iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: VI Month of publication: June 2019

DOI: http://doi.org/10.22214/ijraset.2019.6430

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

Exploring the Data Stream Size using ICBMI and CBMI Algorithm

P. Logeshwari
Assistant Professor, Department of Computer Science, SNMV College, Malumachampatti, India.

Abstract: ICBMI and CBMIboth the algorithms react to changes in the data stream. By entering the second part of the Dataset, as the concept changes, the ICBMI gradually reduces the average of the data stream sizes while the CBMI abruptly reduces the data stream size to remove the obsolete transactions and their effects on the mining result. At this point, for the ICBMI, the average of data stream sizes now contains both these concepts. In fact the ICBMI, for some Missing Data, therefore the max frequency data streams reside in both new and old parts of the Dataset because it considers the complete history of the input stream and does not completely discard the previous concept.

Index Terms--stream Data, Multiple Imputation, Missing Data, ICBMI, and CBMI

I. INTRODUCTION

The first version of CBMI is computationally expensive, because it checks exhaustively all "large enough" sub data streams of the current data stream for possible cuts. Furthermore, the content of the data stream is kept explicitly, with the corresponding memory cost as the data stream grows. To reduce these costs a new version ICBMI is introduced. ICBMI uses ideas developed in data stream algorithms to find a good cut point quickly. This data structure is a variation of exponential histograms, a data structure that maintains an approximation of the number of 1's in a main stream Data Multiple Imputation of length W with logarithmic memory and update time. Also this structure adopted in a way that can provide this approximation simultaneously for about O(logW) sub data streams whose lengths follow a geometric law, with no memory overhead, with respect to keeping the count for a single data stream. That is, our data structure will be able to give the number of 1s among the most recently $t - 1$, $t - b_{cc}$, $t - b_{c2c}$,....,$t - b_{cic}$,… read bits, with the same amount of memory required to keep an approximation for the whole W. Keeping the exact counts for a fixed data stream size is probably impossible in sub linear memory. We go around this problem by shrinking or enlarging the data stream strategically, so what would otherwise be an approximate count happens to be exact.

II. PERFORMANCE OF CBMI AND ICBMI ALGORITHM

In this section we present our ICBMI algorithm. The ICBMI uses simplified Chernoff bound concepts to calculate the appropriate data stream size for mining Missing Data. It uses now the comparison of the two data stream sub-range observations and Data counts when a segment occurs within the data stream and then adjusts the data stream size appropriately. However, for an itemset, its max-frequency data stream moves forward if a data stream with higher frequency is found in the reminder of the input stream. The following table clearly represents the efficiency between ICBMI and the CBMI for different Datasets UNF-BARCLAY'S DATA SET, UNF-BOEING DATA SET, BARCLAY'S DATA SET, and BOEING DATA SET

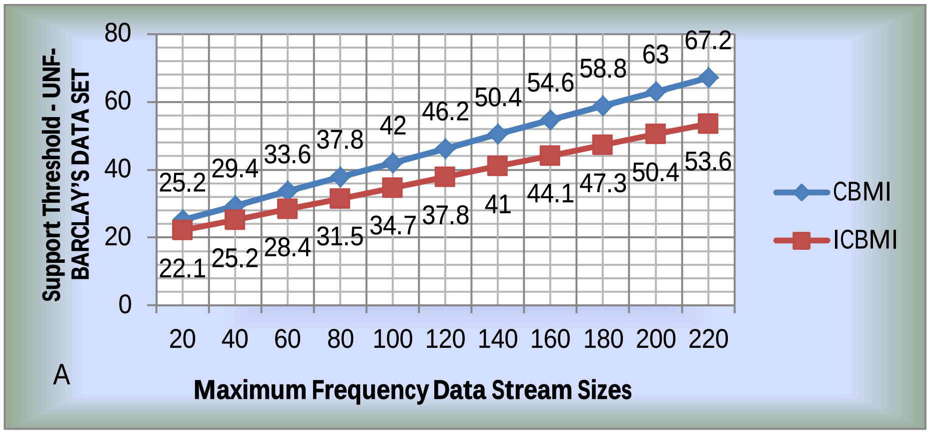
Input: T = set of all observed transaction IDs
δ = required data streams
Output: t = set of all present Missing Data IDs
Initialise: $\forall i \in T, w_i \leftarrow 1$
while(getNextTransaction) **do**
  for (i in T)
processData stream(W_i) →,$p_{i,t}$'s, p_i^{avg},$|S_i|$
    if (itemExist($|S_i|$)
      output i
    end if
$w_i^* \leftarrow$ requiredData streamSize(p_i^{avg} ,δ)
if (itemexists ^ $|S_{2i}| = 0$)
$w_i \leftarrow \max(\min\{w_i/2, w_i^*\}, 3)$
    else if (detectTransaction($|S_i|$,w_i, p_i^{avg}))
$w_i \leftarrow \max\{(w_i - 2),3\}$
    else if ($w_i^* > w_i \wedge |S_i| < w_i p_i^{avg}$)
$w_i \leftarrow \min\{(w_i + 2), w_i^*\}$
    end if
  end for
end while

