A Novel Approach for Association Mining Based on Matrix Factorization and Deep Neural Network

Harvendra Kumar [1], Rakesh Kumar Yadav [2]

IFTM University Moradabad, India.
[1] harvendra.patel82@gmail.com, [2] rkyiftmuniversity@yahoo.com

***Abstract* - Association Rule Mining (ARM) is used for distinguishing proof of relationship between a substantial arrangement of information objects. Because of the vast quantity of information held in databases, businesses are concerned about unauthorized extraction of mining affiliation rules from their databases. This paper present a viable instance of applying market basket analysis on a certifiable deals exchange informational collection utilizing time arrangement grouping, as opposed to utilizing customary affiliation manage mining. Association Rule Mining is used to discover interesting patterns from a large database. Because of the vast quantity of information held in databases, it is difficult to extract useful information. Our proposed grouping process finds numerous arrangements of reciprocal parts, where each arrangement of parts are utilized to make a similar item. Such data is helpful for strategically pitching and estimation. We present a Deep Neural Network based approach for Association Rule Mining (ARM). The analysis of the proposed framework suggests that the usage of Deep Neural Networks with matrix factorization will help in mining association rules that are normally invisible.**

Keywords - Association Mining, Deep Neural Network, Deep Learning, Autoencoder.

I. INTRODUCTION

Data Mining is known as fetching information from enormous group of data [1]. We can say that data mining is the procedure of obtain knowledge from data. The data fetched so can be worn for many of the operations: Analysis of market, Fraud Detection, Customer Retention, Production Control Science Exploration. Over decades retail chains and retail chains have been offering their items without utilizing the value-based information produced by their deals as a wellspring of learning. As of late – over the most recent two decades – organizations began to utilize this information to find data. In the 90's constrained computational capacities made the extraction of learning from a huge number of day by day exchanges unfeasible and just examination with straightforward models and diminished datasets were conceivable. In 1993, (Agrawal, Imielinski, and Swami, 1993 [2]; Agrawal and Srikant, 1994) [3] demonstrated that numerous associations were getting greater databases with value-based information, buyer information, deals records, and so forth. In this manner, they proposed the Apriori calculation (Agrawal and Srikant, 1994) for a vast informational collection for those years.

Association rule mining (ARM) is utilized for distinguishing proof of relationship between vast arrangements of information things. Because of expansive amount of information put away in databases, a few enterprises are getting to be worried in mining affiliation rules from their databases. For instance, the location of fascinating affiliation connections between vast amounts of business exchange information can aid inventory configuration, cross-advertising and different business basic leadership forms. A run of the mill case of affiliation control mining is advertising container examination. This technique looks at client purchasing behaviors by distinguishing relationship among different things that clients put in their shopping crates. The recognizable proof of such affiliations can assist retailers with expanding showcasing systems by picking up knowledge into which things are often obtained mutually by clients. This work goes about as an expansive region for the scientists to build up a superior information mining calculation. This paper displays an overview about the current information digging calculation for market basket investigation.

In this paper, we present a practical case of applying Market Basket Analysis on a real-world sales transaction data set using matrix factorization and deep neural network, rather than using traditional association rule mining. We find that auto encoder neural network can help us to reduce the dimensions on the big data and maps the high dimensional data to lower dimension. Then we applied the matrix factorization to identify the interesting patterns. The result on industrial data shows that the proposed approach can identify the undiscovered interesting patterns.

The rest of this paper is organized as follow: section II 2 surveys related work, section III examines some preliminaries concepts, section IV introduces the proposed approach and in section V illustrates the result, finally section VI concludes paper with a discussion of the ramifications of our work on future research directions.

II. LITERATURE REVIEW

Data mining play an important role to analyzed customer behavior. Searching the hidden knowledge from the data warehouse is the process of DM [4]. Data mining (DM) is the term which is used to describe the process of extracting values from large a database. Typically this involves finding patterns, trends, associations, relationships, dependencies and so on.

In past few years scientist and scholars have proposed various techniques for association rule mining. The very first algorithm for association rules miming was AIS (Agrawal, Imielinski, and Swami) [1993].One of the main drawback of this algorithm is that it required too many over whole database which makes it slow need more space. In 1994 Apriori algorithm was proposed by (Agrawal and Ramakrishan, 1994).This algorithm use breadth first search strategy and known as one of the best algorithm for association rules mining. Another popular approach known as FPGrowth (Han et al., 2004) [5] was proposed at the same time. The basic idea behind as FPGrowth is to first compress the database using FP-tree structure and then it's adapts the divide-and-conquer technique for decompression for rules mining.

