Recurrent Pattern Analysis with Data Slicing and Conceptual Model Used for Supervision based Data Modeling

Suraj Prakash Yadav [1], Dr. A. Y. Suriya [2], Dr. S.B. Kishor [3]

[1] Research Scholar, Gondwana University, Department of Computer Science, Faculty-Science and Technology
[2] Assistant Professor, Department of Computer Science, Janata Mahavidyalaya
[3] HoD, Department of Computer Science, Sardar Patel Mahavidyalaya, Chandrapur

Abstract

The data which is going to be obtained from any standard theory set can be used for various management needs based upon which the decision can be taken and the same decision which is going to be used for any purpose. The major decision-making system for any commerce or management related issues the gathered data is firstly classified based upon its intents. The intents which are used for passing the data or the information pertaining to the contents and the broadcast system will be very much useful for the start-ups where all the real time challenges which a beginners may face before the startups can be minimized. The proposed techniques of the data partition can be done based upon the frequently or recurrent pattern. The recurrent pattern discovery based upon the partition made makes the tasks of the decision support and data mining based on slicing provides the easy way to use the analytics in the area of management. The slicing can be done based upon various attributes or the required parameters which forms the base of the context and the organizational goals. To enhance the security measures the quasi attributes can be used in the concept to be identified as the process knowledge management. The integration of the data will be provided based upon the organizational contexts to justify the cognitive approach and the sliced constructs of the available data to reach to exhibit the significant chain of identifying the extraction of the pure data and the analysis of the obtained determinants of the generalized and the bucketized data can be easily done.

Keywords: Data Slicing, Bucketization, Recurrent Pattern

1. Introduction

The data supervision is very much essential in the perspective of various usages and its implementation. There are many faces and phases of the data handling mechanisms and its varied inclusion in decision making and modeling techniques are used. Recurrent pattern analysis and slicing is one of the factors which can be implemented for various knowledge discovery and decision making [1]. It can be used for any startups, computational research or in any field where the data can be presented in various aspects. Slicing is made in such a way where a significant amount of data can be used and based upon the categorical decision it can be used for subsequent approach [2]. Micro data is typically extracted from huge databases for research purposes, hence one micro data set linked to any specialized area can be selected and data analysis is performed. It is widely divided into many categories based on data analysis, and partitioning pattern in all possible vertical implementation is done once the extracted data is found to be significant in a specified area.

DOI: 10.37082/IJIRMPS.2021.v09si05.015

2. Slicing Architecture

In the overall process of the slicing techniques the data is going to be collected and based upon the type it can be filtered. Type indicates the relevance and usefulness. The slicing can be useful in varied areas therefore the correct selection of the data from a vast database or the dataset can be used. Once the pattern is obtained then it can be extracted based upon the algorithms used for the filtration process or the segmentation process.

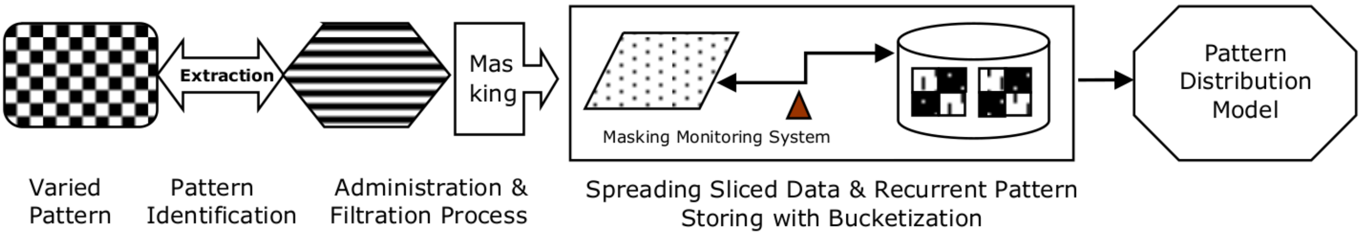
Extraction
Mas king
Masking Monitoring System
Pattern Distribution Model
Varied Pattern
Pattern Identification
Administration & Filtration Process
Spreading Sliced Data & Recurrent Pattern Storing with Bucketization

Figure 1: Initial Phase - Pattern Recognition & Distribution Model

Context 1: The major context which can be executed at the initial phase is the actual identification and the discovery of the pattern learning process. Once the extraction is initiated then the same can be used for the next level of the filtration process. The administration of the overall pattern & data management should be performed on every single element whether it can be obtained before or after the masking process.

