ISSN 2278-3091

Volume 9, No.4, July – August 2020

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse272942020.pdf

https://doi.org/10.30534/ijatcse/2020/272942020

Enhanced Robust Association Rules (ERAR)Method for Missing Values Imputation

Awsan Thabet[1], NurulA. Emran[2], AzahK. Muda[3]

[1,2,3]Centre for Advanced Computing Technologies (C-ACT), Fakulti Teknologi Maklumat Dan Komunikasi,
Universiti Teknikal Malaysia Melaka (UTeM), Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia,
awsanthabet666@gmail.com[1], nurulakmar@utem.edu.my[2], azah@utem.edu.my[3]

ABSTRACT

Missing values or incomplete data is a common problem that occurs in many applications. In most cases, recovering missing values from data sets is necessary to avoid bias conclusions made by omitting missing values. Missing values recovery (that is also known as missing values imputation) is an important research subject in the field of statistics and data mining. In this paper, we present the Enhanced Robust Association Rules (ERAR)method to extract useful association rules and avoid redundant rules. We show the enhancement made on ERAR to improve the imputation performed by the original Robust Association Rules (RAR). ERAR is designed in selecting the frequent items in datasets that are only related to missing values. Therefore, unnecessary frequent items can be ignored in generating the association rules. The result of the experiment shows that ERAR offers better performance in terms of the time taken for the imputation process and the amount of memory used to complete the imputation. In particular, ERAR behaves better in a monotone pattern of missing values than the arbitrary pattern. In terms of imputation accuracy, we found that both ERAR and RAR exhibit a decreasing rate of accuracy as the amount of missing values increases for data of arbitrary pattern, but this is not the case of data of the monotone pattern. With the findings, ERAR contributes to improving how one can deal with incomplete data.

Key words: missing value simputation, data completeness, Robust Association Rules

1. INTRODUCTION

Missing values is a common problem highlighted in data mining, statistics, and database domains [23-25]. Mining techniques are typically unable to deal with missing values and need some kind of pre-processing or workaround [1]. Imputation is a technique used to handle missing values where missing values are replaced by substituted values [26]. Missing values imputation can be seen reported in the application of Mean\Mode, K-Nearest Neighbor, Hot-Deck, Expectation-Maximization[2], and machine learning [26-29]. In addition to these techniques, the association rules technique is commonly used to deal with missing values.

To impute missing values one needs to consider several factors which are the rate of missing values, the missing values mechanism, and the pattern of missing values. It is useful to know the mechanism and the pattern of the missing values in incomplete datasets before one can decide on the ways to treat them [3][4]. The mechanisms of missing values describe the possible relation between the observed and the missing values [5][4], while the pattern of missing values refers to the observed values and the configuration of the missing value in the dataset [3][4].

Agrawal and Srikant (1994)proposed a method to speed up mining association rules in datasets containing a large number of transactions[6]. Based on their method, Ragel and Crmilleux (1998) developed the Robust Association Rules (RAR) method to discover missing values[7]. Since then, more methods for handling missing values were proposed based on the RAR approach, such as Recycle Composed Association Rule (RCAR) [8], Fast Recycle Combined Association Rules (FRCAR) [1], iterative missing-value completion [9] and Association Rule Mining from Data with Missing Values (ARDM) [10].

In data mining, association rules are used to discover all possible relations between the values (called rules) that will be used to predict future behaviors. The same principle is applied in missing values imputation where a set of rules are generated to recover missing values. Most of the existing methods that used association rules for missing values imputation select all possible frequent itemsets in a dataset to generate the rules. The rule will be used later to fill up the missing values. These methods differ in terms of the way frequent itemsets and rules are generated.

In this paper, we propose the Enhanced Robust Association Rule (ERAR) that extends the way frequent itemsets are selected in RAR. In ERAR, the frequent itemsets are filtered to determine frequent itemsets that relate to the missing values. Through this process, as the size of the frequent itemsets is reduced, the number of rules generated will also be reduced. This will offer improvement in the imputation process in terms of imputation speed and the amount of memory saving.