In 1997 Zaki proposed four new algorithms [6]. These algorithms were based clustering and lattice traversal scheme. At the same time the database was growing exponentially due to fast web development. The tradition technique were unable to extract all the rules from large databases in real-time. In order to cope with these problem different approaches were proposed recently. These algorithms try to reduce the number of passes over the database, utilize the parallelism and use some advanced concepts.

Many scientists see association rule mining as an optimization algorithm. Metaheuristic algorithms provide a sufficiently good solution to an optimization problem. Keeping this in mind scientist has used Metaheuristic algorithms for association rule mining. GENAR and GAR are the two very popular algorithm based on genetic algorithm proposed by Mata et al. (2001, 2002) [7, 8].These algorithm have some shortcoming like inefficient representation of the individual. To overcome these limitations many algorithms based on generation algorithm were proposed that can represents the solutions in efficient manners. One of them is ARMGA was developed by Yan and Zhang (2005) [9] which is very effective for global searching. In improvement over ARMGA was proposed by Wang et al. (2009) [10], namely AGA. These techniques used the intelligent mutation and crossover operators. Some scientist also proposed hybrid algorithms. One such approach proposed by Liu (2010) [11]. It is a hybrid of GA and Simulated Annealing. Here mutation and crossover are performed with the help of Simulated Annealing. Indira and Kanmani (2012) [12] tried to analyses the performance of association rule mining based on genetic algorithm. PSO (particle swarm optimization) is another popular metaheuristic technique. PSOARM is a PSO based technique proposed by Kuo et al. (2011). Another metahurisctic based technique was proposed by Kuo and Shih (2007) [13]. This technique combines the clustering with ant colony optimization (ACO). Another approached called ACOR proposed by Parisa et al. (2011) [14]. Inspired by Newtonian gravity and the law of the motions a new algorithm called ARMBGSA proposed by Fariba et al. (2011) [15]. This algorithm model association rules as mass and as we already know all the masses attract each other's by the law of the motion. The heaviest masses are selected in order to influence the new masses.

III. PRELIMINARIES

A. Document Auto Encoder

In recent years, deep neural network (DNN) has attracted an increasing attention of many researchers in applying different types of DNN related algorithms in a wide variety of application areas, such as, such as speech processing, pattern recognition ,classification ,medical applications, etc. (Hinton, Geoffrey, et al 2012) [16]. A representative method is Auto Encoders (AE) (Hinton, Geoffrey, et al 2011) [17]. An Auto Encoder (AE) is a feed forward artificial neural network that consists of one input layer, one hidden layer and one output layer.

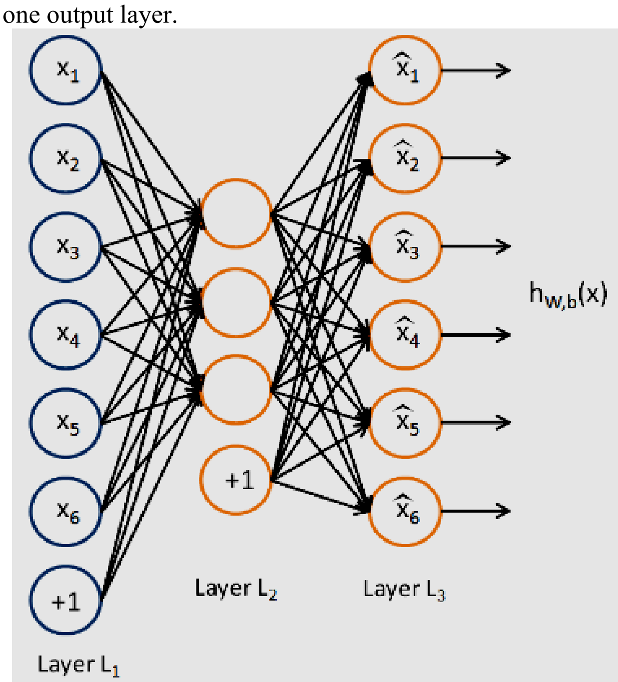

Figure 1: Autoencoder

The encoder maps the input data from a high-dimensional space into codes in a low-dimensional space, and the decoder

DOI 10.5013/IJSSST.a.20.01.25

ISSN: 1473-804x online, 1473-8031 print

reconstructs the input data from the corresponding codes. In other words, the objective is to approximate an identity function in various ways, which makes the output as close to the input as possible. An illustration of an autoencoder is displayed in Fig.1. Note that the dimensionality of the output layer is equal to that of the input layer.

