A Lightweight Association Rules Based Prediction Algorithm (LWRCCAR) for Context-Aware Systems in IoT Ubiquitous, Fog, and Edge Computing Environment

Asma Abdulghani Al-Shargabi
Faculty of Computing & IT
University of Science & Technology
Sana'a, Yemen
a.alshargabi@gmail.com

Francois Siewe
Software Technology Research Laboratory (STRL)
De Montfort University
Leicester, UK
fsiewe@dmu.ac.uk

Abstract- Proactive is one main aspect of ubiquitous context-aware systems in IoT environment. The power of artificial intelligent is employed to realize this high-end aspect. Ubiquitous context-aware systems in IoT environment needs a light-weight intelligent prediction techniques especially within fog and edge computing environment where technologies capabilities are poor. On the other hand, in big data area, the amount of data used to train ubiquitous context-aware systems is huge. This paper suggests a light-weight prediction algorithm to help such systems to work more effectively. The proposed algorithm is an improvement of the RCCAR algorithm that we developed in previous work. RCCAR utilizes association rules for prediction. The contribution of this paper is to minimize the number of association rules by giving a priority to associations that produced of high order. The prediction is scored and formulated mathematically using the confidence measure of association rules. A real-world dataset is used to evaluate the proposed algorithm in various scenarios. The results show that the proposed algorithm achieves better prediction scores. For future work, extensive experiments with many datasets is recommended.

Keywords- *RCCAR; Prediction; Pervasive Computing; Ubiquitous Computing; Fog Computing, Edge Computing, Internet of Things (IoT), Context-aware systems, Association Rules, Data Mining, Big Data Analytics.*

I. INTRODUCTION

In Internet of Things (IoT) era, many intelligent technologies become available to facilitate tasks and improve human-computer interaction especially in fog computing and edge computing [1]. As a consequence, a huge volume, high velocity and more complex data is produced daily. IoT applications with a numerous different collection of sensors, wearable devices, and smart phones is one rich source of this data [2]. A valuable knowledge is embedded implicitly in this big data.

Context-Aware Systems (CAS) is one vital component in the wide range of pervasive/ubiquitous computing within IoT environment. CASs are systems that are aware of their situation (or context) in their physical, virtual (ICT) and user environment. CASs gather context based on big and variant sensors network. To realize CASs functions perfectly, CASs are in need to prediction to be proactive and to ensure that the gathered context is accurate, especially in real-time CASs [22].

In such environment, many algorithms have been developed for mining useful patterns to improve IoT CASs and its different applications and make it more intelligent and proactive. These algorithms encompass a different amazing collection of serial, parallel, distributed, and MapReduce-based algorithms run in local computers, distributed and parallel environments, clusters, grids, clouds and/or data centers [1][2]. Data mining on big IoT data has been a basic component in pervasive computing to discover interesting patterns for many IoT-related applications such as smart cities, urban analytics and mining social network data to detect communities [2].

Main data mining tasks in IoT big data include: (1) mining frequent pattern; (2) classification or supervised learning. After discovering frequent patterns, they can be used to form association rules or build associative classifier for

classification and/or prediction. Alternatively, other classifiers such as decision trees and random forests can be built [2].

This research employs association rules (AR) to develop predictive systems in ubiquitous, fog, and edge computing. In Fog and edge computing, we need to eliminate the data used in building the predictive models. RCCAR algorithm employs AR to build a predictive model using a little data as possible [7][8][9]. The proposed algorithm is an improvement of RCCAR algorithm. It is more efficient lightweight version to work in IoT Fog and edge and computing environment.

The rest of this paper is layout as follows: Section II introduces the related work. An overview of RCCAR algorithm is introduced in Section III. Section VI is devoted to introduce the proposed algorithm. Section V shows the algorithm implementation, experiments, and findings. Finally, Section IV conclude the work and discusses the potential future work.

II. RELATED WORK

Association Rules Mining is a technique that finds frequent patterns from given database. Patterns are represented as rules with two main measures, rule support, and rule confidence. These rules called association rules that describes the patterns in data. As name describes, these rules describes the associations between data value, which values come together [11][12]. Association rules are formulated as follows [12]:

$$\boldsymbol{x \Rightarrow y}, \quad [\boldsymbol{Support} = \boldsymbol{a}\%, \ \boldsymbol{confidence} = \boldsymbol{b}\%]$$

which means the occurrence of x value comes associated with occurrence y with support value a% and confidence value b%. Rule support and confidence respectively reflect the usefulness and certainty of discovered rules. A support of a% means that a% of all the transactions under analysis show that x and y come together. A confidence of b% means that b% of x occurrences comes with y.

Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. These thresholds can be set by users or domain experts [11] [12]. Support and confidence are calculates as follows [12]:

$$\text{Support } (\boldsymbol{x \Rightarrow y}) = \frac{count(X \cap Y)}{n}$$

$$\text{Confidence } (\boldsymbol{x \Rightarrow y}) = \frac{count(X \cap Y)}{count(X)}$$

Where $\boldsymbol{n}$ is number of transactions in database.

AR proves significant success in different IoT applications. For instance, authors in [1] used an algorithm that uses AR for intelligence in the Internet of Vehicles (IoV). In [3] an occupational adaptive classifier is designed based on association rules using LinkedIn data to help other areas of research and industrial applications, such as career adaptability judgment, interest training, career development, and career recommendation.

In [4], an association classification algorithm has been introduced and employed to help solving one important cybercrime web security problem, which is phishing websites. In education area, [5] applied AR to find useful knowledge for supporting admission policies planning. In [6], AR is applied to introduce a predicting method for a city's overall traffic state. As case study, the traffic system in Shanghai is considered. Authors in [14], introduced a method to analyze users workflow preferences using association rules. The method suggested filtering the rules using rules confidence and rules lift value to obtain preferred workflow for a specific user. Lift value is association rule measure that reflect how much occurrences of two items affect each other.

Authors in [1] developed an accurate weighted interest degree recommendation algorithm using association rules. However, this algorithm still use all parameters in data. In [4], an association classification algorithm has been introduced. The algorithm is called PWCAC. It introduced a new method for generating the association rules and suggested a new method for predicting the class of new data [4]. In [13], a new enhancement of AR is introduced. The authors in [13] combined the popular AR algorithm that called Apriori with the probabilistic graph model to improve the accuracy of prediction. An interesting contribution is proposed in [15], where a novel association classification and regression algorithm based is introduced.

The main contribution of this research is to minimize the time complexity of association rules technique and improve the AR efficiency significantly, where efficiency is one main drawback of AR. At first, the clustering algorithm is used to discretize the transaction data. Then, the discrete transaction database is transformed into Boolean matrix based on matrix operation. A new frequent itemsets tree structure is proposed, and all frequent itemsets can be obtained when the tree structure is constructed. Finally, the rules are reconstructed. So, the database is scanned once in comparable with Apriori algorithm that scan the database k time. Where k is the maximum order of combinations of items in database.

A similar contribution is introduced in [16], an improvement of Apriori algorithm is suggested to enhance time complexity. The method transforms the transaction database into the upper triangular matrix (UTM) UTM with support count. Using this method, the frequent itemsets can be generated by scanning database only once, which greatly improves the efficiency of the algorithm and reduces the computational complexity. Authors in [17] introduced a new method to filter ARs in interesting manner. The Association rule is given a weight to indicate its importance based on its both support and confidence measures collectively. This method in weighting achieves a good success in prediction accuracy.

In [18], Authors introduced an algorithm called MDPrefR to minimize the number of important ARs. The main idea of this algorithm is creating a reference rule that realizes the best AR three measures support, confidence, and pearl, and then calculating the degree of similarity of all rules one and finally removing rules that are not similar to the reference rule. An interesting method, which is called PSTMiner is introduced by [19] to prune AR model and so improve the efficiency. This method used chi-square statistical measure to prune the redundant rules. In [20] authors introduces a modification of classic Apriori algorithm by filtering the produced rule based on minimum support measure value while generating the candidate itemsets. [21] Proposed n-cross validation technique to reduce association rules which are irrelevant.

III. OVERVIEW OF RCCAR ALGORITHM

RCCAR is an algorithm that uses association rules in a particular way to outweigh one of many conflicting values of the same data item that comes from different sensors [7]. So, it is used basically for resolving conflicts in sensors data. It is developed as a light-weight algorithm to be used in context-aware systems within ubiquitous computing environment. However, RCCAR is then improved to introduce a solution for many sensors data shortcomings beside conflicts such as missing values and erroneous values [7][9]. RCCAR predicts the value or outweigh the correct value using association rules. On the other hand, RCCAR algorithm is improved to simplify the algorithm time complexity to be more efficient in working in ubiquitous computing environment. This improvement was by choosing the most important variables using decision tree before producing associations [8].