CBMI Algorithm

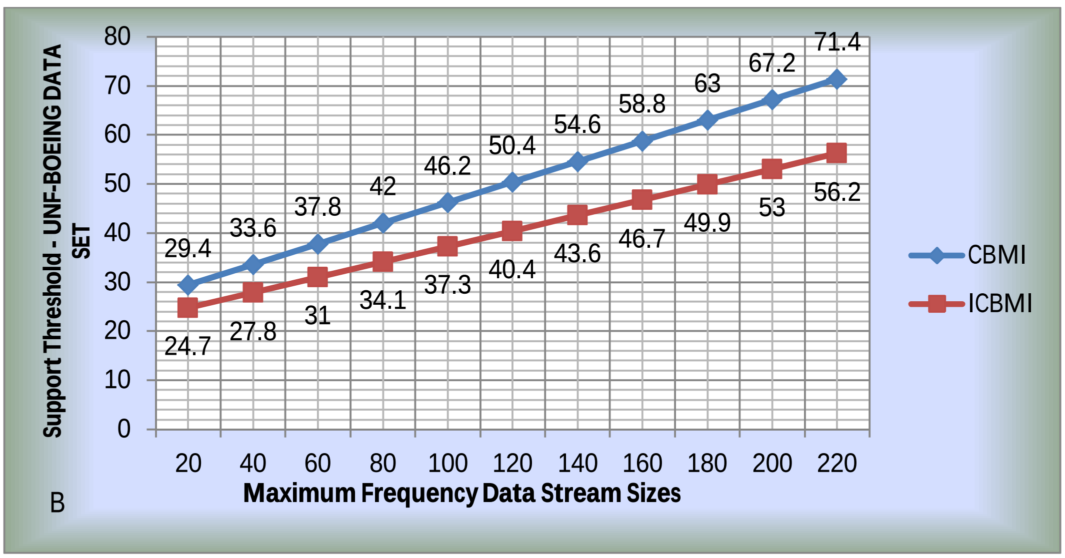
International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