Context 2: In the masking the allotment of the pattern token distribution and the allotment is performed. If it is found that any pattern has not undergone the masking or identification phase then there is one more time given by the monitoring system which is used before storing the data which is spread across the pattern formation. It can be then bucketized based upon the selected or specified attributes after which a significant model is used which allows the pattern is ready for the distribution.

Occurrence Process "D": Once the distribution (D) is ready to forward the pattern for its generalization or the bucketization then it can be used for the knowledge discovery process. The high dimensional data should be handled at varied process of implementation based upon which anonymization can be performed [5]. The randomly linking and the view on each segment of data is based upon the protection of the administered process of the privacy.

3. Data Characterization

Tuples are clustered based on quasi identifiers once attributes have been generated, as identifiers are unique and cannot be utilized as clustering elements if masking is not performed with clarity in pattern analysis. Clustering is done with sensitive attributes because they are the part which requires the protection from access from at this moment from the unauthorized use.

Data Volume	Fragment	Segmented	Pattern
500	100	1-7	Positive
200	200	2-8	Negative
50	300	6-11	Negative
600	400	1-9	Positive
700	500	1-4	Positive
900	600	2-4	Positive
20	100	4-5	Negative
1010	700	3-8	Positive
1500	800	9-11	Positive

Figure 2: Data Categorization

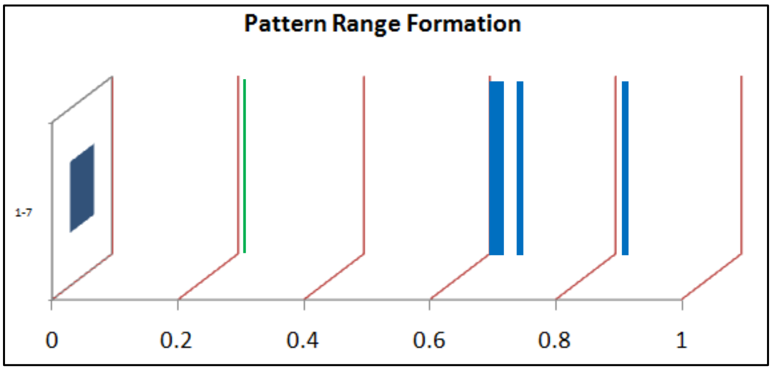

Figure 3: Pattern Matching Points Level

The characterization should be performed based upon the masking techniques used. As the quasi identifier is used means a part of the sensitive data which can be either accessed by the authorized users or it can be prevented from the access [15].

Base 1: In supervision based modeling the techniques used are at number of instances which are created during the clustering process. The model can be provided with the required attributes at Clustering (I-IV) Process Model. In the Clustering (I) the intensified data is carried out for pattern analysis i.e. C (I), once the identification has a pivot value in the positive dimension then the next level of the clustering process can be carried out. Clustering (II) involves the segmentation and it should be based on the analyzed data which showcase the size of the data consumed or used for the tuple generation. It should be also demonstrating the positive approach [16].

Base 2: Volume, Variety, Value and Visualization for the characterization can also be done. The structured data can be used for the storing in any data storage mechanisms [9]. The pattern matching levels are very important to decide the inclusion in the bucketization process. Every level is identified on every points where the pattern matching points are falling in the specified range or not and based upon this we can move to another level of clustering.

Base 3: The further analysis and the pattern classification and the modeling involves the implementation of Clustering (III) which significantly verifies the partitioned data also the masking strategy is being identified. The overall identification is verified based upon the masking identification which is being provided for every module which undergoes the process. Clustering (IV) is the phased where the decision can be initiated and thus the segments can be then stored in the tables or the required database and this will be the base of the distribution of the data for the classification of pattern and the sliced data now is ready for the decision making.

4. Modules Analysis

More information is preserved when many sets of exact values are used rather than generalization. Also, because both approaches keep the precise values of each attribute but break the relationship between them inside one bucket, this multi set based generalization is equal to a basic slicing scheme where each column holds exactly one attribute [21]. The referred pattern and the used fragmentation is then used for the categorization phase where the data pattern elements can be significantly used for providing the pattern accumulation in multi set.