In the next section, related work on the missing values mechanisms, patterns, and association rules will be presented. Section 3 consists of the details of the proposed model, Section 4 covers the experimental results, and finally, Section 5 concludes the findings of this research.

2. RELATED WORK

2.1 Missing Values Mechanisms

According to Enders (2010), the mechanisms of missing values reveal the potential relationships between the observed variables and the missing value probabilities. Missing values mechanism is usually classified into three types: *missing completely at random* (MCAR), *missing at random* (MAR) and *missing not at random*(NMAR)[13][4][14][15]. Kaiser (2014)pointed out that determining missing data mechanisms is a challenging process[13]. The following are the characteristics of the above mentioned missing data mechanisms.

1) *Missing Completely at Random (MCAR)*

The MCAR comes in a high degree of randomness. There is no explanation for missing the data [15]. The missing values in MCAR do not rely on the data observed or the data not observed in an attribute. according to Allison, if any missing variable X does not depend on any other variable Y and itself we call it MCAR [11][2]. The missing variable X in the dataset can not be determined from any other variables. With all the missing variables, the prediction is the same.

2) *Missing at Random (MAR)*

A less stringent assumption is needed about why data are missing in The MAR mechanism. MAR happens when the missing values are related to other values but not to values itself [16]. Simplify, The value on missing supposed variable X will be predicted depending on the other variable Y in a specific data set., this is completely different than the MCAR [17][16].

3) *Missing Not at Random (MNAR)*

According to Aljuaid (2016)[18], the MNAR happens if it can depend on the value of an attribute for the probability of a record having a missing value. Simplify, if we have X variable contain missing values when X is related to the values of X itself we call this MNAR. MNAR is called “non-ignorable” because you have to model the missing data mechanism itself when treating the missing data.

In summary, it is important to emphasize that the missing values mechanisms are not the property of a whole dataset, but rather hypotheses for different analyses. As a consequence, the same data set can produce MCAR, MAR or MNAR analyzes depending on which variables the analysis includes [13][19].

2.2 Missing Values Pattern

The missing values pattern defines only the position of the missing data and does not describe why the data is missing. Knowing the pattern of missing values in advance helps to determine suitable imputation techniques that will be used to treat the missing values. According to Peugh and Enders (2004), there are six prototypical missing data patterns, but the most popular ones are as follow :

1.Univariate Pattern

This contains a single attribute with missing values. For certain fields, a Univariate trend is very uncommon but can occur for experimental studies. Figure 1 shows a univariate pattern, where Y4 is the incomplete variable, and Y1 to Y3 are manipulated variables.

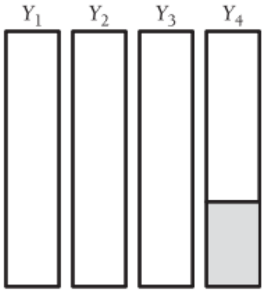

Figure 1: Univariate Pattern[18]

2.Monotone Pattern

The monotony pattern is like a hierarchy so that subsequent measurements are often missing for cases with missing values on a specific evaluation. Monotone missing data patterns if known in advance can greatly minimize the maximum likelihood mathematical complexity and multiple imputations, and the need for iterative estimation algorithms [1][4]. Figure 2 illustrates the monotone pattern.

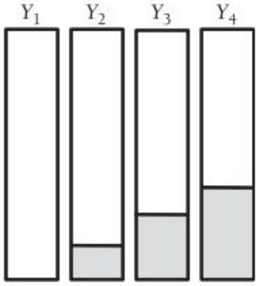

Figure 2: Monotone Pattern[18]

3) *Arbitrary Pattern*

Arbitrary pattern or generalized pattern they name it in some papers is perhaps the most common configuration of missing values. In this category, there is no specific pattern in missing values in data structure. The arbitrary pattern includes missing values distributed randomly around the data matrix

[12][20]. Figure 3 shows the arbitrary pattern, where missing values are distributed in a non-systematic fashion.

