International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 12 (2017) pp. 3483-3486

Efficient Approach for Mining Consistent Profit Based Pattern over Large Dynamic Data Stream

Vijay Kumar Verma[1] and Kanak Saxena[2]

[1]Computer Science & Engineering Department Lord Krishna College of Technology, Indore, Madhya Pradesh, India .
[2]Computer Application Department Samrat Ashok Technological Institute, Vidisha, Madhya Pradesh, India.

[1]ORCID: 0000-0001-5442-0113

Abstract

In present scenario information systems are continuously updated in several applications. Consistent pattern or regular pattern play important role in data mining. Efficient mining of consistent pattern is crucial task. Consistent pattern are those pattern which appear in the transaction after a fixed interval. Maximum existing methods are based on generation of huge applicants and avoiding this costly stupid candidate generation is a difficult and challenging task. This paper introduce an inexpensive method which not only reducing costly candidate generation but also minimize computational cost. This paper used heuristic information based approach to mine consistent pattern with high profit.

Keywords: Consistent Pattern, Interval, Inexpensive, Computation, profit

INTRODUCTION

Database technology has been rapidly changed day to day to facilitate the storage and usage of massive data which are generated from governments, industry, and business organizations and scientific and research organizations. To identify frequent pattern form database is measured with the support of frequency of an itemset. Maximum algorithm applies the traditional approach for pattern generation which is based up words closure property. There is another approach based on downward closure property.

Downward closure property provides the powerful discarding policy. In this process once identified a useless candidates then there is no need to check its permutations of the item set.

BACKGROUND

Let Y contain items $x_1, x_2, x_3 \ldots x_n$ denoted as $Y=\{ x_1, x_2, x_3, \ldots, x_n \}$.

Daily purchasing records are denoted by

$PR= \{x_j, , \ldots , x_m\}$ is a subset of Y,

Daily purchasing records are composed into a database which denoted by D. Every item in Y has a profit value in the profit table. Each purchasing records is identify by a unique number (PR) and is a subset of Y.

a. Consistent pattern

A pattern *is* said to be consistent pattern its period is less than or equal to user given threshold.

Let PR^Y_j *and* PR^Y_{j+1} are two consecutive purchasing records. The difference between these two successive transactions can be defined as a difference

$$PR^X(i.e., PR^X= PR^X_{j+1} - PR^X_j).$$

b. Local profit

The local profit of an item x_i in Y denoted $lp(x_i, Y)$, and calculated by

$$lp(x_i, Y)= \sum_{k=0}^{n} pf(xi, Y)$$

c. *Profit of an item set*

The profit of an itemset, denoted *pf* is the sum of the local profit of every item in Yin all records containing item set and calculated by

$$pf(Y) = \sum_{k=0}^{n} pf(xi, Y)$$

d. Purchasing record based profit

The profit of purchasing record denoted as *pf(PR)* is the sum of the utilities of all the items in Y and calculated by

$$pf(PR) = \sum_{k=1}^{n} (i, PRi)$$

REVIEW OF RELATED WORKS.

In 2004Yao, Howard and proposed theoretical analysis of profit mining. They two types of profit for items, record based profit and outside profit [4].

In 2005 Liu etc. proposed algorithm which contain 2 parts. In the first part they calculate record based profit and in second part they identify the high profit item [5].

In 2007 Erwin etc. proposed a new algorithm called which is based on a tree structure and used the compressed method. They tested proposed algorithm on several sparse and dense data sets [6].

In 2010 Syed etc. proposed a novel tree structure called (RPS-tree). RPS-tree captures the stream content, and with an efficient tree updating mechanism it constantly processes exact stream data. Extensive experimental analyses show that our RPS-tree is highly efficient in discovering *r* patterns from a high-speed data stream [7].

In 2011Vijay etc. proposed VDRP method to mine complete set of consistent patterns in incremental databases for a user given threshold. They also introduce IncVDRP-table for updated database [9].

In 2012 Pavan Kumar etc. proposed new approach to find the complete set of consistent patterns. The proposed method is very simple to use with simple operations like arrays, unions, intersection, deletion etc[10].

In 2013 Vijay Kumar etc. They proposed a method mining regular Pattern using Bit Stream Sequence with sliding window to generate the find consistent pattern over a data stream at a user given threshold. The proposed bit-stream sequence representation of items to enhance the performance of the proposed method over the previous methods[11].

In 2014 Philippe etc. proposed proposing a novel strategy based on co-occurrences to reduce the number of join operations. A proposed FHM (Fast High-Utility Miner) reduces the number of join operations by up to 95 % and is up to six times faster than the HUI-Miner[12].

THE PROPOSED METHODS

We proposed an efficient approach (name) to find consistent pattern with high profit using an efficient removing (pruning) strategies. A pattern is consistent when it occurs at fixed intervals in a database at a user given consistent threshold.

Let $Y= \{x_1, x_2, \ldots, x_n\}$ be a set of items. A set $X = \{x_j. \ldots .x_k\}$subset of Y, where j ≤ k and j, belongs[1, n].

A database D contains Y is a collection all purchasing records containing items $DPR= \{PR_1, \ldots, PR_m\}$ size of DPR, is defined the total number of purchasing records. Let the first transaction records are denoted PR_f and the last transaction records is denoted by PR_l. The difference between two records can be defined as a period of *X*, say $DPR^x= PR^x_{j+1}– PR^x_j$, where $f<=j<=l$.

For large number of transaction we consider the in the *DPR* as $TR_f= 1$ and the last records is the $TR_l=m$.

Now the set of difference can be represented as. $DPR^x=\{DPR_1^X, \ldots, DPR_s^x\}$, where *s*is the total number of difference *X*.

Consider a simple purchasing record and corresponding profit table

Decreasing search space

In the proposed our objective is to reduce search space. So in the first we start with the coupling of purchasing records by using a unique id. Coupling of purchasing records can represented as

$CPR= \{(PR_1,PR_2,).(PR_1,PR_3)\ldots(PR_n,PR_{n-1})\}$

Table 1: Purchasing records

S.No.	x_1	x_2	x_3	x_4	x_5	x_6
PR_1	1	0	5	7	4	0
PR_2	2	0	5	6	8	7
PR_3	4	0	8	0	5	0
PR_4	0	0	3	6	7	0
PR_5	3	0	5	0	4	
PR_6	7	0	0	0	0	4
PR_7	0	3	6	7	6	0
PR_8	0	4	5	6	7	0
PR_9	6	4	7	5	0	0

Table 2: Profit table

Item	Profit
x_1	3
x_2	2
x_3	10
x_4	8
x_5	5
x_6	2

During coupling we also calculate the number of items. After coupling we arrange purchasing record on the basis of number of items from max to min. Some item appear commonly in the records for those items we merge records and remove duplications. consider an example suppose $\{x_1, x_3, x_5\}$ are three item appear in

$\{(PR_1,PR_3),(PR_1,PR_5),(PR_2,PR_3),(PR_3,PR_5)\}$.

We takes common and removes duplicates

$\{(PR_1, PR2,PR3,PR_5\}$

It will helps us for finding frequent pattern and consistent pattern in. consistent pattern are calculated on the basis of are calculated by subtracting two successive transaction by using the formula

Eliminate unpromising applicant

In the proposed approach we eliminate unpromising applicant very efficiently. For example consider a set of item X containing n items $X=\{x_1, x_2, x_3, x_4\ldots\ldots\ldots x_n\}$. For n item we need 2^ncoupling. Suppose there is k unpromising applicant.

We have to check $(2)^{(n-k)} - 1$ subsets.

By the proposed approach we need create $n(n-1)/2$ applicant.

THE PROPOSED ALGORITHM

Input: Let $D= \{PR_1, PR2\ldots\ldots\ldots\ldots\ldots PR_n\}$ be set of all purchasing records.