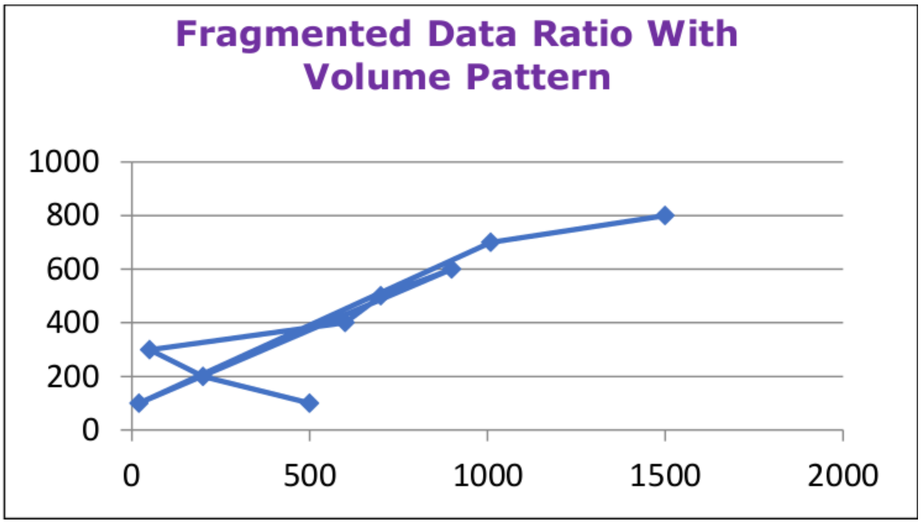

Figure 4: Actual Process of Fragmentation – Volume & Segments

4.1. Volume Administration

The separation of the two properties is not required for slicing. Slicing core model and concepts is to break the relationship between columns while preserving the association inside each column. Data's dimensionality is reduced, while its utility is preserved. Slicing divides the data in two directions: horizontally and vertically. High-dimensional data can also be handled using data slicing. The decision on the overlapping segments or the fragments can be then initiated for the only after the C (IV) strategic implementation.

Although a variety of anonymization techniques based on C (III) have been developed, now the main issues arises how we can proceed towards anonymization process of data which remains unsolved. In our experiments, we generate connections between bucket column values at random. Sometimes the process may reduce the usefulness of the data which should be avoided as much possible we can so that only the useful and the customized available fragments can be used for the attribute preservation.

4.2. Clustering Formation

The usefulness of the obtained data pattern is then leads to the process of the formation of the each and every clusters which can be implemented with the every data bit present in the main fragmented or the sub-fragmented based column. During the overall process the effective model will be selected and then the

DOI: 10.37082/IJIRMPS.2021.v09si05.015

clustering can be initiated. In every segment selection process the steps to be carried out are:

Step 1: Classify the Data Categorization (DC) Process
Step 2: Every Positive Data Fragments (PDF) should be marked for Masking
Step 3: Pattern Categorization (PC) should be formed.
Step 4: Clustering is done based as:
PDF < 0 (Fail) Do not proceed.
PDF > 0 (Pass) Proceed with fragment identification.
Step 5: Comparing PC value before and after clustering.

We also need to take proper measures to secure privacy by breaking the linkage of uncorrelated attributes and preserve data utility by keeping the association of highly correlated attributes by splitting attributes into columns. Only the obtained values should be used for incorporation of clusters and its pattern identification.

If all the above specified procedure is not followed then it may lead to an inconsistency implementation of fragments and pattern may arise. The data utility of the generalized data is significantly reduced as a result of this. Correlations between different qualities are lost in the generalized table because each attribute is generalized separately. This is a difficulty with generalization in general if not followed the proper assigned procedures. Bucketization is more useful for data than generalization, although it does not rule out membership or association disclosure. We need to assure that every execution should be giving so accurate generation of the pattern which can be used for decision making and implementation.

6. Conclusion

Slicing can be done based on a variety of features or required requirements, which serve as the foundation for the context and corporate objectives. The quasi characteristics can be employed in the notion to be identified as the process knowledge management to improve security measures. Every characterized pattern can be used in varied areas. It can be a health sectors, knowledge discovery and management decision making strategy or for any other pattern analysis approach as applicable. The data will be integrated based on organizational contexts to justify the cognitive approach and sliced constructs of the available data to show the significant chain of identifying the extraction of the pure data and the analysis of the obtained determinants of the generalized and bucketized data.

