International Journal of Computer Engineering & Technology (IJCET)
Volume 7, Issue 6, November–December 2016, pp. 50–54, Article ID: IJCET_07_06_006
Available online at
http://www.iaeme.com/ijcet/issues.asp?JType=IJCET&VType=7&IType=6
Journal Impact Factor (2016): 9.3590(Calculated by GISI) www.jifactor.com
ISSN Print: 0976-6367 and ISSN Online: 0976–6375

DESIGN OF TRANSACTIONAL PREDICTION USING PLAN MINE AND GENETIC ALGORITHMS

R.P.S. Manikandan
Assistant Professor, Department of IT, Sri Shakthi Institute of Engineering & Technology, Coimbatore, India

Dr. A.M. Kalpana
Professor, Head/Department of CSE, Government College of Engineering, Salem, India

ABSTRACT

In this paper, we represent Plan Mine Algorithm to discover the prediction problem in the context of plan failure. The Existing remedies for the problem create failures, by filtering out the frequent patterns but it also having uninteresting events. Hence by removing the irrelevant and uninterested data by means of Plan Mine techniques, the remaining data becomes more dominant and crisp in the data set. The Frequent Transactions in the online shopping makes the business more effectively and successively.

Key words: Genetic Algorithm, PlanMine, Pruning, Frequent Pattern, Mutation.

Cite this Article: R.P.S. Manikandan and Dr. A.M. Kalpana, Design of Transactional Prediction using Plan Mine and Genetic Algorithms. *International Journal of Computer Engineering and Technology,* 7(6), 2016, pp. 50–54.
http://www.iaeme.com/ijcet/issues.asp?JType=IJCET&VType=7&IType=6

1. INTRODUCTION

In Plan Mine, we introduce Genetic Algorithms. In order for the initial classification of sequence rules, the techniques called production system introduced. Then Pruning is applied in the following phases.

The Dataset can be classified into relevant and interesting data called as good data set and relevant but uninteresting data; we called it as bad dataset. So in order to improvise the business, the good events are to be collected and need to be stored and at the same time the bad events are to be retained. Though the good events data sets are not used for prediction, they act as a reference in the subsequent pruning stages.

The second pruning phase removes patterns which correspond to the sequence that provides support in the dataset of worst plan. Moreover, it provides very good level of support in the dataset of good plans. The Third pruning phase removes the redundant patterns. A sequence is said to be redundant if it contains a sub sequence having the same support value as itself for both data sets. At last the same kind of patterns was removed, if it contains any sub sequence that kind of patterns were captured.

2. FREQUENT PATTERN PREDICTION

The discovery of frequent patterns is irrelevant for many applications such as in error discovery. Because the error occurred are rare events, the statistical support of such a sequence is low. Hence if we restrict the search space to frequent patterns, it will be hard to find rare events. So if we increase the threshold value for the frequent patterns, we can able to distinguish the interesting patterns from any trivial sequence and possible to inspect the result also. If the events are time stamped, then it contains only categorical and numerical transaction data set. So the Problem of Predicting the failure is very much important. Hence we used Genetic type of Algorithms, in order to inspect these kinds of problems and produce the result in an efficient way.

2.1. Sequence List of Ordered Items

The frequent pattern is a sequence whose frequency exceeds some user specified threshold. A sequence is a list of ordered items.

A sequence $\beta_1 \rightarrow \beta_2 \rightarrow \rightarrow \beta_q$

Where β_i is the sequence of items.

The rationale behind frequent patterns represented in detecting precedence and casual relationships that make them statistically remarkable.

Customer	P1	P2	P3	P4	P5
Cust1	1	1	1	0	0
Cust2	0	1	1	0	1
Cust3	1	1	0	0	0
Cust1	1	1	0	0	0
Cust1	1	0	1	0	1
Cust2	1	0	0	0	0
Cust1	1	1	1	1	0
Cust1	0	1	1	0	1

From the Data set, the first customer transaction can be retrieved.

Customer	P1	P2	P3	P4	P5
Cust1	1	1	1	0	0
Cust1	1	1	0	0	0
Cust1	1	0	1	0	1
Cust1	1	1	1	1	0
Cust1	0	1	0	0	0

Sequence 1: (P1, P2, P3) → (P1, P2) → (P1, P3, P5) → (P1, P2, P3, P4) → (P2)

Sequence2: (P2, P2, P5) → (P1)

Sequence 3: (P1, P2)

Normally, a simple sequence mining represents the flow of transaction.

2.2. Drawbacks in the Existing System

The spade exploits the lattice structure of the set of all elements to decompose the problem into sub problems [1]. When we consider the list of sequence in the set of items, the lattice structure is not preserved. The main drawback is that, there is no unique k-sequence resulting from the migration of two sequences.

The hyper-lattice is to be unbounded at the top. The product A can generate an infinite sequence. But when the transactions are of finite length, and the set of transaction in a sequence is bounded from the above, the hyper lattice structure is also to be bounded. It always starts with the least element of the lattice and gradually works in a bottom up way to generate all frequent sequences. It maintains the vertical as well as horizontal structure as it proceeds from the least element to the maximum elements.

3. OUR TECHNIQUES

In order to optimize the memory requirement and efficiency, sequential pattern discovery using class decomposes the original lattice what we can called as original search space into smaller sub spaces. This process can be done independently in the primary memory.

The key idea of decomposition is a equivalence relation, in which the list of sequence are distinct and they share the same prefix. It is quite convenient to see that, for a fixed value of length, say k, en equivalence relation can be defined as W_k. This relation induces a partition on the set of sequences and each equivalent class is to be treated as a sub lattice in this structure.