The left half of the AE is called the encoder, whose input is the input of the AE and output is the output of the hidden layer of the AE. The encoder converts a given input vector into a code, which is intended to be a more efficient representation of the input vector.

Given the training example $X = \{x_1, x_1, \dots, x_m\}$, let the encoder transforms the input vector into a hidden representation example $h = \{h_1, h_1, \dots, h_m\}$ such that:

$$h = f(X) = S(WX + b)$$

Where S is a nonlinear activation function, typically a logistic sigmoid function W is a $d' \times d$ weight matrix and b is the bias unit of size *d*.

$$S(z) = \frac{1}{1 + e^{-z}}$$

The right half of the AE is called the decoder, whose input is the output of the hidden layer (c) and output ˆ x is the output of the AE.

$$y = g(h) = S(W'X + b')$$

The training of AF aims to optimize the parameter set $\theta = \{W, b\}$, $\theta' = \{W', b'\}$ minimizing the reconstruction error:

$$\theta^*, \theta'^* = argmin_{\theta,\theta'} = \frac{1}{n}\sum_{i=1}^{n} L(x^{(i)}, z^{(i)})$$

Where L is mean square error (MSE) is usually used as the standard autoencoder loss function

$$L(x, z) = \|x - z\|^2$$

After the normalization process it is expected to change the clients and things framework range to the utilized Autoencoder enactment work run. Afterward, a Denoising Autoencoder is prepared for clients highlights utilizing the standardized client lattice and another is prepared for the thing highlights utilized the standardized thing framework. Both is prepared utilizing normal backpropagation calculation and have a defilement level v to every stacked layer, sigmoid actuation work, the learning rate γ, a L2 regularization λ and squared error loss function.

B. Non-Negative Matrix Factorization

NMF is a matrix factorization algorithm with non-negative constraint. It has been investigated by many researchers, e.g. PAATERO and TAPPER (1994) [18]. However, it is popularized by the work of LEE and SEUNG (1999) [19]. Based on the point that the negativity is meaningless in human perception, they proposed a smart algorithm to find proper non-negative representations of non-negative data or images. The basic NMF problem is stated as follows: given a matrix $A_{m\times n}$ with non-negative values, and then factorize it into two matrices $W_{m\times k}$ and $H_{k\times n}$ as well as possible. The process can be described as follows:

$$A_{m\times n} \approx W_{m\times k} H_{k\times n}$$

Additionally, the reduced rank k is generally chosen as $(m + n)k < m + n$, hence the compression effect is accomplished. As a result, V is able to be estimated as a linear combination of the vectors of the basis matrix W and gains matrix H. As the key characteristic of NMF, non-negativity makes the representation purely additive. It is quite different from the other factorization techniques, such as PCA and ICA, whose elements may be negative. In practice, the amplitude of frequency spectrum presented by negative components can not represent any physical meanings.

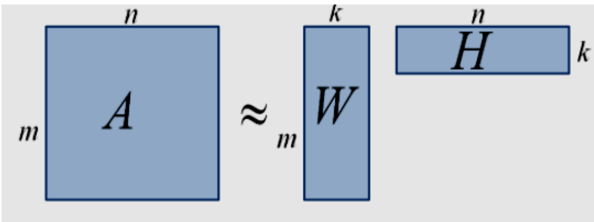

Figure 2. NMF problem.

IV. THE PROPOSED APPROACH

The Figure 3 shows proposed approach for association rule mining. The proposed model has two key components phases. The first component is DAE that takes the transactional database as input and compute the probabilities for each possible basket x based on learned conditional distribution. The NMF is second component, applied to the output obtained by the previous step. The NMF approximate the product vector across all transactions as linear combination of pattern feature vectors and pattern strengths. The output of NMF provides the frequent items.