Knowing the missing values mechanisms and patterns can offer performance advantage in missing values imputation algorithms. Nevertheless, it is difficult to find a method that works best in all cases[1]. For this reason, finding missing values imputation methods that can offer optimal performance for specific cases is an open problem.

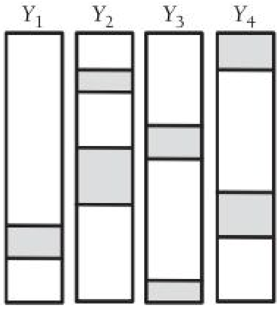

Figure 3: Arbitrary Pattern [18]

2.3 The Association Rules

In a relational database context, an association rule is an expression X→Y, where X is a collection of attributes in a table and Y is normally an attribute. It means that if all of the attributes in X are available on a transaction in a tuple, Y is also most definitely in the tuple. Agrawal and his co-workers have suggested several mining methods to find the association rules [21]. The mining method was divided into the following steps:

1- Count the frequencies for the candidate itemsets by scanning the dataset.If the number of tuples in which the set appears is equal to or greater than a threshold value (known as minimum support), the itemset will be considered as frequent.
2- After selecting the frequent items at the first level, the larger itemsets will be processed by merging two frequent items to create the candidates inthe form of two items. Based on the support values, the frequent itemsets will be selected from the created candidates' itemsets. This process is iterative, level by level until all large itemsets in the dataset are found.
3- Form and calculate the confidence value for all possible relationships (association rules) between the items in the large dataset created in the previous step.The process is repeated on all created itemsets.
4- Selectthe rules with confidence values that are equal to or larger than the pre-defined confidence threshold.

In 1998, Ragel and Crmilleux proposed the Robust Association Rules (RAR) technique to mine associations rules within incomplete datasets [7]. The method deals with the problem of generating rules with the presence of missing values. With RAR, the tuples containing missing values will be partially disabled and not being deleted.In this way, the number of rules generated is not reduced even though missing values are present. The method also modified the *Support* and *Confidence* equations to calculate the frequent itemsets and rules. The rules mined by the RAR method are then used to recover the missing values in a dataset[7]. Formally, let the tuple set $\sigma(x)$ for an itemset $x$be defined as follows:

$$\sigma(x) = \{Tup|x \subseteq Tup, Tup \in D\}, \qquad (1)$$

where *D*is the dataset and *Tup*is the tuples.

The disabled missing values with itemset *x*is notated by *Dis(x)* defined as follows:

$$Dis(x) = \{Tup|\exists\ A \in x, A = ?, x \subseteq Tup, Tup \in D\} \qquad (2)$$

The sign '?' is the missing value in the attribute. With this definition, *A* is a missing attribute belonging to *x*. The *support* equation for itemset *x*, based on RAR is defined as:

$$Sup(x) = \frac{|\sigma(x)|}{|D|-|Dis(X\cup Y)} \qquad (3)$$

The equation for the confidence is defined for an association rule X→Y based on the RAR approach as follows:

$$Conf(X \to Y) = \frac{|\sigma(X\cup Y)|}{\sigma(X)-|Dis(Y)\cap\sigma(X)|} \qquad (4)$$

In order to retrieve several missing values in a dataset, Ragel and Cremilleux suggested an approach for Missing-Values Completion function in the RAR process [22]. RAR is used to discover all rules of the association, and then the most suitable rule discovered by RAR will be applied to replace a single missing value in the tuple. In the next section, we will present the extension made in RAR to improve missing values imputation.

3. THE ENHANCED ROBUST ASSOCIATION RULES (ERAR)

Based on the existing works on the RAR that deal with association rules discovery within incomplete datasets (refer[7][22]), in this section we propose the Enhanced Robust Association Rules (ERAR) that will offer better missing values imputation performance.

ERAR maintains the same steps as in RAR where in the first phase, it will collect the frequent itemsets from the candidate values in an incomplete dataset based on the threshold of the support value. In the second phase, the rules are generated based on the selected frequent itemsets in the first phase and by use threshold of the confidence value. In the final phase, the Missing-Values Completion function to fill the missing values in incomplete datasets based on the generated rules in the second phase.

Unlike RAR, the ERAR method will enhance the first phase by collecting the frequent itemsets that are related to missing values only. This step will avoid calculating the rest of the frequent itemsets within the incomplete dataset. With the enhancement, faster computation and less amount of memory usage can be offered by ERAR. The details of the steps are as follow:

1. Define the candidate itemsets in the tuples that consist of at least one attribute with missing values.
2. Calculate the frequent items found in step 1 for candidates based on the support threshold (we refer to the selected frequent items in this step as F1).
3. Locate the candidates in every column that consists of at least one attribute with missing values. This will avoid repeating calculating frequent itemsets from the previous steps.
4. Calculate the frequent items for candidates selected in step 3 based on the support threshold (we refer to the selected frequent items in this step as F2).
5. In this step, the pure frequent items will be selected to generate the rules that will be produced. It depends on the amount of intersecting rows between F1 and F2. The conditions that will be checked are as follow:
   - If the current value in F1 intersects with any value in F2, the intersected values in F1 and F2 will be added into the pure frequent items group. We refer to the selected frequent items in this step as F3 (where F3 will contain the frequent items needed to generate the rules).
   - If there is no intersection between the current value in F1 with any F2 values, then move to the next value in F1and compare it with F2 values. The process continues until all frequent items in F1 is compared with every frequent item in F2.

   At the end of this step, F3 will contain all frequent items needed to generate rules. The process will avoid adding frequent items that are already added to F3 in step 5. Thus, this will require a small amount of memory and the time taken for rules generation against a small number of frequent items will become shorter.
6. Missing-Values Completion function will be applied to impute the missing values based on the rules generated in the previous step.

4. EXPERIMENTAL SETUP

In the experiment, we set to observe RAR and ERAR against two factors which are the rate of missing values and the pattern of missing values for data sets with MAR mechanism. The experiment was set on a 4 GB RAM, Intel Core i7 (2.9 GHz), and PC operating Windows 7. The proposed method was implemented on two real datasets, Zomato and Restaurants on Yellow pages, which were taken from the Kaggle. The description of the two datasets as well as the thresholds used in experiments are shown in Table 1. The missing amounts of 5 to 30 percent are set for both data sets where *arbitrary* patterns have been randomly assigned to all Zomato attributes. The Restaurants data set was assigned with a *monotone* pattern.

Table 1: Characteristics and threshold values used of the two experimental datasets

Dataset	Tuple No	Attr No	Missing Pattern
Zomato	9551	21	Arbitrary
Restaurants onYellow pages	6000	11	Monotone

RAR was executed with the same data set as a comparison to ERAR. To evaluate the performance ERAR, four performance indicators are used namely the number of frequent itemsets, memory usage, time taken for imputation processes, and the accuracy of imputation.

5. EXPERIMENTAL RESULTS

5.1 The Amount of Frequent Itemset

Figure 4 shows the number of frequent items selected by RAR and ERAR over a range of missing values rates(in %) on the Zomato dataset with arbitrary missing data patterns. Based on the figure, ERAR exhibitsa lower number of selected frequent items than RAR. Note that, as the rate of missing ration increases, the gap in the number of selected frequent items between RAR and ERAR has decreased.

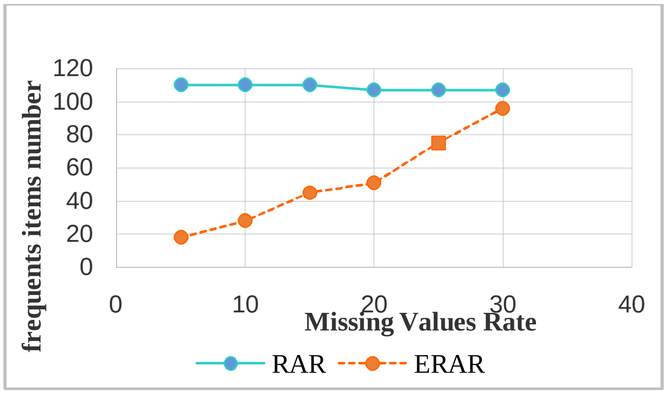

Figure 4: Frequent items number for ERAR and RAR on Zomato (Arbitrary pattern)