References

[1] Tiancheng Li, NinghuiLi,JiaZhang,and Ian Molloy, Slicing: A New Approachfor Privacy Preserving Data Publishing", Proc.ieee transactions on knowledgeand data engineering, Vol.24,No.3,March 2012

[2] Alina Campan, Traian Marius Truta, Nicholas Cooper, P-Sensitive K-Anonymity with Generalization Constraints",transactions on data privacy,Vol.65,No.3,2010

[3] Wei, D.; Li, C.; Naheman,W.;Wei, J.; Yang, J. Organizing and Storing Method for Large- scale Unstructured Data Set with Complex Content. In Proceedings of the 5th International Conference on Computing for Geospatial Research and Application,Washington, DC, USA, 4–6 August 2014; pp. 70–76

[4] Abelló, A. Big Data Design. In Proceedings of the 18th International Workshop on DataWarehousing and LAP, Melbourne, Australia, 23 October 2015; pp. 35–38

[5] G.Ghinita,Y.Tao,and P.Kalnis.On the anonymization of sparse high-dimensional data. In ICDE,pages 715-724,2008

[6] J.Li,Y.Tao and X.Xiao.privation of proximity privacy in publishing numerical sensitive data. In SIGMOD,pages 473-486,2008

[7] C. Dwork. Differential privacy. In ICALP, pages 1–12, 2006. [9] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In TCC, pages 265–284, 2006.

[8] J. Brickell and V. Shmatikov. The cost of privacy: destruction of data-mining utility in anonymized data publishing. In KDD, pages 70–78, 2008

[9] I. Dinur and K. Nissim. Revealing information while preserving privacy. In PODS, pages 202–210, 2003

[10] L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. Int. J.Uncertain. Fuzz., 10(6):571–588, 2002

[11] D. J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J. Y. Halpern. Worst-case background knowledge for privacy-preserving data publishing. In ICDE, pages 126–135, 2007

[12] M. Terrovitis, N. Mamoulis, and P. Kalnis. Privacy-preserving anonymization of set- valued data. In VLDB, pages 115–125, 2008

[13] Yogendra Kumar Jain, Vinod Kumar Yadav, Geetika S. Panday , An EffcientAssociation Rule Hiding Algorithm for Privacy Preserving Data Mining", IJCSE,Vol.4,No.18 ,2011

[14] Alina Campan, Traian Marius Truta, Nicholas Cooper, P-Sensitive K-Anonymity with Generalization Constraints",transactions on data privacy,Vol.65,No.3,2010

[15] Bee-Chung Chen, Daniel Kifer, Kristen LeFevre and Ashwin Machanavajjhala, "Privacy-Preserving Data Publishing", Foundationsand Trends in Databases Vol.2,No.12,2009

[16] Li.T and Li.N, “On the Tradeoff between Privacy and Utility in Data Publishing,” Proc.ACM SIGKDD Int"lConf.Knowledge Discovery and Data Mining (KDD), 2009

[17] K. LeFevre, D. DeWitt, and R. Ramakrishnan.Mondrian multidimensional k-anonymity. In ICDE, page 25, 2006

[18] A. Inan, M. Kantarcioglu, and E. Bertino, In ICDE, Using anonymizeddatafor classify"Vol.28,No.30,2009

[19] Yeye He, Je_rey F. Naughton,Anonymization of SetValued Data via TopDown,Local Generaliztionl.5,No.7,2009

[20] Wardani, D.; Küng, J. Semantic Mapping Relational to Graph Model. In Proceedings of the 2014 International Conference on Computer, Control, Informatics and Its Applications, Bandung, Indonesia, 21–23 October 2014; pp. 160–165

[21] Raju, N.V.S.L.; Seetaramanath, M.N.; Srinivasa Rao, P. A Novel Dynamic KCi - Slice Publishing Prototype for Retaining Privacy and Utility of Multiple Sensitive Attributes. Int. J. Inf. Technol. Comput. Sci. 2019, 4, 18–32

[22] Agarwal, S.; Sachdeva, S. An Enhanced Method for Privacy-Preserving Data Publishing, In Innovations in Computational Intelligence. Studies in Computational Intelligence; Panda, B., Sharma, S., Batra, U., Eds.; Springer: Singapore, 2018

[23] Brickell.J and Shmatikov, “The Cost of Privacy:Destruction of Data Mining Utility in Anonymized Data Publishing”, Proc.ACM SIGKDD int’l conf. Knowledge Discovery and Data Mining (KDD), 2008

[24] R. C.-W. Wong, A. W.-C. Fu, K. Wang, and J. Pei. Minimality attack in privacy preserving data publishing. In VLDB, pages 543–554, 2007