RCCAR depends on the fact that the occurrence of any data item/feature item will be associated with occurrences of other data items/feature items. RCCAR introduced a light solution for context data prediction in ubiquitous context-aware systems. It employed the association rules to predict the context data items by studying the associations that link this target data item and other occurred items. RCCAR is introduced and formulated in [7][8]. The mathematical model is built on the association rules mathematical model. RCCAR uses confidence scores of association rules that are produced previously off-line for the data under investigation. It uses all association rules that related to the investigated context data. The predicted value is that value with largest confidence score. The confidence score comes from summing all association rules confidence collectively. It uses all orders of associations to get the confirmation. Formula 1 describes this mechanism [7]:

$$\boldsymbol{Prediction\ Score}(\boldsymbol{x}) = \sum_{\boldsymbol{k=2}}^{\boldsymbol{w}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{d_k(x)}} \boldsymbol{confidence}(\boldsymbol{y_i(k)} \Rightarrow \boldsymbol{x})$$

Where x is the predicted value, y is the other variables values, and w is the maximum number of itemsets according to occurrences in context database for the

current values of context, and d is the number of associations. k is starting with 2 because the associations start with two elements. The equation shows that the associations which will be considered in the summation are just those with x in the right side [7].

IV. THE PROPOSED ALGORITHM

The proposed algorithm is called Lightweight RCCAR algorithm "LWRCCAR". The introduced solution is improvement of RCCAR algorithm which is based on Association rules. The introduced solution aims at simplifying RCCAR mechanism by using just the larger item set of association rules in prediction. This method should achieve good prediction and also should improve the efficiency of RCCAR by simplifying the calculations and minimizing the time complexity. In proposed approach we do not need for all associations of all orders, just the largest order associations will be considered. It would be useful to have more efficient version of RCCAR algorithm. This version can be efficiently employed in and edge computing and ubiquitous computing environment in general where time is a critical factor.

The proposed approach also uses the association rules and consequently its measures (confidence) to formulate the solution. The predicted value with the best/maximum prediction score is formulated as shown by Formula 2, and Formula 3:

$$\boldsymbol{Predicted_value}(\mathbf{x}) = \boldsymbol{x_j}$$

$$\boldsymbol{Prediction_Score}(\boldsymbol{x_j}) = \boldsymbol{Max}(\sum_{\mathbf{i=1}}^{\boldsymbol{n}} \boldsymbol{confidence}(\boldsymbol{y_i}) \Rightarrow \boldsymbol{x_j})$$

Where x is the predicted value, $\boldsymbol{x_j}$ is all values of variable x, n is the number of association rules with the larger order that is related to $\boldsymbol{x_j}$ and $\boldsymbol{y_n}$ is association rule itself.

V. IMPLEMENTATION AND EXPERIMENTS

The proposed solution is implemented using WEKA tool. It is a simple and light tool for most data mining tools. Many Experiments were conducted to prove the proposed solution. Experiments include prediction using the largest order association (as the new approach proposed), the next order associations, and the 1st order associations to compare the results and show that there are no need to use all association rules as original RCCAR did. The used dataset is well-known dataset. It is appropriate for the nature of approach. It is real weather historical data with long time depth. It also has a suitable volume for prediction and also has many dimensions/features.

The dataset is Southampton monthly weather historical data (1855-1999) [10]. It is officially recorded by Southampton Weather Station. This dataset contains 1744 instance. Variables are year, month, temperature max degree, temperature min degree, air frost, rainfall, and sunshine hours. The data is cleaned and preprocessed before experiments. The experiments used three variables to predict the temperature max degree, these variables are month, temperature min degree, and rainfall. The continuous variables are transformed to be nominal variables as AR requires.

The foundings shows interested results. The proposed approach achieves a remarkable succes. As shown by Figure 1, using 3-order associations acheieves the best results againts the less orders associations. It is the best even with different data time depth. As concequently, we are not in need to use the confirmation using all associations as RCCAR did, the bigger order is enough. This results will improve the effeciency of RCCAR.

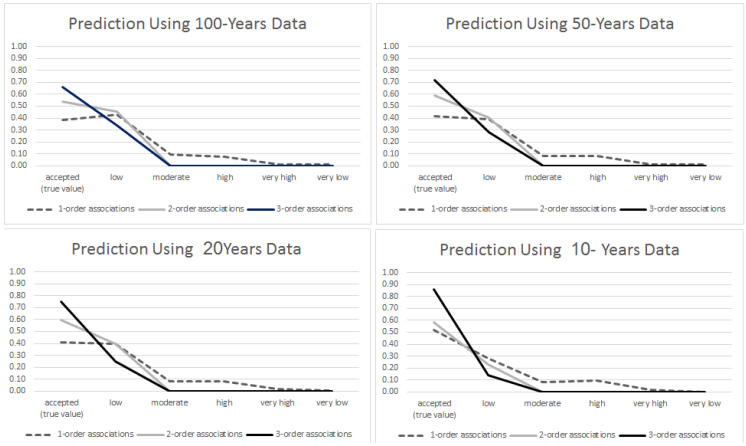


Figure 1. Prediction of "temperature max degree" using different orders of associations of other variables

The prediction results using 10 years' time depth inspires conducting further study using smaller time depth. The experiments included 5, 4, 3, and 2 years. Obviously, the best result was when using 2-Years' time depth. This result was expected where the weather is a natural phenomenon that change slowly over the years. And by examine the results in Figure 3, the best prediction was with 2 years, 5 years, and 50 years. This result is consistent with our simple information about climate where the climate has a natural cycle over every 50 years.

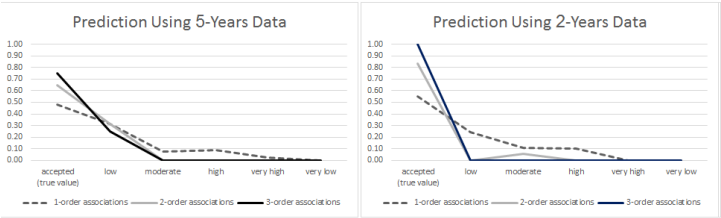


Figure 2. Prediction of "temperature max degree" using small data time depth

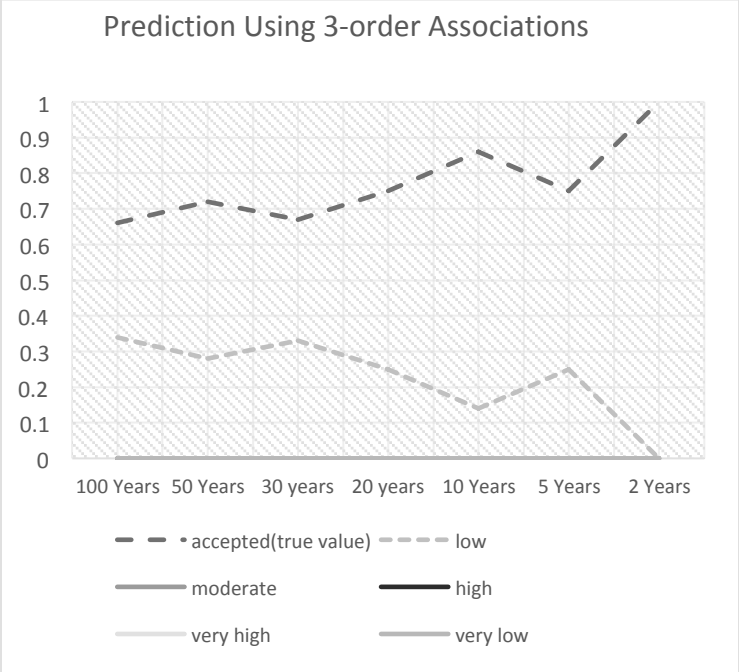


Figure 3. Prediction of "temperature max degree" using 3-order associations with other variables and different data time depth.

VI. CONCLUSIONS & FUTURE WORKS

As shown in results, the proposed approach has achieved good success against RCCAR. By examining the results carefully, some points can also be concluded: (1) Choosing the appropriate time depth/frame for the prediction improves the results; (2) the prediction using proposed approach does not depend on the adjacent previous data, so it is expected to work well with any type of data within any context. In other word, the data does not need to have a linear regression between the target feature and other different features; (3) the produced prediction model is more efficient than the original RCCAR.

For future works, it is recommended to compare the introduced solution with other prediction approaches. In addition, it will be useful to examine the solution with other different types of data and different context.

Finally, because of RCCAR is introduced for prediction in ubiquitous context aware systems, the proposed approach will work better within this environment in terms of efficiency. It is expected to work well even within edge computing applications.

REFERENCES

[1] FUHONG LIN, YUTONG ZHOU, ILSUN YOU, JIUZHI LIN, XINGSHUO AN, XING LÜ, "Content Recommendation Algorithm for Intelligent Navigator in Fog Computing Based IoT Environment", IEEE Access, SPECIAL SECTION ON COLLABORATION FOR INTERNET OF THINGS, VOLUME 7, 2019.

[2] Peter Braun, Alfredo Cuzzocrea, Carson K. Leung, Adam G. M. Pazdor, Joglas Souza, Syed K. Tanbeer, " Pattern Mining from Big IoT Data with Fog Computing: Models, Issues, and Research Perspectives ", IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), 2019.

