© 2017 IJSRSET | Volume 3 | Issue 8 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Mining of Frequent Maximal Itemsets Using Neural Network

Gagan Madaan[1] ,Chahat Monga[2]

[1]Assistant Professor, Department of Computer Science & Application S.U.S. Panjab University Constituent College, Guru Harsahai, Punjab, India

[2]Assistant Professor, Department of Computer Science & Application, Guru Nanak College, Ferozepur, Punjab, India

ABSTRACT

The task of discovering itemsets in databases was introduced in 1993 by Agrawal and Srikant3 as large itemset mining, but it is nowadays called frequent itemset mining (FIM). This paper proposes the Mining of Frequent maximal Itemsets using the technique of Optical Neural Networks. Since optical neural network performs many optical computations in nanoseconds, the time complexity is very low as compared to other data mining techniques. The data is stored in such a way that it minimizes space complexity to a large extent as database is scanned only once and stored in the form of weight matrix as in neural networks. The frequent patterns are then mined from this weight matrix using optical inputs. This approach discovers the frequent patterns quickly and effectively mines the potential association rules. It discovers frequent patterns by using the best features of data mining, optics and neural networks. This paper focuses on how this model can be helpful in generating frequent patterns for various applications.

Keywords: Frequent Itemsets, Patterns, Maximal Itensets, Data Mining, Association Rules, Optical Neural Network.

I. INTRODUCTION

Data mining is the auspicious field of database due to its wide and significant applications in industry and is a key step in the knowledge discovery process in large databases. It consists of applying data analysis and discovery algorithms that under limitation of acceptable computational efficiency, produce a particular enumeration of patterns over the data.Due to massive amount of data generated from business transactions, there arose a need for

an efficient techniques to discover new interesting patterns in less time from these large databases in order to derive knowledge for quick and effective decision making. One of the problems of data mining is to discover association rules from the database from which we need to determine frequent itemsets. The problems of finding these frequent itemsets are fundamental in data mining, and from the applications, fast implementations for solving the problems are needed. Many algorithms have been implemented for finding frequent patterns for data mining. In large databases the problem of mining frequent patterns gets multifold, since the database needs to be scanned several times.

One of the important development in area of association rule mining was development of Apriori algorithm but candidate key generation remained the unsolved issue in that too.. It was improved by partition and sampling , but both of these approaches were inefficient when the database was dense. The use of optical neural network for mining frequent patterns with only a single database scan seems to be the most optimized technique. The parallel computation of frequent patterns makes mining faster. Traditional association rule algorithms adopt an

IJSRSET1841123 | Received : 01 Nov 2017 | Accepted : 20 Nov 2017 | November-December-2017 [(3)8: 1100-1104]

iterative method to discovery, which requires very large calculations and a complicated transaction process[7]. This approach discovers the frequent itemsets by making use of the best features of optics and neural networks. It makes use of maximal pattern mining to save computational time and reduce the number of candidate generation. The maximal patterns are determined and used for further mining their frequent subsets.

II. PROPOSED MODEL

The model suggests for mining frequent maximal patterns using an optical neural network.

2.1 Artificial Neural Networks : Artificial neural networks are inspired by the operation of the human brain. It is a model of the biological neuron as a circuit component to perform computational tasks.

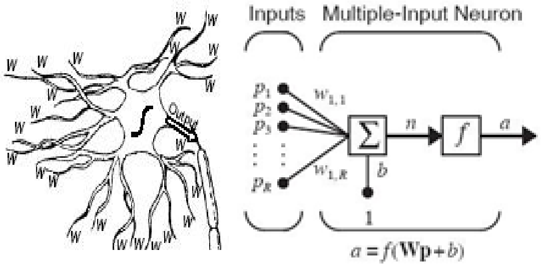


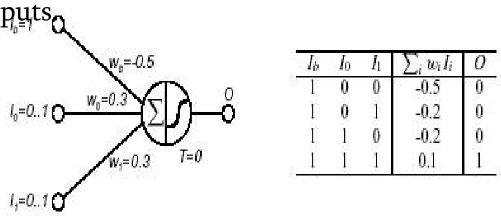
Figure 1. Human nerve cell and Artificial Neuron Mode

The function of a neuron can be described in mathematical form with:

$$O = \mathrm{F}\ (\ \square\ \mathrm{wi}\ .\ \mathrm{Ii}\ )$$

i

where $o$is the output signal of the neuron and *Ii* are the input signals to the neuron, weighted with a factor *wi* . F is some nonlinear function representing the threshold operation on the weighted sum of inputs.