Output: Set of consistent pattern with max profit

Procedure:

1. Let $PR_i \subseteq DPR$ a k-item set
2. Calculate profit of each item and records based profit
3. For each i (i= 1, 2, 3, 4…n)

   Do coupling with PR_{i+1}until no candidates generation possible.
4. For each pair counts items and arrange them in using number of item and delete duplicates PR_i .
5. Find the difference of

   $DPR_i = PX_{i+1} - PX_i$
6. Find max value in set of difference $(DPR) = max\ (DPR_i)$
7. If DPR<=Consistent threshold and profit >= Profit threshold then Items are consistent with max profit
8. Else
9. Delete x_i
10. Repeat

EXPERIMENTAL ANALYSIS.

We implements propose approach using visual studio Dot net 2013 as front and SQL server R2 2008 as data base storage net our proposed method have. We used different parameter for experimental analysis like number of records and execution time.

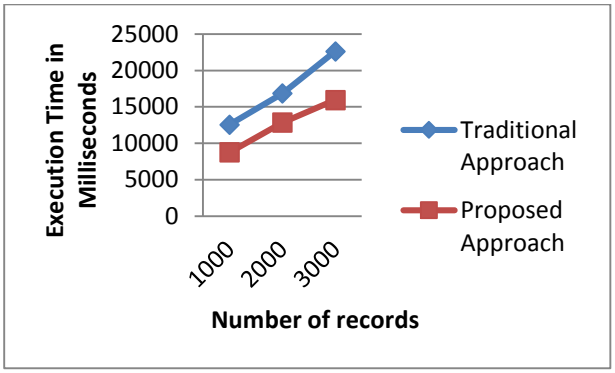

Figure 1: Number of consistent patterns with high utility

REFERENCES

[1] Han camber Data Mining: Concepts and Techniques, 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.

[1] Arun Pujari Data Mining Techniques Universities Press, 2001.

[2] Data Mining and Analysis: Fundamental Concepts and Algorithms, by Mohammed Zaki and Wagner MeiraJr, to be published by Cambridge University Press in 2014.

[3] Hong Yao, Howard J. Hamilton, and Cory J. Butz "A Foundational Approach to Mining Itemset Utilities from Databases" University of Regina , SK, Canada

[4] Ying Liu "A Fast High Utility Itemsets Mining Algorithm "UBDM '05, August 21, 2005, Chicago Copyright 2005.

[5] Alva Erwin "A Bottom-Up Projection Based Algorithm for Mining High Utility Itemsets" Copyright 2007, Australian Computer Society, Inc. 2nd Workshop on Integrating AI and DataMining (AIDM 2007).

[6] Syed Khairuzzaman "Mining Regular Patterns in Data Streams "H. Kitagawa et DASFAA 2010, Part I, LNCS 5981, 399413, and 2010. Springer-Verlag Berlin Heidelberg 2010.

[7] Yi-Cheng Chen "An Efficient Algorithm for Mining Time Interval-based Patterns in Large Databases" CIKM'10, October 2630, 2010, Toronto, Ontario, Canada. Copyright 2010 ACM 978-1-4503-0099-5/10/10.

[8] G. Vijay Kumar, M. Sreedevi, NVS Pavan Kumar "A Vertical Format to Mine Regular Patterns inIncremental Transactional Databases" Journal of

computing, Volume 3, issue 11, November 2011, ISSN 2151-9617 .

[9] NVS Pavan Kumar etc. "Mining Regular Patterns in Data Streams Using Vertical Format" International Journal of Computer Science and Security (IJCSS), Volume (6) : Issue (2) : 2012.

[10] Vijay Kumar Verma, KanakSaxena" Mining Regular Pattern Over Dynamic Data Stream Using Bit Stream Sequence" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-3, Issue-7, December 2013

[11] Philippe Fournier-Viger "FHM: Faster High-Utility Itemset Mining usingEstimated Utility Co-occurrence Pruning" Viger, P., Wu, C.-W., Zida, S., Tseng, V. S. (2014) FHM: Faster High-Utility Itemset Mining using Estimated Utility Co-occurrence Pruning. Proc. 21st International Symposium on Methodologies for Intelligent Systems (ISMIS 2014), Springer, LNAI, pp. 83-92.