III. COMPARISON OF ICBMI AND CBMI

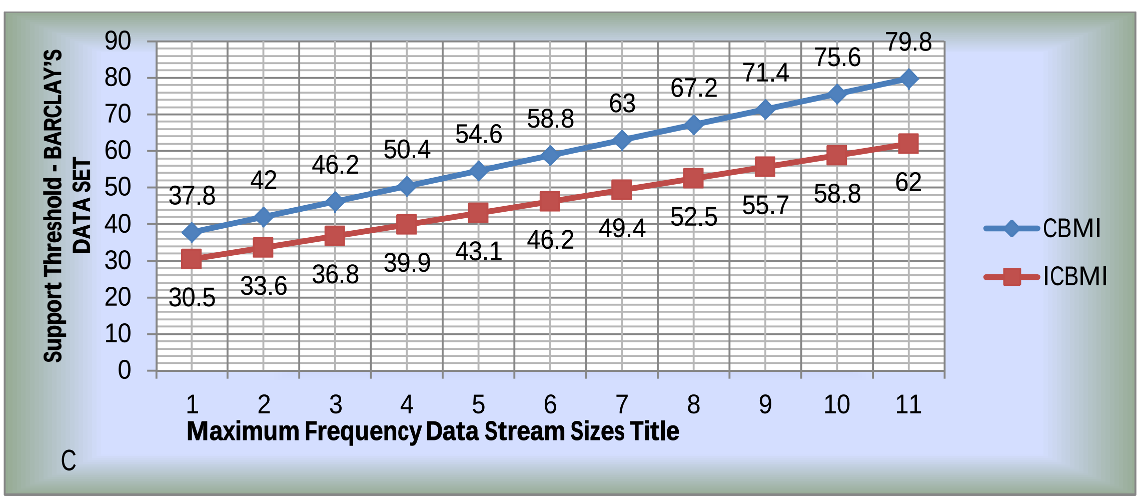
The Data and its max-frequency data stream information are now deleted, if its support falls below the Support Threshold. Through experiment, after each segment the, average of the Maximum Frequency Data Stream Sizes of the ICBMI was compared to the current data stream size of CBMI.



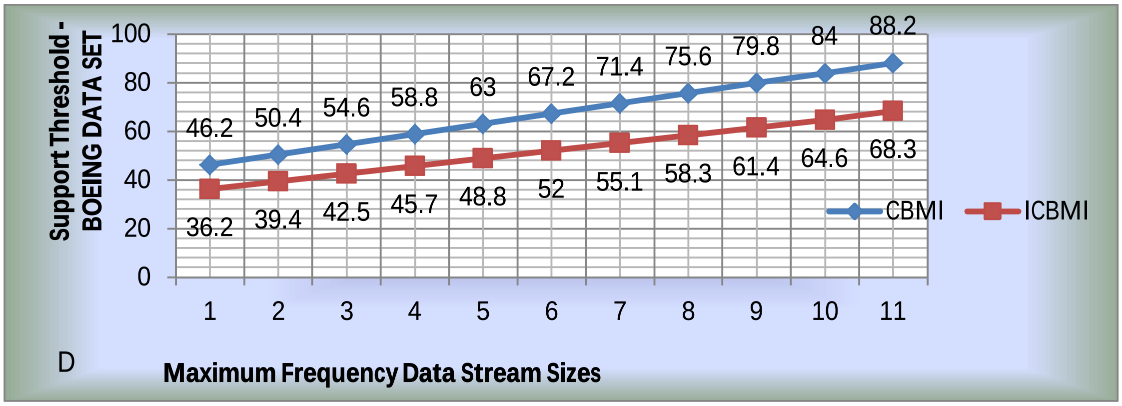
(a)



(b)



(c)



(d)

Fig. 2.Comparison of ICBMI and CBMI

The above figure shows how the data stream sizes of both the algorithms change as new segments are received. As shown in this figure1 (a), (b), (c) and (d) both the algorithms react to changes in the data stream. By entering the second part of the Dataset, as the concept changes, the ICBMI gradually reduces the average of the data stream sizes while the CBMI abruptly reduces the data stream size to remove the obsolete transactions and their effects on the mining result. At this point, for the ICBMI, the average of data stream sizes now contains both these concepts. In fact the ICBMI, for some Missing Data, therefore the max frequency data streams reside in both new and old parts of the Dataset because it considers the complete history of the input stream and does not completely discard the previous concept.

IV. CONCLUSION

Through experiment, after each segments the, average of the Maximum Frequency Data Stream Sizes of the ICBMI was compared to the current data stream size of CBMI. The ICBMI gradually reduces the average of the data stream sizes while the CBMI abruptly reduces the data stream size to remove the obsolete transactions and their effects on the mining result. At this point, for the ICBMI, the average of data stream sizes now contains both these concepts.In fact the ICBMI, for some Missing Data.