Figure 5 shows the number of frequent items selected by the RAR and ERAR over a range of missing data rates(in %) on the Restaurant dataset with a monotone pattern. Like Zomato's data set, ERAR performs better than RAR in terms of having a lower number of frequent itemsets. Nevertheless, unlike Zomato, even though the rate of missing value increases, ERAR's performance is steady towards the end. This shows ERAR behaves better in monotone patterns than the arbitrary pattern.

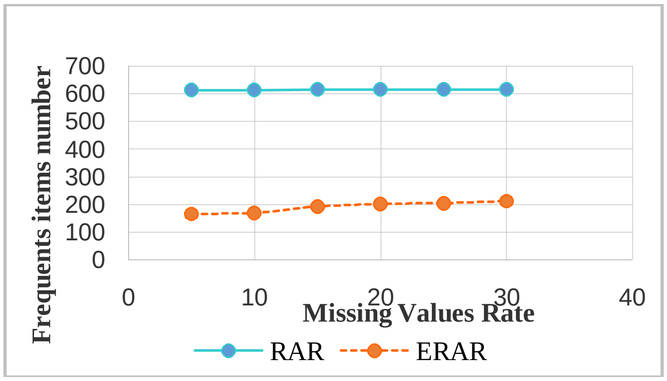

Figure 5: Frequent items number for ERAR and RAR on Restaurant data set (Monotone pattern)

5.2 Memory Usage

Figure 6 and Figure 7 show the memory usage during the imputation process on the Zomato data set and Restaurant dataset respectively on a range of missing values rate. The results show that there is no significant difference between ERAR and RAR for the Zomato data set especially for missing values from 15 to 30%. But for the Restaurant data set, we can see ERAR consistently beats RAR in terms of the amount of memory saving.

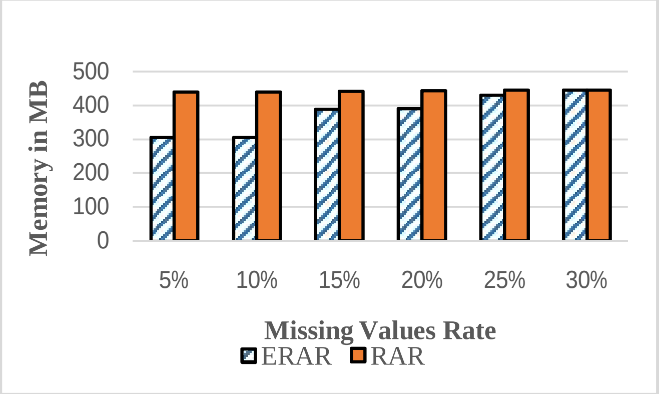

Figure 6: Memory usage for ERAR and RAR on Zomato (Arbitrary pattern)

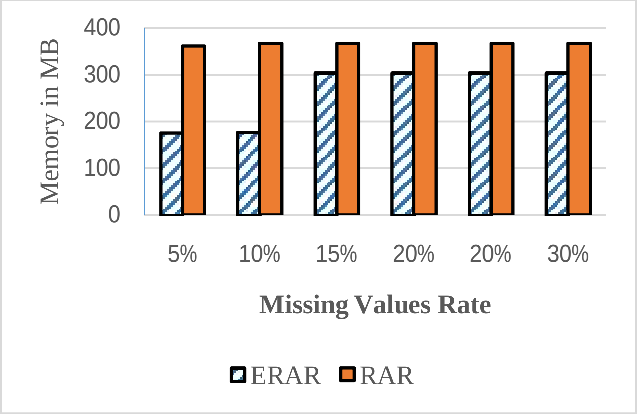

Figure 7: Memory used for ERAR and RAR on Restaurant data set (Monotone pattern)

5.3 Imputation Time

Figure 8shows the time taken for imputation by RAR and ERAR for the Zomato data set. The results show that there is a huge time difference in imputation exhibited by ERAR as compared to RAR especially from 5 to 20 percent missing values ratio. After 20 percent, there is a sharp increase in the imputation time taken by ERAR and at 30 percent, the performance of ERAR and RAR is equal. This can be explained by the decreasing gap in the amount of frequent itemset as the rate of missing value increases, as shown in Figure 4.