[3] Huayou Si, Haopeng Wu, Li Zhou, Jian Wan, Naixue Xiong, Jilin Zhang, " An Industrial Analysis Technology about Occupational Adaptability and Association Rules in Social Networks ", DOI 10.1109/TII.2019.2926574, IEEE Transactions on Industrial, 2019.

[4] Mohammed Alqahtani, " Phishing Websites Classification using Association Classification (PWCAC)", IEEE International Conference on Computer and Information Sciences (ICCIS), 2019.

[5] Pornthep Rojanavasu, "Educational Data Analytics using Association Rule Mining and Classification", IEEE The 4th International Conference on Digital Arts, Media and Technology and 2nd ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering, 2019.

[6] Chengjue Yuan, Xiangxiang Yu, Dewei Li , and Yugeng Xi, " Overall Traffic Mode Prediction by VOMM Approach and AR Mining Algorithm With Large-Scale Data ", IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 20, NO. 4, APRIL 2019.

[7] Asma A. Al-Shargabi, Francois Siewe, "Resolving Context Conflicts Using Association Rules (RCCAR) to Improve Quality of Context-Aware", 8th IEEE International Conference on Computer Science & Education (ICCSE 2013), Colombo, Sri Lanka, 2013.

[8] Asma A. Al-Shargabi, Francois Siewe, "An Efficient Approach for Realizing RCCAR to Resolve Context Conflicts in Context-Aware Systems", 2nd Conference on Control, Systems & Industrial Informatics (ICCSII 2013), Bandung, Indonesia, 2013.

[9] Asma Abdul Ghani Al-Shargabi, Francois Siewe, "A multi-layer framework for quality of context in ubiquitous context-aware Systems", International Journal of Pervasive Computing and Communications, Emerald, UK, 2018.

[10] http://www.southamptonweather.co.uk/sotonhist.php

[11] Pooja Akulwar, Sujata Pardeshi, Arvind Kamble, "Survey on Different Data Mining Techniques for Prediction", Proceedings of the Second International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2018), 2018.

[12] Jiawei Han & Micheline Kamber, Jian Pei, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, Elsevier, 3rd edition, 2016.

[13] Gehao Sheng, Huijuan Hou, Xiuchen Jiang, Yufeng Chen, " A Novel Association Rule Mining Method of Big Data for Power Transformers State Parameters Based on Probabilistic Graph Model ", IEEE TRANSACTIONS ON SMART GRID, VOL. 9, NO. 2, MARCH 2018.

[14] Mohd Anuaruddin Bin Ahmadon, Shingo Yamaguchi, " User Workflow Preference Analysis Based on Confidence and Lift Value of Association Rul", IEEE 7th Global Conference on Consumer Electronics (GCCE), 2018.

[15] Ling Wang, Hui Zhu, and RuiXia Huang, "A Novel Association Rule Prediction Algorithm for Classification and Regression", IEEE Conference on Decision and Control (CDC), Miami Beach, USA, 2018.

[16] Mengmeng Cao, Chaoyou Guo, " Research on the Improvement of Association Rule Algorithm for Power Monitoring Data Mining", IEEE 10th International Symposium on Computational Intelligence and Design, 2017.

[17] Mohammed Hayel Refai, Yuhanis Yusof, " Group-based Approach for Class Prediction in Associative Classification", IEEE 3rd International Conference on Information Retrieval and Knowledge Management, 2016.

[18] Mouhir Mohammed, Gadi Taoufiq, Balouki Youssef, El Far Mohamed, " A New way to select the valuable association rules", IEEE 7th International Conference on Knowledge and Smart Technology (KST), 2015.

[19] K.Prasanna Lakshmi, C.R.K.Reddy, " Fast Rule-Based Prediction of Data Streams using Associative Classification Mining ", IEEE 5th International Conference on IT Convergence and Security (ICITCS), 2015.

[20] Lalit Mohan Goyal, M. M. Sufyan Beg, "An efficient filtration approach for mining association rules", IEEE International Conference on Computing for Sustainable Global Development (INDIACom), 2014.

[21] K.Rameshkumar, M.Sambath, S.Ravi, "Relevant Association Rule Mining from Medical Dataset Using New Irrelevant Rule Elimination Technique", IEEE International Conference on Information Communication and Embedded Systems (ICICES), 2013.

[22] A.A.Q Al-Shargabi, F. Siewe, A. Zahary, "Quality of Context in Context-Aware Systems ", International Journal of Pervasive Computing and Communications, UK, 2018.