I_b	I_0	I_1	$\sum_i w_i I_i$	O
1	0	0	-0.5	0
1	0	1	-0.2	0
1	1	0	-0.2	0
1	1	1	0.1	1

Figure 2. A neural implementation of a logical AND function and the corresponding truth table including weighted sum of inputs of the neuron.

2.2 Optical Neural Networks

In Optical neural networks the neurons are interconnected with light beams. No insulation is required between signal paths. The light rays pass through between each other without interlacing. The density of transmission path is limited only by the spacing of light sources, the effect of divergence and the spacing of detectors. As a result all signal paths operate simultaneously, which results in a true data rate. The strengths of weights are stored in holograms with high density. These weights can be modified during operation to produce a fully adaptive system. Weighted addition and transformation are the main operations of a neuron and since optoelectronic devices are most effective for realization of vector-matrix multiplication, the main calculation load can be put on them .The proposed model uses electro-optical matrix multipliers where optics is used for its massive parallelism and the input and output data are defined in the electronic domain.

2.3 Electro-optical Matrix Multipliers

Electro-optical Matrix Multipliers provide a means for performing matrix multiplication in parallel. The network speed is limited only by the available electro-optical components. The computational time is potentially in the Nanosecond range[8]. The speed is independent of the size of array. This makes the network to be scaled up without increasing the time required for computation. Theses nets provide a means for performing matrix multiplication in parallel. The network speed is limited only by the available electro-optical components. The computational time is potentially in the Nanosecond range[8]. Figure2. shows the electro-optical vector matrix multiplier. The system is capable of multiplying a 5-element input vector by a 5 * 6 matrix, which produces 6-element NET vector. The column of light sources passes its rays through a lens, such that each light illuminates a single row of weight shield. The weight shield is a photographic film in which the transmittance of each square is proportional to the weight. There is another lens which focuses the light from each column of the shield to a corresponding photo detector. The NET is calculated by, NET = w_{ik} x_i where NET_k - NET, w_{ik} - weight from neuron i to neuron k, x_i - input vector component I. The output of each photo detector will represent the dot product between the input vector and the weight matrix. The set of outputs is a vector equal to the product of the input vector with weight matrix. Hence matrix multiplication is done in parallel. The speed is independent of the size of array. This makes the network to be scaled up without increasing the time required for computation. For weights, instead of photographic film, liquid crystal light valve may be used.

III. MINING MAXIMAL FREQUENT ITEMSETS

In the suggested model, each transaction is represented by rows of the weight matrix and the presence and absence of any item is stored as weights 1 and 0 respectively. A sample database D and its corresponding weight matrix W is given.

Table 1. Sample DataBase (D)

Rid	**Items**
1.	**A C D F**
2.	**B C E F**
3.	**A B C E F**
4.	**B E F**
5.	**A B C E F**

Table 2. Weight Matrix (W)

A	B	C	D	E	F
1	0	1	1	0	1
0	1	1	0	1	1
1	1	1	0	1	1
0	1	0	0	1	1
1	1	1	0	1	1

Electro-optical implementation of the suggested model uses an array of light emitting diodes (LEDs) to represent the logic or neurons of the network. The array of LEDs represent the input vector, and an array of photodiodes (PD) is used to detect the output vector. Multiplication of the input vector by the matrix W is achieved by horizontal imaging and vertical smearing of the input vector that is displayed by the LEDs on the plane of the mask W. The output obtained is the net input to the neuron ie NET = w_{ik} x_i. It is then thresholded and compared with the minimum support required for an item to be frequent. If the NET>=min-sup, the item is frequent. In this way a single weight matrix can find all frequent-1 item sets in the database.

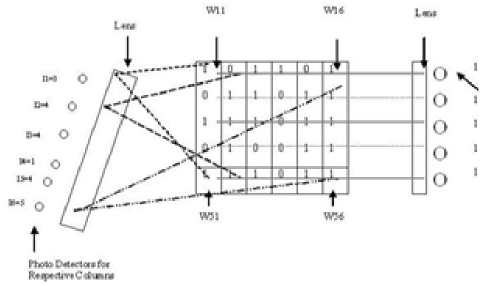


Figure 2. Model mining all 1-items

The proposed model goes through the following phases:

3.1. Transform database to weight matrix: The mined transaction database is D, with D having m transactions and n items. Let T={T1,T2,…,Tm} be the set of transactions and I={I1,I2,…,In}be the set of items. We set up a weight matrix Wm*n, which has m rows and n columns. Scanning the transaction database D, if item Ij is in transaction Ti , where 1=j=n,1=i=m, the element value of Wij is '1' otherwise the value of Wij is '0'.