REFERENCE

[1] ERTE PAN, (Student Member, IEEE), MIAO PAN, (Member, IEEE), AND ZHU HAN, (Fellow, IEEE)Tensor Voting Techniques and Applicationsin Mobile Trace Inference, IEEE Access SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE ENABLE NETWORKING,VOLUME 3, 2015 Received October 30, 2015, accepted November 16, 2015, date of publication December 24, 2015, date of current version January 7, 2016.

[2] Ms.R.Malarvizhi and Dr.Antony Selvadoss Thanamani Cluster Based Mean Imputation International Journal of Research and Reviews in Applicable Mathematics& Computer Science.Vol 2.No.1,2012,

[3] Bayesian Learning of Noisy Markov Decision Processes,ACM Transactions on Modeling and Computer Simulation Vol. 23, No. 1, Article 4, Publication date: January 2013.SUMEETPAL S. SINGH, University of Cambridge

[4] Yosio Edemir Shimabukuro, Jukka Miettinen, René Beuchle, Rosana Cristina Grecchi,Dario Simonetti, and Frédéric Achard Estimating Burned Area in Mato Grosso, Brazil,Using an Object-Based Classification Method on a Systematic Sample of Medium Resolution Satellite Images,IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 8, NO. 9, SEPTEMBER 2015,

[5] Xuanyu Zhao, Huafeng Zhou, Di Shi, Huashi Zhao, Chaoyang Jing, Chris Jones On-Line PMU-Based Transmission Line Parameter Identification,CSEE JOURNAL OF POWER AND ENERGY SYSTEMS,VOL. 1, NO. 2, JUNE 2015,

[6] Ms.R.Malarvizhi and Dr.Antony Selvadoss Thanamani Cluster Based Mean Imputation,International Journal of Research and Reviews in Applicable Mathematics& Computer Science.Vol 2.No.1,2012,

[7] Ms.R.Malarvizhi and Dr.Antony Selvadoss Thanamani.K-NN Classifier Performs Better Than K-Means Clustering in Missing Value Imputation,International Journal for Research in Science & Advanced Technologies,Vol 1.Issue-2,2013,

[8] S.Kanchana and Dr.Antony Selvadoss Thanamani.Classification of Efficient Imputation Method for Analyzing Missing Values,International Journal of Computer Trends and Technology(IJCTT),Vol 12.No.4-Jun 2014 ,

[9] S.Kanchana and Dr.Antony Selvadoss Thanamani Multiple Imputation of Missing Data Using Efficient Machine Learing Approach,International Journal of Applied Engineering Research,Vol 1.No.1 ,2015,