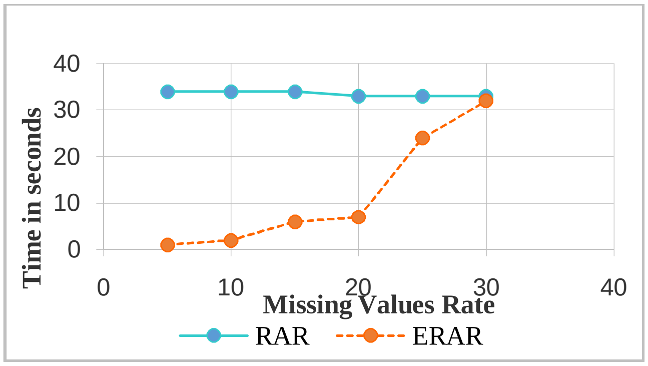

Figure 8. The duration on the imputation process for ERAR and RAR on Zomato(Arbitrary pattern)

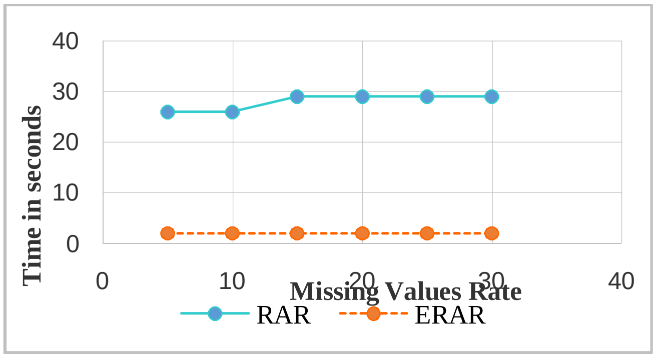

Figure 9: The duration on the imputation process for ERAR and RAR on Restaurant data set (Monotone pattern)

In comparison, for Restaurant data set ERAR's performance is better than ERAR even though the rate of missing ratio increases. The time taken for ERAR to impute missing values is consistently less than 5 seconds, as shown in Figure 9. This shows that ERAR offers faster imputation than RAR especially for data of monotone missing values pattern. This is supported by the size of frequent itemsets which is consistently small for ERAR as shown in Figure 5. More time was taken for imputation involving the Zomato data set due to the size and the arbitrary pattern factors.

5.4 Imputation Accuracy

The results yielded on the imputation accuracy show that both ERAR and RAR exhibit the same level of accuracy. As shown in Figure 10, there are differences in the level accuracy within a range of missing values rates. The imputation is 100% accurate for missing values from 5 to 25% for the Zomato data set with an arbitrary pattern. The accuracy drops at 30% of the missing value rate. As for the Restaurant data set with the monotone pattern, we can see an inconsistent level of accuracy shown by both algorithms as the rate of missing value increases. As expected, we can say that the accuracy of ERAR and RAR is decreasing as the amount of missing values increases for data of arbitrary pattern, but this is not the case of data of the monotone pattern.

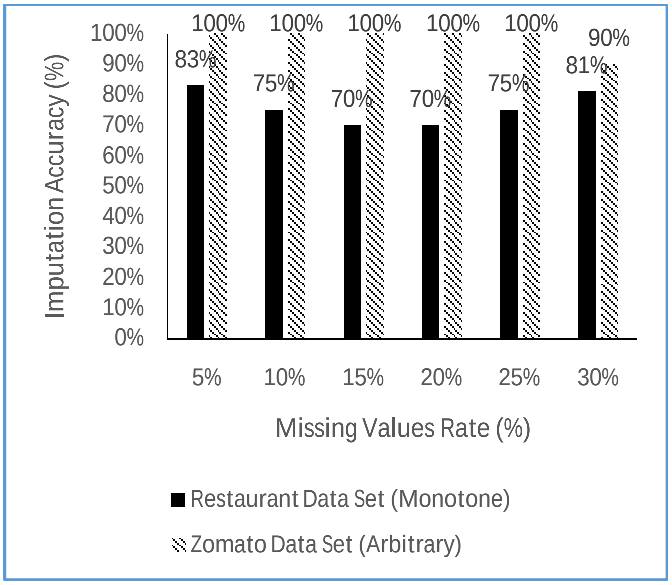

Figure 10: Imputation Accuracy for ERAR and RAR on Restaurant and ZomatoData Sets

6. CONCLUSION

In conclusion, in this paper, we have described our proposal on the enhancement made on missing value imputation using RAR in ERAR. The results show that ERAR offers better performance in terms of the time taken for the imputation process and the amount of memory usage, especially for the monotone pattern. This is possible as ERAR is designed in selecting the frequent items in datasets that are only related to missing values. Therefore, unnecessary frequent items can be ignored in generating the association rules. In terms of imputation accuracy, we found that both ERAR and RAR exhibit a decreasing rate of accuracy as the amount of missing values increases for data of arbitrary pattern, but this is not the case of data of the monotone pattern. With the findings, ERAR contributes to improving the imputation processes in terms of the speed in missing values imputation, and also the amount of memory used. For future work, we may consider the iterative missing-value completion method in the imputation process where we will observe whether the inclusion of the method will affect ERAR's performance.

ACKNOWLEDGEMENT

We would like to thank the Fakulti Teknologi Maklumat Dan Komunikasi, Universiti Teknikal Malaysia Melaka (UTeM) and the Centre for Advanced Computing Technologies (C-ACT), for supporting this research.

REFERENCES

1. A. A. Chavan and V. K. Verma, "**Treatment of Missing Values for Association Rules : A Recent Survey**," *Int. J. Comput. Appl.*, vol. 70, no. 26, pp. 1–4, 2013. https://doi.org/10.5120/12228-8274
2. H. Uenal, B. Mayer, and J. B. Duprel, "**Choosing Appropriate Methods For Missing Data In Medical Research : A Decision Algorithm On Methods For Missing Data**," *J. Appl. Quant. Methods*, vol. 9, no. 4, pp. 10–21, 2014.
3. J. L. Peugh and C. K. Enders, "**Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement**," *Rev. Educ. Res.*, vol. 74, no. 4, pp. 525–556, 2004.
4. T. Aljuaid and S. Sasi, "**Proper Imputation Techniques for Missing Values in Data sets**," in *In Data Science and Engineering (ICDSE), 2016 International Conference*, 2016, pp. 1–5.
5. B. Suthar, H. Patel, and A. Goswami, "**A Survey : Classification of Imputation Methods in Data Mining**," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 1, 2012.
6. R. Agrawal and R. Srikant, "**Fast Algorithms for Mining Association Rules**," in *very large data bases, VLDB, vol. , pp.*, 1994, pp. 487–499.
7. A. Ragel and B. Crmilleux, "**Treatment of Missing Values for Association Rules**," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 1998, vol. 1394, pp. 258–270.
8. J. Shen and M. Chen, "**A Recycle Technique of Association Rule for Missing Value Completion**," in *In 17th International Conference on Advanced Information Networking and Applications, 2003. AINA 2003. . IEEE.*, 2003, pp. 526–529.
9. T. P. Hong and C. W. Wu, "**Mining rules from an incomplete dataset with a high missing rate**," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 3931–3936, 2011. https://doi.org/10.1016/j.eswa.2010.09.054
10. K. Rameshkumar, "**A Novel Algorithm For Association Rule Mining From Data With Incomplete And Missing Values,**" *ICTACT J. Soft Comput.*, vol. 1, no. 4, pp. 171–177, 2011.
11. P. D. Allison, "**Missing Data**," 2002.
12. C. K. Enders, ***Applied Missing Data Analysis***. Guilford press, 2010.