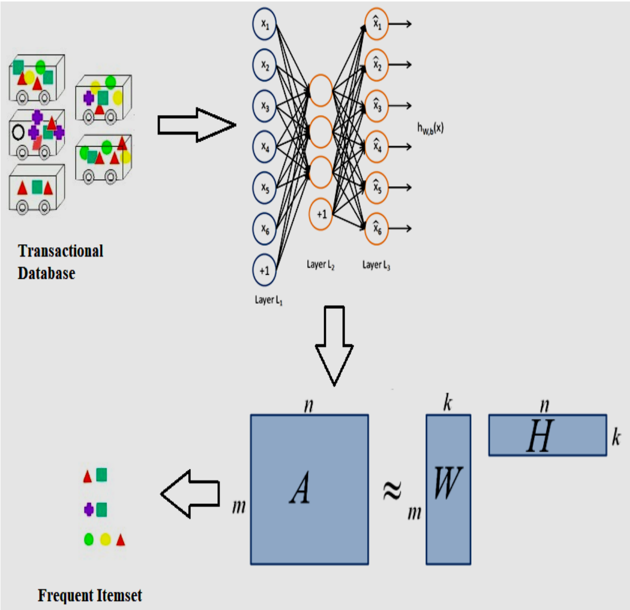

Figure 3. Proposed Architecture.

V. RESULTS AND DISCUSSION

For analyzing purpose, the groceries dataset was used in this study. This dataset contains 9835 rows and 169 columns and the density, which is the percentage of columns that are not empty, is 2.6%. This may seem small but remember that the number of purchases varies from person to person so this affects how many empty columns there are. The plot that is produced gives you an idea of what people were purchasing.

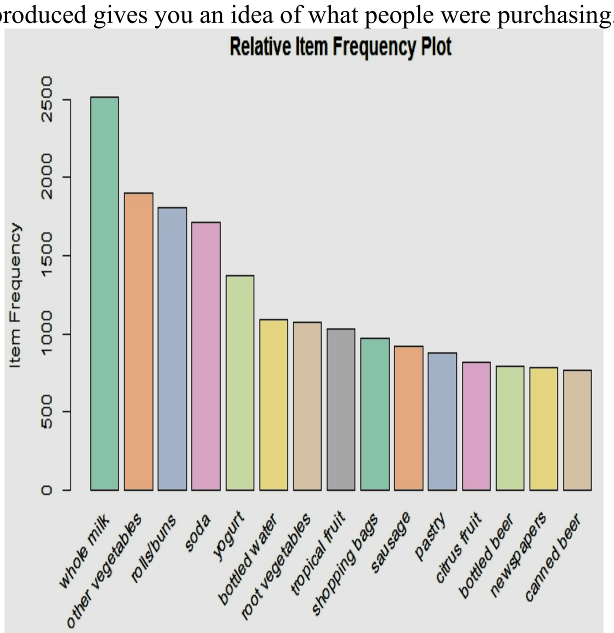

Figure 4. Results

To evaluate the importance of a discovered rule support and confidence are the very popular measures. These measures can be defined as follows:

Support (A): Support of item is the number of times an item occurs in transactions in a database.

Confidence: The likelihood that an exchange that contains the things on the left-hand side of the administer (in our precedent, pencil and paper) likewise contains the thing on the right-hand side (an elastic). The higher the certainty, the more prominent the probability that the thing on the right-hand side will be bought or, as such, the more prominent the arrival rates you can expect for a given run the show.

TABLE I. THE TOP RULES WITH SUPPORT AND CONFIDENCE VALUE

Association Rule
{yogurt}=>{whole milk} support=0.06 confidence=0.40
{yogurt}=>{white bread} support=0.01 confidence=0.06
{yogurt}=>{whipped/sour cream} support=0.02 confidence=0.15
{whole milk}=>{whipped/sour cream} support=0.03 confidence=0.13
{whole milk}=>{white bread} support=0.02 confidence=0.07
{whole milk}=>{sugar} support=0.02 confidence=0.06
{whole milk}=>{tropical fruit} support=0.04 confidence=0.17
{whole milk}=>{waffles} support=0.01 confidence=0.05
{whole milk}=>{soda} support=0.04 confidence=0.16
{tropical fruit}=>{soda} support=0.02 confidence=0.2
{yogurt}=>{tropical fruit} support=0.03 confidence=0.21
{whole milk}=>{whipped/sour cream} support=0.03 confidence=0.13
{soda}=>{root vegetables} support=0.02 confidence=0.11

The Table I show the top rules with support and confidence value. According to the table the top rule obtained is {yogurt} => {whole milk} .This rule have support=0.06 and confidence=0.40. The support as well as the confidence of this rule is highest among the all rules. It means that these two items are frequently bought. From the table very interesting can be made for examples the rule {whole milk} => {white bread} and {whole milk} => {sugar} have low support and confidence than {whole milk}=>{tropical fruit}. It means people buy tropical fruit with milk more frequently as compared to white bread or sugar.