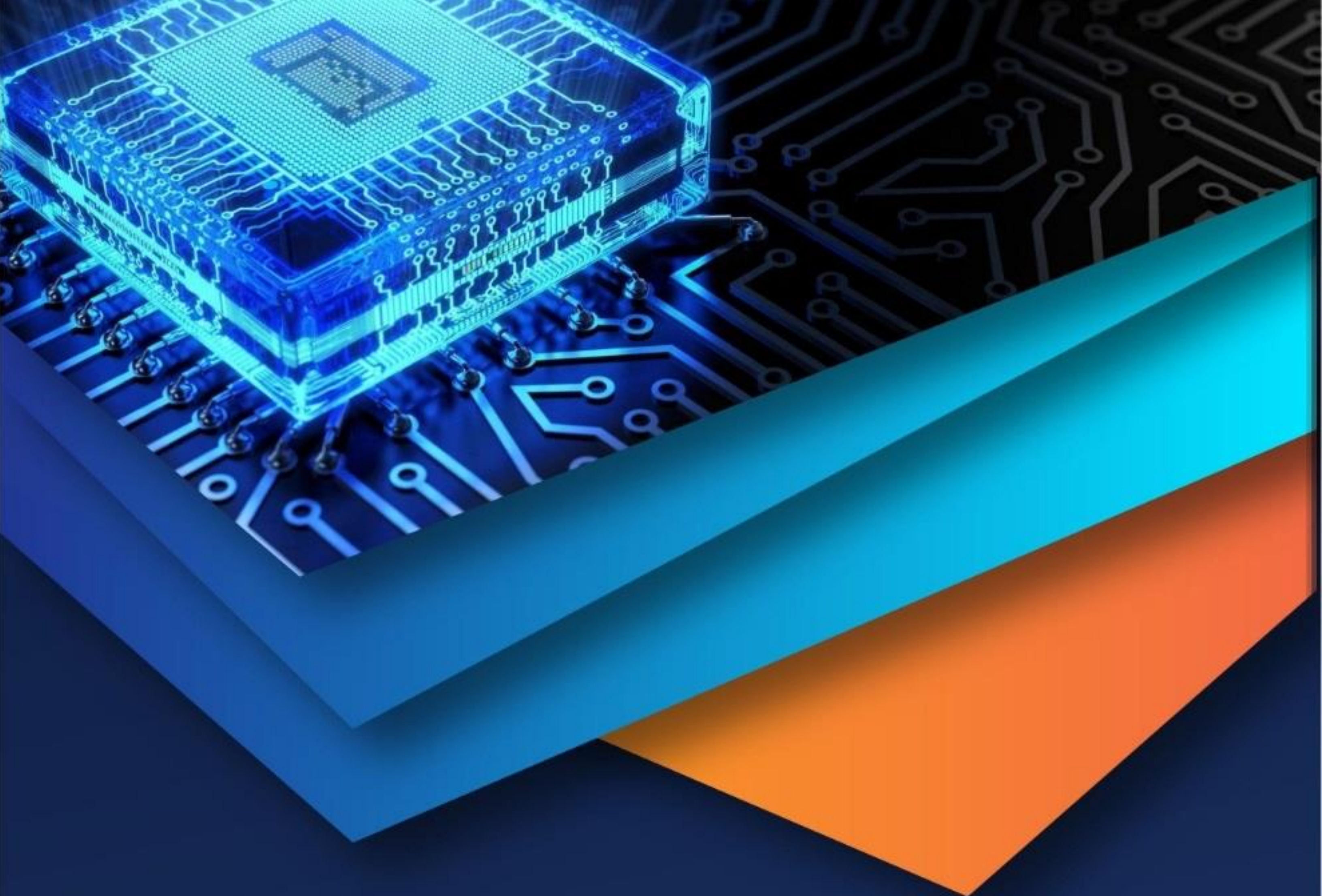
[10] Ms.R.Malarvizhi and Dr.Antony Selvadoss Thanamani..K-NN Classifier Performs Better Than K-Means Clustering in MissingValue Imputation Journal: International ournal for Research in Science & Advanced Technologies,Vol 1.Issue-2,2013,

[11] Mrs. P.Logeshwari and Dr.Antony Selvadoss Thanamani.A Survey On Missing Data And Methods To Find TheMissing Values International Journal For research In Science And Technology Volume 1, 2015,

[12] Mrs. P.Logeshwari and Dr.Antony Selvadoss Thanamani.Assignable Algorithms Available forMissing Data for Finding MV, International Journal Of Advanced Networking and Applications (IJANA), Special Issue,2015,

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue VI, June 2019- Available at www.ijraset.com

Mrs. P. Logeswari received her MCA., degree in computer Science from Sree Saraswathi Thiyagaraja College of arts and science, Pollachi, India in 2010. She completed her M.Phil., degree in computer Science from Sree Saraswathi Thiyagaraja College of arts and science, Pollachi, India on 2012 . She completed her PhD (Full Time) degree in Computer Science in NGM College (Autonomous), Pollachi under Bharathiar University, Coimbatore. She served as a Faculty of Computer Science at Government Arts College Udumalpet, from 2012 to 2013 and she served as a Faculty of Computer Science at Sree Ramu College of Arts and Science, NM Sunggam,Pollachi,India.from April 2013 to August 2014.Currently she is working as a Assistant Professor in SNMV college of arts and science, Coimbatore. She has presented papers in International/National conferences and published two papers in International journal. Her research focuses on Data Mining.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY