ISSN:0975-9646

A Neural Network Approach for Fragmentation in Distributed Databases

M.Ram Bhupal [1], A. Raga Deepthi[2¶]
[1]*Computer Science & Engineering , Aditya Engineering College , Kakinada,Andhrapradesh,India.*
[2]*Computer Science & Engineering , Sri Sai Aditya Institute of Science and Technology , Kakinada, Andhrapradesh,India.*

Abstract **- In this paper , Neural Network approach is proposed for fragmenting and defragmenting the distributed database for Aditya Educational groups in Andhra Pradesh. Collaboration of different colleges like Engineering, Pharmacy, PG exists. Hence, we need a distributed database environment for designing sub databases and fragmenting them on the sites which are geographically separated. All the fragments in return should be reconstructed as one database. Therefore different techniques are considered for a database fragmentation. A multi-layered neural network architecture is used to implement the fragmentation and defragmentation.**

Keywords **- Fragmentation, Defragmentation, Multilayer architecture, Distributed database, Collaboration.**

I. INTRODUCTION

Different colleges collaboration requires the combination of the theoretical and practical information. Therefore we need a distributed database environment that integrates theoretic and practical values. Distributed database systems generally include more than two geographic remote sites. Inter-related sites has a partial connection in both its hardware and software, which includes the necessity of database management systems and applications[1].

Distributed database design process performs global conceptual design and then local conceptual design that fragments databases on the sites[2] and reconstruction of relations. Distribution design includes requirement analysis, view analysis and integration analysis. These inputs are oriented towards the distributed data base design[3]. Fragmentation and allocation issues simultaneously are considered and integrated in the work[4].

This can be useful in distributed databases where each group of attributes can contain data which have common geographical properties. The Mixed fragmentation is a technique which combines both the horizontal and vertical fragmentation techniques.

There are many database fragmentation techniques developed for organizing data physically in storage devices. Every technique firstly divides the data into groups then assigns those groups to the physical pages. There are two types of fragmentations: horizontal fragmentation and vertical fragmentation. There are, however, some rules which must be followed when defining fragments :

Completeness condition: All the data of the global relation must be mapped into the fragments.

Reconstruction Condition: It must always be possible to reconstruct each global relation from its fragments.

Disjointness Condition: It is convenient that fragments be disjoint, so that the replication of data can be controlled explicitly at the allocation level. This disjointness condition requires that qualifications be mutually exclusive.

Fragments are indicated by a global relation name with an index(fragment index); for example, R_i indicates the i^{th} fragment of global relation R. Physical images can be indicated by a global relation name and a site index. To distinguish them from fragments, we will use a superscript; for example, R^j indicates the physical image of the global relation R at site 'j'.

Horizontal fragmentation consists of portioning the tuples of a global relation into subsets; where each subset can contain data which have common geographical properties. Formal approaches for horizontal fragmentation explained in [5].

The horizontal fragmentation of a relation cannot be based on a property of its own attributes, but is derived from the horizontal fragmentation of another relation is called derived horizontal relation.

The Vertical fragmentation is a global relation which is the sub division of its attributes into groups, fragments which are obtained by projecting the properties. The Mixed fragmentation is a techniques which combines both the horizontal and vertical fragmentation techniques.

II. MULTILAYER NEURAL NETWORK

A multi layer neural network architecture consist of input layer, hidden layers and the output layer. Every layers' output is set as input to the next layer. The final output of the architecture is obtained by the output layer. There can be any number of hidden layers within a feed forward network. There can be any number of neurons in each layer, and it all depends on the problem.

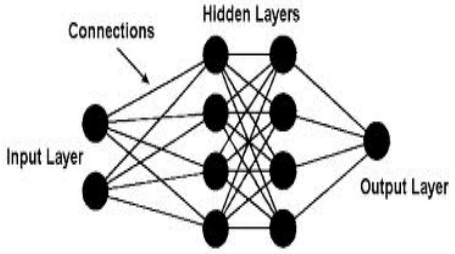

Figure 1. Example of Multilayer Architecture

Figure 1 shows two inputs which are transmitted to the two hidden layers and four outcomes of the hidden layer which are propagated to the output layer. Hence single final output is obtained from the single neuron in the output layer. A simple multi layer architecture computes the output by using the weight value.

Inputs $\{x_1, x_2\}$

Weights $\{w_{11}, w_{21}, w_{31}, w_{41}\}$
Computations in the first hidden layer are
$h_1 = \Sigma\, x_i w_i \quad i = 1....n$ where h_1 represents the output of the first neuron in the first hidden layer. Similarly, we obtain the output of remaining neurons. The obtained outputs h_1, h_2, h_3 and h_4 are set as inputs for the neurons in the second hidden layer. Finally the output of the architecture is
$Y = \Sigma\, h_{2i} w_i \quad i = 1...n$
where h_{2i} is the output of second hidden layer
w_i represents the weight values.

III. RELATED WORK

Educational groups which consist many colleges offering different courses maintain data bases. The database needs to be fragmented and sent to different sites which are geographically separated. To implement this we take an example relation 'R' as shown in table1 which should be fragmented and reconstructed by using multi layer neural network. The following steps shows the fragmentation and reconstruction of global relation 'R'(for simplicity in this example we took only a single relation 'R')

Step1: The relation 'R' is vertically fragmented into R_1, R_2, R_3.... R_n Using Projection (PJ).

Step2: Each sub relation R_1, R_2, R_3.... R_n are further fragmented horizontally i.e. R_1 fragmented into R_{11}, R_{12}, R_{13}.... R_{1n} Using Selection (SL).

Step3: The sub relation R_1 is reconstructed by Union (UN) of R_{11}, R_{12}, R_{13}.... R_{1n}.Similarly R_2, R_3.... R_n are formed.

Step4: The original global relation R is reconstructed by joining (JN) of R_2, R_3.... R_n.

The above steps can be implemented by using multilayer architecture as shown in the Figure 2.

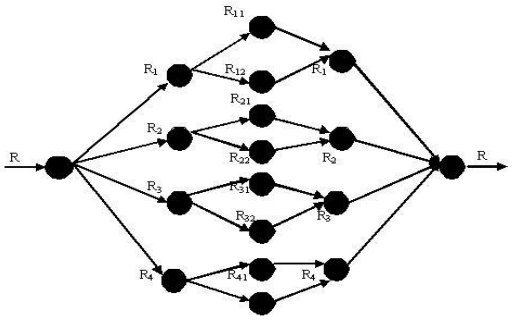

Figure 2

The input to the architecture is the relation 'R' as shown in TABLE I.

The relation 'R' is sent to the neuron in first layer. The neuron in input layer performs the following Projection(PJ) operations to get the R_1,R_2,R_3 and R_4 respectively(for simplicity we fragmented to four sub relations only):

PJ (Emp_No , Emp_Name , Sal , Tax);

TABLE II. SUB RELATION R_1

Emp_No	Emp_Name	Sal	Tax
324	M.Rambhupal	3.5	Nill
325	A.Deepthi	3.5	Nill
326	R.Srinivas	6.2	1.2

PJ (Emp_No , Qualification);

TABLE III. SUB RELATION R_2

Emp_No	Qualification
324	M.Tech
325	M.Tech
326	M.Tech

PJ (Emp_No , Dept_No , Dept_Name);

TABLE IV. SUB RELATION R_3

Emp_No	Dept_No	Dept_Name
324	05	CSE
325	12	IT
326	05	CSE

The above vertical fragment is obtained by using projection. Here, the weight values are Dept_No, Emp_No and Dept_Name.

Similarly, the remaining vertical fragments are derived by using the appropriate weight values.

PJ(Emp_No,Emp_Name,age,Gender,PhoneNo);

TABLE V. Sub relation R_4

Emp_No	Emp_Name	Age	Gender	PhoneNo
324	M.Rambhupal	30	M	9989102261
325	A.Deepthi	27	F	9908910199
326	R.Srinivas	31	M	8106482777

Each sub relation R_1,R_2,R_3 and R_4 are fragmented horizontally by first layer hidden neurons into R_{11}, R_{12} using Selection(SL) and similarly remaining sub relations.

$SL_{sal<4.0}R_1$;

TABLE VI. R_{11}

Emp_No	Emp_Name	Sal	Tax
324	M.Rambhupal	3.5	Nill
325	A.Deepthi	3.5	Nill

$SL_{sal>4.0}R_1$;

TABLE VII. R_{12}

Emp_No	Emp_Name	Sal	Tax
326	R.Srinivas	6.2	1.2

TABLE I. RELATION R

Emp_No	Emp_Name	Dept_No	Dept_Name	Qualification	Sal	Tax	Age	Gender	PhoneNo
324	M.Rambhupal	05	CSE	M.Tech	3.5	Nill	30	M	9989102261
325	A.Deepthi	12	IT	M.Tech	3.5	Nill	27	F	9908910199
326	R.Srinivas	05	CSE	M.Tech	6.2	1.2	31	M	8106482777

The union operation can be performed on horizontally fragmented relations to obtain the relations R_1, R_2, R_3, R_4 in the third hidden layer.