3.2. Generate the set of frequent 1-itemsets :The weight matrix Wm*n is scanned and weight masks are set to 1 or 0. Initially to mine all frequent 1-itemsets, the input vector will consist of all 1s. This is because we need the product of w and x to be either 1 or 0 to represent presence or absence of the item. The sum of each column is read and compared with the minimum support. If it is found frequent it is stored in a list of frequent 1-itemsets F1.

3.3 Prune and join using apriori property: Pruning means omitting infrequent items for further consideration. Only frequent items are joined to get candidate 2-itemsets. To do his, the corresponding values of the columns of frequent items in weight matrix are multiplied. Here 1*1 will only give the output as 1 which shows that both transactions contain that item otherwise a 0 in any one or both transactions will give a 0 indicating that the item is not present in both the transactions. Thus, we get a matrix Am*q, where q is the number of candidate 2-itemsets.

3.4 Generate the set of frequent 2-itemsets. The weight matrix Am*q is scanned and weight masks are set. The input vector consisting of all 1s is fed. The sum of each column is read and compared with the minimum support. If it satisfies the minimum threshold the 2-itemset is frequent and added to the list of frequent 2-itemsets F2.

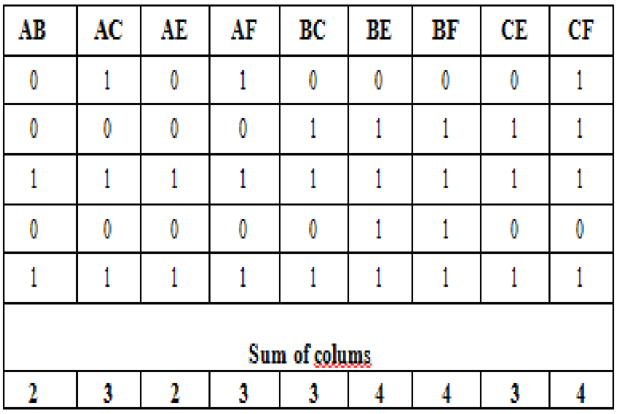
3.5. Generate candidate maximal itemsets from frequent 2-itemsets: A frequent itemset P is maximal if P is included in no other frequent itemset[9]. All 2-itemsets that satisfy the criteria for joining are joined together to form a candidate maximal itemset. For example, {BC}, {BE}, and {BF} can be joined to give a candidate maximal itemset, i.e., {BCEF}. All such itemsets are generated by joining all frequent 2-itemsets.

3.6. Mine frequent maximal itemsets from frequent 2-itemsets: In order to mine frequent maximal itemsets, the candidate maximal itemsets are fed as input to the transpose of the initial weight matrix, i.e., W^Tm*n from which the column corresponding to the infrequent item is removed. Let us call this weight matrix as M. When a candidate maximal k-itemset is fed as input vector I_m, the output k received at each photo-detector shows the presence of that itemset in the transaction. If the output is not k, the candidate maximal itemset is not present in the transaction. Now, the total number of photo-detectors giving the output k is the support count of that candidate maximal k-itemset. The input vector for which the value of k is more than the threshold, i.e., min-sup, is frequent maximal itemset.

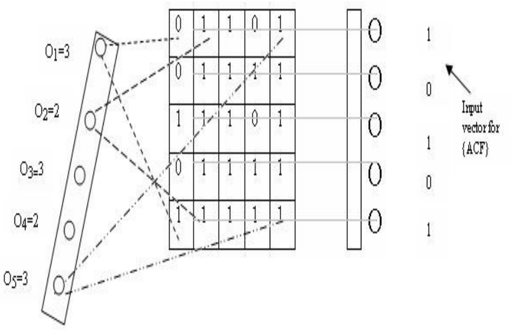
3.7. Mine all subsets of frequent maximal itemsets. The other smaller itemsets are then mined from these frequent maximal itemsets.

IV. EXAMPLE

F2 is {AC} {AF}, {BC} {BE} {BF}, and {CE} {CF}. Candidate Maximal itemsets are {ACF}, and {BCEF}.Thus we get M as follows: We feed {ACF}, i.e., 10101 as input M that contains rows corresponding to A, B, C, E, F. Since {ACF} is a 3-itemset, therefore, the expected output from each column is not more than 3. Since {ACF} is present in the first column indicating the first transaction, it gives the output O_1 =3 for the first column. Since, the value for O_2 and O_5 is also 3, the number of transactions containing 3 are 3 which is equal to min-sup. Therefore, the maximal itemset{ACF} is a frequent maximal itemset.The same procedure can be repeated for other candidate maximal itemsets. The other smaller itemsets can then be mined from these frequent maximal itemsets.



AB	AC	AE	AF	BC	BE	BF	CE	CF
0	1	0	1	0	0	0	0	1
0	0	0	0	1	1	1	1	1
1	1	1	1	1	1	1	1	1
0	0	0	0	0	1	1	0	0
1	1	1	1	1	1	1	1	1
Sum of colums								
2	3	2	3	3	4	4	3	4



0	1	1	0	1
0	1	1	1	1
1	1	1	0	1
0	1	1	1	1
1	1	1	1	1

Figure 3. Electro-optical Vector Matrix Multiplier implementing a 5 by 5 matrix and input vector for the candidate itemset {ACF}.

V. OTHER APPLICATION AREA OF THIS MODEL

Optical computing is suitable to many problems due to its Fan-in efficiency, efficiency in interconnection complexity, and energy efficiency.The model presented in this paper finds scope in many areas like incremental data mining, classification, prediction, maximal approach, distributed approach, online data stream mining clustering etc.

VI. CONCLUSION

In this paper, maximal frequent itemsets are mined using an optical neural network model. It stores all transaction data in bits, so it needs less memory space and can be applied to mining large databases. This model accesses the database only once to store the transaction–ids in the Electro-optical MVM and all supports are determined in parallel thus making it much faster than the other available techniques. This model can further be improved by replacing the electronic threshold by optical threshold. Optical thresholding maintains the spatial optical parallelism and avoids opto-electronic inter-conversions . The 2-itemsets and higher can also be mined using the same model that will futher reduce computation time. The model finds future scope in incremental data mining and online data stream mining.

VII. REFERENCES

[1]. Han, J., Kamber, M., "Data mining: Concepts and Techniques" Morgan Kaufmann Publishers, San Francisco (2001)

[2]. Fayyad U.,Piatetsky-Shapiro G.,Smyth P.,and Uthrusamy R 9Eds.). "Advances in Knowledge Discovery and Data Mining" AAAI press, Menlo Park, CA, ISBN:0-262-56097-6, pages:611,1996.

[3]. Takeaki Uno, Masashi Kiyomi, Hiroki Arimura: LCM ver 2. Efficient mining algorithms for Frequent/ closed/ maximal itemsets.

[4]. Agrawal R. and Srikant R. : Fast algorithms for mining association rules in large databases. In Proc. 20th VLDB, pages 478-499, Sept. 1994.

[5]. Sarasere A.,Omiecinsky E., and Navathe S. : An efficient algorithm for mining association rules in large databases. In Proc. 21st VLDB, pages 432-444, Sept. 1995

[6]. Toivonen H. L sampling large databases for association mining rules . In Proc. 22nd VLDB, pages 134-145, Sept. 1996

[7]. An Optical Neural Network Technique for Mining Frequent Maximal Itemsets in Large Databases Divya Bhatnagar1 and A. S. Saxena2 2011.

[8]. Hanbing Liu and Baisheng Wang: An association rule mining algorithm based on a Boolean matrix. Data Science Journal, Volume 6, Supplement, 9 September 2007.

[9]. Shivanandam, S. N., Sumathi, S., Deepa, S. N.: Introduction to Neural Network using MATLAB 6.0. TATA Mc.Graw Hill.

[10]. Takeaki Uno, Masashi Kiyomi, Hiroki Arimura: LCM ver 2. Efficient mining algorithms for Frequent/ closed maximal itemsets.

[11]. I. Saxena, P. Moerland, E. Fiesler, A.R. Pourzand, and N Collings: An optical Thresholding Perceptron with Soft Optical Threshold.

[12]. Tenanbaum M., Augenstein M.J., Langsam Y. "Data Structures using C and C++" Prentice-Hall India, Edition

[13]. Pujari A.K. "Data Mining:Technique" Universities Press

[14]. Mos, Evert C. "Optical Neural Network based on Laser Diode Longitudinal Mode