First of all, we can solve the sequence mining problems and send it to the other techniques for the execution. In order to solve {A}, we may have to solve {AB}, {A→B}, etc. and at the same time, should not go more than three level deep search in the recursion process. Because there may be the possibility of Performance overhead problem, if repeated db scan performs.[2]

In most cases, the size of the sub program rapidly decreases and it becomes very small, so that it can able to handle in the main memory. Always it begins from the bottom elements of this lattice and move in a upward direction either in Breadth first or Depth first manner.

This bottom up movement is carried out by incrementing k by 1 at every step and at every iteration pruning takes place. It appears to be an efficient method for frequent pattern mining. But when we take huge elements in the dataset while online shopping cart, the result is not to be accurate as determined because we expect only good and interesting relevant data rather than uninterested failure data. So in order to overcome the drawback, the frequency patterns can be identified using GA.

4. GA METHODOLOGY

GA executes in a package of separate elements or items what we can term as population. The Genetic algorithm maps strings of numbers to each potential solution. Each and every solution becomes an individual in the population and each string becomes a representation of an individual.

Genetic Methodology works in a cyclic model:

In the cycle, the first one represents the population.

That is Creation and Evolution of Population of strings

Apart from the First cycle, in a consecutive next stage, the individuals were analysed, and the corresponding descriptions are given, then its performance is to be evaluated which is correlated to the same kind of population.

4.1. Mutation

While invoking the algorithm, a new block of useful data have been formed in population. Since the modification is totally random, it creates different structures. We can able to pick the appropriate one from the list of different structures based on complexity.

5. CONCLUSION

While handling the larger data sets, in order to overcome the drawback of performance bottleneck, we proposed the concepts using genetic algorithm by optimization technique. Hence it can be used for online shopping with accuracy of handling big data and overall performance gets improved.

REFERENCE

[1] Risto Vaarandi, "A Data Clustering Algorithm for Mining Patterns From Event Logs," in Proceedings of the 2003 IEEE Workshop on IP Operations and Management, pp. 119-126

[2] Risto Vaarandi, "A Breadth-First Algorithm for Mining Frequent Patterns from Event Logs," in Proceedings of the 2004 IFIP International Conference on Intelligence in Communication Systems, LNCS Vol. 3283, Springer, pp. 293-308.

[3] Wei Xu, Ling Huang, Armando Fox, David Patterson and Michael Jordan, "Mining Console Logs for Large-Scale System Problem Detection," in Proceedings of the 3rd Workshop on Tackling Computer Systems Problems with Machine Learning Techniques, 2008.

[4] Adetokunbo Makanju, A. Nur Zincir-Heywood and Evangelos E. Milios, "Clustering Event Logs using Iterative Partitioning," in Proceedings of the 2009 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1255-1264.

[5] James J. Treinen and Ramakrishna Thurimella, "A Framework for the Application of Association Rule Mining in Large Intrusion Detection Infrastructures," in Proceedings of the 2006 Symposium on Recent Advances in Intrusion Detection, LNCS Vol. 4219, Springer, pp. 1-18.

[6] Chris Clifton and Gary Gengo, "Developing Custom Intrusion Detection Filters Using Data Mining," in Proceedings of the 2000 IEEE Military Communications Conference, pp. 440-443.

[7] Jon Stearley, "Towards Informatic Analysis of Syslogs," in Proceedings of the 2004 IEEE International Conference on Cluster Computing, pp. 309–318.

[8] Adetokunbo Makanju, Stephen Brooks, A. Nur Zincir-Heywood and Evangelos E. Milios, "LogView: Visualizing Event Log Clusters," in Proceedings of the 6th Annual Conference on Privacy, Security and Trust, pp. 99-108, 2008.

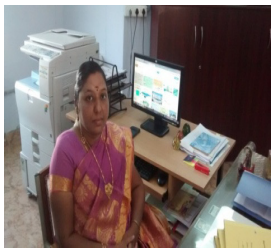
[9] Vijay Kumar, Jagdev Singh, Yaduvir Singh and Sanjay Sood . Optimal Economic Load Dispatch Using Fuzzy Logic & Genetic Algorithms . *International Journal of Computer Engineering and Technology (IJCET)*, 7 (1), 201 6, pp. 62-77.

[10] Mr.Vijay Kumar, Dr.Jagdev Singh, Dr.Yaduvir Singh, Dr.Sanjay Sood. Design & Development of Genetic Algorithms for Economic Load Dispatch of Thermal Generating Units. *International Journal of Computer Engineering and Technology (IJCET)*, 3 (1), 2012, pp. 59-75.

[11] S.V.S. Ganga Devi. Financial Forecasting Using Genetic Algorithms. *International Journal of Computer Engineering and Technology (IJCET),* 4 (6), 2013, pp. 423-435.

ABOUT THE AUTHOR

R.P.S. Manikandan currently working as a Assistant Professor in Sri Shakthi institute of Engineering and Technology, Coimbatore. He is having almost 5 years of experience in teaching field. He Completed Master Degree in Computer science and Engineering from Sasurie College of Engineering, Vijayamangalam, Tirupur Dist, Tamilnadu in 2012 and Bachelor of Engineering in Computer science and Engineering from Velalar College of Engineering and Technology, Thindal, Erode Dist, Tamilnadu in 2010 and secured First class with Distinction grade in both the graduation. He is interested to learn new Software's and to innovate technological ideas in the field of Computer science and Engineering. He is a member in International Association of Engineers.



Dr. A.M. Kalpana is currently the heading the department of CSE in Government College of Engineering, Salem, Tamilnadu, India. She received her doctorate from Anna University, Chennai. She is a member of IEEE, ISTE and CSI. She has published more than 15 papers in International Journals and 100 papers in national and international conferences. Her research interests include Software Engineering, Software testing and data mining.