13. J. Kaiser, "**Dealing with Missing Values in Data**," *J. Syst. Integr.*, vol. 5, no. 1, pp. 42–51, 2014.
14. G. Chhabra, "**A Comparison of Multiple Imputation Methods for Data with Missing Values**," *ResearchGate*, vol. 19, no. 10, p. 8, 2017.
15. S. Van Buuren, ***Flexible Imputation of Missing Data.***, Second Edi. CRC press., 2018.
16. K. S. Ingersoll, "**Missing Data**," in *Applied Missing Data Analysis*, Guilford Publications, 2013, pp. 3–12.
17. J. L. Schafer and J. W. Graham, "**Missing data: Our view of the state of the art**," *Psychol. Methods*, vol. 7, no. 2, pp. 147–177, 2002.
18. T. Aljuaid and S. Sasi, "," in *International Conference on Data Science and Engineering (ICDSE)*, 2016, no. September 2017, pp. 1–5. **Proper Imputation Techniques for Missing Values in Data sets** https://doi.org/10.1109/ICDSE.2016.7823957
19. A. N. Baraldi and K. C. Enders, "A**n introduction to modern missing data analyses**," *J. Sch. Psychol.*, vol. 48, no. 1, pp. 5–37, 2010.
20. J. Wu, Q. Song, and J. Shen, "**Missing Nominal Data Imputation Using Association Rule Based on Weighted Voting Method**," in *2008 International Joint Conference on Neural Networks (IJCNN 2008)*, 2008, pp. 1158–1163.
21. R. Agrawal, H Verkamo. Mannila, R. Srikant, H. Toivonen, "**Fast discovery of association rules**," *Advances in knowledge discovery and data mining*, vol. 12. pp. 307–328, 1996.
22. A. Ragel and B. Cremilleux, "**missing valuesC — a Preprocessing Method To Deal With Missing Values**," *Res. Dev. Expert Syst. XV. Springer, London*, vol. 12, p. 159–170., 1999.
23. F. N. M. Leza and N. A. Emran, "**Data accessibility model using QR code for lifetime healthcare records**," *World Appl. Sci. J.*, vol. 30, no. 30 A, pp. 395–402, 2014
24. N. A. Emran, S. Embury, and P. Missier, "**Measuring Population-Based Completeness for Single Nucleotide Polymorphism (SNP) Databases**," in *Advanced Approaches to Intelligent Information and Database Systems*, J. Sobecki, V. Boonjing, and S. Chittayasothorn, Eds. Cham: Springer International Publishing, 2014, pp. 173–182.
25. N. A. Emran, S. Embury, P. Missier, and A. K. Muda, "**Measuring Data Completeness for Microbial Genomics Database**," in *ACIIDS 2013 Part 1*, 2013. https://doi.org/10.1007/978-3-642-36546-1_20
26. N. Z. Abidin, A. R. Ismail, and N. A. Emran, "**Performance analysis of machine learning algorithms for missing value imputation**," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 6, 2018.
27. N. M. Said, Z. M. Zin, M. N. Ismail, and T. A. Bakar, "**Comparative analysis of missing data imputation methods for continuous variables in water consumption data**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.6 Special Issue, pp. 471–478, 2019.
28. F. H. Mausor, J. Jaafar, and S. Mohdtaib, "**Fuzzy C means imputation of missing values with ant colony optimization**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1 Special Issue 3, pp. 145–149, 2020. https://doi.org/10.30534/ijatcse/2020/2191.32020
29. W. Y. Lai, K. K. Kuok, S. Gato-Trinidad, and K. X. L. Derrick, "**A study on sequential K-nearest neighbor (SKNN) imputation for treating missing rainfall data**," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 3, pp. 363–368, 2019. https://doi.org/10.30534/ijatcse/2019/05832019