VI. CONCLUSION

The exponential development of PC equipment and framework programming innovation led to a huge supply of advanced and smart PCs. This innovation gives rise to huge data archives accessible for the exchange of administration data recovery and information examination. Manual/physical investigation of this substantial quantity of information is extremely time consuming. This has prompted the need for information mining algorithms and programs. Affiliation lead mining and arrangement systems to locate the related data in vast databases is critical in the present situation.

The expansive amount of data gathered through the arrangement of affiliation standards can be utilized not only to delineate the connections in the database, but additionally utilized for separating various types of information classes in a database. Our research reported here highlights a portion of the current information mining work to aid database studies. The examination of existing calculations recommends that the use of affiliation control mining calculations to aid database investigation will help in better grouping of the vast quantity of information embedded in databases.

REFERENCES

[1] Dianfang Gao, Qiang Liu,“An Improved Simulated Annealing Algorithm for Process Mining”, IEEE 13th International conference on computer supported cooperative work in design, 2009.

[2] Agrawal, R., Imielinski, T. and Swami, A., “Mining Association Rules between Sets of Items in Large Databases”, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington DC, May 1993, 207-216.

[3] Agrawal, R. and R. Srikant, “Fast algorithms for mining association rules in large databases”, Proceedings of the 20th International Conference on Very Large Data Bases, September 12-15, 1994, San Francisco, USA., pp: 487-499.

[4] Frans Coenen, “Data Mining: Past, Present and Future”, The Knowledge Engineering Review, Cambridge University Press, Vol. 26:1, 25-29, 2011.

[5] Han, J., J. Pei, Y. Yin and R. Mao, “Mining frequent patterns without candidate generation: A frequent-pattern tree approach”, Data Mining Knowledge Discovery, 2004, 8: 53-87.

.

[6] M. J. Zaki , S. Parthasarathy, Wei Li , M. Ogihara, “Evaluation of sampling for data mining of association rules”, Proceedings of the 7th International Workshop on Research Issues in Data Engineering (RIDE '97) High Performance Database Management for Large-Scale Applications, p.42, April 07-08, 1997.

[7] J. Mata, J. Alvarez, J. Riquelme, "An evolutionary algorithm to discover numeric association rules", Proc. ACM Symp. Appl. Comput, pp. 590-594, Mar 2002.

[8] J. Mata, J. Alvarez, J. Riquelme, "Mining numeric association rules with genetic algorithms", Proc. 5th Int. Conf. Artif. Neural Netw. Genetic Algorithms, pp. 264-267, Apr 2001.

[9] X. Yan, C. Zhang, S. Zhang, "ARMGA: Identifying interesting association rules with genetic algorithms", Appl. Artif. Intell., vol. 19, no. 7, pp. 677-689, 2005.

[10] Yan, X. and Zhang, C., “Genetic algorithm based strategy for identifying association rule without specifying minimum support”, Expert System with applications, Vol. 36, No. 2, pp3066–3076, 2009.

[11] Liu, D., “Improved genetic algorithm based on simulated annealing and quantum computing strategy for association rule mining”, Journal of Software, Vol. 5, No. 11, pp.1243–1249, 2010.

[12] Indira, K. and Kanmani, S., “Performance analysis of genetic algorithm for mining association rules”, International Journal of Computer Science Issues, Vol. 9, No. 1, pp.368–376, 2012.

[13] Kuo, R.J. and Shih, C.W., “Association rule mining through the ant colony system for national health insurance research database in Taiwan”, in Journal of Computers and Mathematics with Applications, Vol. 54, No. 11, pp.1303–1318, 2007.

[14] Kuo, R.J., Chao, C.M. and Chiu, Y.T., “Application of particle swarm optimization to association rule mining”, Journal of Applied Soft Computing, Vol. 11, No. 1, pp.326–336, 2011.

[15] Fariba, K., Baraani, A. and Kamran, Z., “Efficient mining of association rules based on gravitational search algorithm”, International Journal of Computer Science Issues, Vol. 8, No. 2, pp.51–58, 2011.

[16] Hinton, Geoffrey, et al., "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." IEEE Signal processing magazine 29.6 (2012): 82-97.

[17] Hinton, Geoffrey E., Alex Krizhevsky, and Sida D. Wang, "Transforming auto-encoders." International Conference on Artificial Neural Networks. Springer, Berlin, Heidelberg, 2011.

[18] Paatero, Pentti, and Unto Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values." Environmetrics5.2, 1994, 111-126.

[19] Lee, Daniel D., and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization", Nature 401.6755 (1999): 788.

.