Now, the original relation 'R' is reconstructed by performing join operation on R_1, R_2, R_3, R_4 relations at the output layer. Hence the reconstruction of original relation is done.

IV. CONCLUSION

The first feature to consider in design of a distributed database system is the fragmentation of the central databases into sites with the minimum cost. By trainable neural network model, it is possible to calculate results with the help of weight constants instead of applying fragmentation tests for all the tables and relations of the system. Due to this, determination of a fragmentation technique becomes very simple after training the neural network.

REFERENCES

[1] Ceri, S., Pelagatti, G. (1986). Distributed Databases Principles and Systems, McGraw Hill

[2] Özsu, M., T., Valduriez,P. (1991). Principles of Distributed Database Systems, Prentice Hall, New Jersy.

[3] Ceri,S.,Navathe, S., Widerhold, G.(1983). Distribution Design of Logical Database Schemas, IEEE Transactions on Software Engineering, Vol.9,No.4,487-504.

[4] Tamhankar,A.,M.,Ram,S. (1998). Database Fragmentation and Allocation: An Integrated Methodology and Case Study, IEEE Transactions On Systems, Man, and Cybernetics, Vol.28,No.3,288-305.

[5] .J.B. Rothine et al., "Introduction to a system for Distributed Databases(SDD-1),"ACM-TODS, 5:1,1980.

[6] Gruenwald, L. & Eich, M.(1993). Selecting a Database Partitioning Technique, Journal of Database Management, Summer, 27-39.

[7] Gelenbe E., Stafylopatis A., "Global behavior of homogeneous random neural systems", Vol.15, pp. 534-541, 1991

[8] Gelenbe E., "Learning in the recurrent random neural network", Neural Computation, Vol. 5, No. 1, pp 154-164, 1993.

[9] Haykin S.(1999). Neural Networks Second Edition, Prentice Hall Inter. Inc., New York.

[10] Ambrose Goicoechea. 2000. Requirements Blueprint and Multiple Criteria For Distributed Database Design. International Council on Systems Engineering (INCOSE) 2000.

[11] Yin-Fu Huang, and Jyh-Her Chen. 2001.Fragment Allocation in Distributed Database Design. *Journal of Information Science and Engineering 17, 491-506 (2001).*

[12] Cyrus Shahabi, Latifur Khan, and Dennis McLeod. 2000. A Probe-Based Technique to Optimize Join Queries in Distributed Internet Databases.Knowledge and Information Systems (2000) 2: 373-385

[13] T. Kohonen, Adaptive associative and self-organizing functions in neural computing, Appl. Optics 26 (1987) 4910-4918.

[14] H. Chen, Machine learning for information retrieval:Neural networks symbolic learning and genetic algorithms, J. Am. Societyfor Info. Sci. 46 (3) (1995) 194-216.

[15] S. Fahlman, Fast-learning variations on back-propagation: An empirical study, in: Proceedings of the 1988 Connectionist Models Summer School, 1988, pp. 38-51.

[16] D.H. Fisher, K.B. McKusick, An empirical comparison of id3 and backpropagation, in: Proceedings of the 11th Joint Conference of Artificial Intelligence, 1989, pp. 788-793.

[17] S.Y. Kung and J.N. Hwang, "Parallel Architectures for Artificial Neural Nets", Proc. IEEE Int. Conf. On Neural Networks, San Diego, California, 1998, pp. II-165 – II-172.

[18] R.P. Lippman, "An Introduction to computing with Neural Nets", IEEE ASSP Magazine, April 1987,pp. 4-22.

AUTHOR'S PROFILE

Mr.M.Rambhupal received the M.Tech degree in Computer Science and Engineering from the JNT University, Kakinada, Andhra Pradesh. Presently he is working as a Computer Engineering Faculty at Aditya Engineering college since five years and the total teaching experience is eight years. His area of interest is "Distributed Databases."

Ms.A.Raga Deepthi received the M.Tech degree in computer Science and Engineering from the JNT University, Kakinada, Andhra Pradesh. From 2007 she is working as an Assistant Professor at Sai Aditya Institute of Science and Technology, Kakinada. Her area of interest is "Neural Networks."