules exploration" [C], *Working Conference on Advanced Visual Interfaces, AVI 2004*, Gallipoli, Italy, May. 2004:381-384
- [18] Inselberg, A., Reif, M., Chomut, T, "Convexity algorithms in parallel coordinates" *Journal of the ACM* 34, 765–801 (1987)
- [19] Inselberg, A, "Parallel coordinates: A tool for visualizing multi-dimensional geometry", In: *Proc. 1st IEEE Conf. on Visualization*, San Francisco, CA, pp. 361–375(1990)
- [20] Inselberg, A, "Visualizing high dimensional datasets and multivariate relations (tutorial)", In: *Proc. 6th ACM SIGKDD Inter. Conf. on Knowledge Discovery and Data Mining (KDD 2000)*, Boston, MA (2000)

- [21] Martin, A., Ward, M.O, "High dimensional brushing for interactive exploration of multivariate data". In: Proc. IEEE Conf. on Visualization, Atlanta, GA, 1995,pp. 271-278
- [22] L. Yang, "Visual Exploration of Frequent Itemsets and Association Rules", In: S. J. Simoff, M. H. Böhlen and A. Mazeika, Eds., Visual Data Mining: Theory, Techniques and Tools for Visual Analytics, Springer Berlin Heidelberg, Berlin, 2008, pp. 60-75.
- [23] L. Yang, "Pruning and Visualizing Generalized Association Rules in Parallel Coordinates", IEEE Transactions on Knowledge and Data Engineering, Vol.17, No.1,2005,pp.60-70.
- [24] Michael Hahsler, Sudheer Chelluboina,"Visualizing Association Rules: Introduction to the R-extension Package arulesViz" <https://cran.r-project.org/web/packages/arulesViz/vignettes/arulesViz.pdf> ,2011.
- [25] Utts, Jessica M. (2005). "Seeing Through Statistics", 3rd Edition, Thomson Brooks/Cole, 2005, pp 166-167
- [26] Sitanggang I S, "Spatial Multidimensional Association Rules Mining in Forest Fire Data" [J], Journal of Data Analysis & Information Processing, 2013, 01(4):90-96.
- [27] W.S.Cleveland and R.McGill, "Graphical Perception: The Visual Decoding of Quantitative Information on Statistical Graphs", Journal of the Royal Statistical Society, Vol. 150, No. 1, 1987, pp. 192-229. doi:10.2307/2981473
- [28] C. Ware, "Information Visualization: Perception for Design", 2nd Edition, Morgan Kaufmann, Burlington, 2004.
- [29] M. Hahsler, S. Chelluboina, K. Hornik and C. Buchta, "The Arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets", Journal of Machine Learning Research, Vol. 12, No. 1, 2011, pp.2021-2025
- [30] W. S. Cleveland and R. McGill, "Graphical Perception: The Visual Decoding of Quantitative Information on Statistical Graphs", Journal of the Royal Statistical Society, Vol. 150, No. 1, 1987, pp. 192-229. doi:10.2307/2981473

- [31] Hahsler M, Chelluboina S, "Visualizing Association Rules in Hierarchical Groups" [J], Journal of Business Economics, 2011:1-19.
- [32] Pak Chung Wong, Paul Whitney, Jim Thomas, "Visualizing Association Rules for Text Mining", Information Visualization, 1999. (Info Vis '99) Proceedings. 1999 IEEE Symposium on, p120 - p123, p152.
- [33] Mafruz Zaman Ashrafi, David Taniar, and Kate Smith, "Redundant Association Rules Reduction Techniques", AI 2005, LNAI 3809, pp.254–263,2005. Springer-Verlag Berlin Heidelberg 2005
- [34] Waler A.Kosters, Elena Marchiori and Ard A.J. Oerlemans, "Mining Clusters with Association Rules", Advances in Intelligent Data Analysis (IDA-99) (D.J.Hand, J.N.Kok and M.R. Berthold, Eds.), Lecture Notes in Computer Science 1642, Springer, 1999, pp. 39-50.
- [35] H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hatonen, and H. Mannila, "Pruning and grouping discovered association rules", In Proc. ECML-95 Workshop on Statistics, Machine Learning, and Knowledge Discovery in Database, April 1995,pp 47-52.
- [36] Pi Dechang and Qin Xiaolin, "A new Fuzzy Clustering algorithm on Association rules for knowledge management", Information Technology Journal 7(1), 2008, pp. 119-124.
- [37] Alipio Jorge, "Hierarchical Clustering for thematic browsing and summarization of large sets of Association Rules", In Proceedings of the 4th SIAM International Conference on Data Mining. Orlando, FL, 2004, pp. 178-187.
- [38] S.Kannan¹ and R.Bhaskaran², "Association Rule Pruning based on Interestingness Measures with Clustering", IJCSI International Journal of Computer Science Issues, Vol. 6, No. 1, 2009
- [39] Muzammil Khan, Sarwar Shah Khan. "Data and Information Visualization Methods, and Interactive Mechanisms: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 34– No.1, November 2011
- [40] Daniel Pineo, Colin Ware. "Data Visualization Optimization via Computational Modeling of Perception", IEEE transactions on visualization and computer graphics, VOL. 18, NO. 2, 2012

- [41] H. Jänicke¹ and M. Chen. "A Saliency-based Quality Metric for Visualization", Eurographics /IEEE-VGTC Symposium on Visualization 2010. Volume 29 (2010), Number 3
- [42] Freitas C M D S, Luzzardi P R G, Cava R A, et al, "On Evaluating Information Visualization Techniques", Published and presented in AVI' 02 – Advanced Visual Interfaces. Trento, Italy, May 2002.
- [43] B. Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations", Proceedings of the IEEE Symposium on Visual Languages, Boulder, 3-6 September 1996, pp. 336-343.[doi:10.1109/VL.1996.545307](https://doi.org/10.1109/VL.1996.545307).
- [44] Fisher A. Ronald, "The use of multiple measurements in taxonomic problems", Annual Eugenics, vol. 7, no. Part III, pp. 179-188, 1936.