

Modified Map Reduce Model using Fuzzy Neural Network

Vatsal Gor¹

¹Department of Computer Science & Engineering

¹Narayan Shastri Institute of Technology

Abstract— The MapReduce paradigm is now the de facto standard for processing and generation large scale datasets. Generally in MapReduce paradigm, user specifies a map function that processes a key/value pair to generate a set of intermediate key-value pairs & reduce function which merges all intermediate values associated with the same intermediate key. This paradigm is enough to solve many real world problems. Here fuzzy neural network is applied in the MapReduce paradigm to improve its efficiency. It is also referred as Neuro-fuzzy which is combination of artificial neural network and fuzzy logic. Now, here the key/value generated pair via map function is feed as input in neuro fuzzy system or model to improve efficiency of current map reduce paradigm. The proposed Fuzzy Neural Network model aims at training the fuzzy Neural Network model in MapReduce programming model. Since the MapReduce programming model has the ability to rapidly process large amount of data in parallel and can use the outcome of training process for the prediction purpose. This can produce better prediction accuracy. In order to train large data in distributed mode it can be effective in reducing the processing time.

Key words: BigData, fuzzy Neural Network, ANFIS

I. INTRODUCTION

MapReduce is introduced firstly by Jeffrey Dean and James Ghemawat at Google Inc[1]. In 2004 firstly. It is used to processed hundreds of peta bytes of structured or unstructured or semi structured verity of data in parallel y. This process is divided in to map phase, then shuffle and then finally in reduced phase which gives the summarization. These phases are stateless mainly in nature mainly. Input files are divided in to chunks of files and fed to the different mappers. Then intermediary key value pairs are generated and then fed to the shuffle phase. This shuffle phase combines related date means value with similar keys and finally these shuffled data are finally fed to reduce phase which summarize the data.

Fuzzy Neural Network is a combination of two words Fuzzy logic and Neural Network. Neural Network are inspired from biological neurons or central nervous system in brains. Here artificial neural network which is interconnected group of nodes consisted in several layers. In simple model they are basically 3 layers as one is input last is output and middle one is hidden layer having adaptive weights. Traditional logic system results in in binary means as True or False, but instead of that we need results that showing the effect of intermediary values that lies between these 1 and 0 or true of false value boundaries. So fuzzy logic represents the impact of intermediary states for a logic as many values logic. So in Fuzzy Neural Network choose parameters from fuzzy rules and combine with neural nodes. [4]

II. SUMMARIZATION OF CURRENT ISSUES IN MAP REDUCE

In general there are various issues in current hadoop/Map Reduce are such as Performance Issues, Programming model issues, Configuration and Automation Issues, Performance Optimization.[2] Here also one shortcoming is static procedure of map-reduce-shuffle which is materialized lack of support for iterative data sets with minor changes as well as major changes and balancing of state between jobs too.

We take an example for social networking sites like google+ or facebook[3] which suggest suggestion of events, ads, friends activities and relative post as per users' profiles. Now smaller change in that users' profiles or interest need to again run these whole map reduce process which is time consuming and costly. So for these types of situation if we use already computed map function output for regenerate suggestions for any changes in input, so we have to compute only modified data instead of whole process again.

As shown below comparison table for Hadoop Variations.[2]

	Optimization Type	Major Contribution	Open-source/ Available to use	Transparent to existing Applications
MapReduce Online	Performance, Programming model	Pipelining , Online aggregation	Yes	Yes
EARL	Performance	Fast approx. query result	Yes	No
Hadoop++	Performance	Performance gain for relation operation	No	Yes
HAIL	Performance	Performance gain for relation operation	No	No
MRShare	Performance	Concurrent work sharing	No	No
ReStore		Reuse of previously computed results	No	No
Skew/Tune	Performance	Automatic skew mitigation	No	yes
CoHadoop	Performance	Communication minimization by data co-locations	No	No
HaLoop	Programming model	Iteration Support	Yes	No
Incoop	Programming model	Incremental processing support	Yes	No

Starfish	Tuning, performance	Dynamic self-tuning	No	Yes
Maninal	Tuning, performance	Automatic data aware optimization	No	Yes

Table 1: Hadoop Variations.[2]

III. PROPOSED METHODOLOGY

In traditional map reduce phase, input chunk of data is divided in map phases. Now Map phase generate key, value intermediate pairs From these fed as input in fuzzy neural

network. Here rules are generated from similar keys are combine together in initial nodes of fuzzy neural network as shown. Then final summarization output is generated as output.

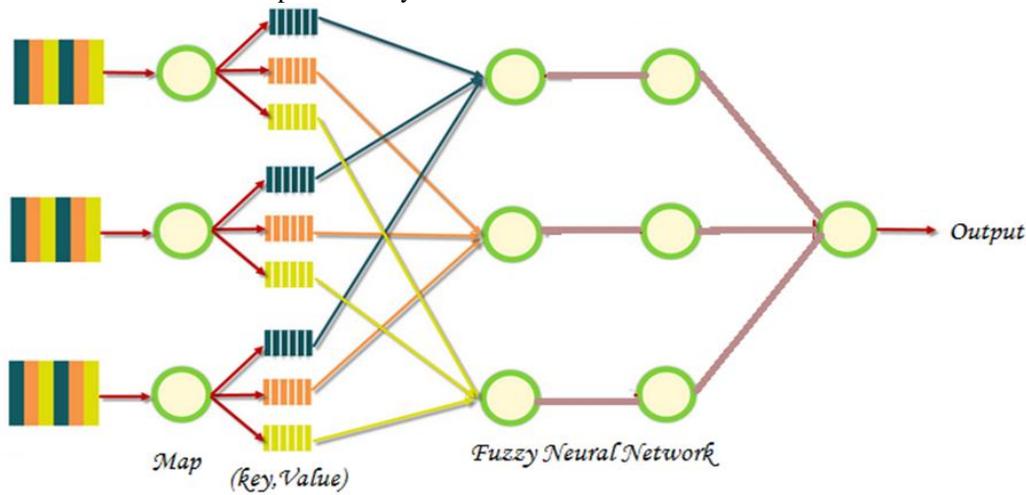


Fig. 1: Proposed methodology

Here suggested model we can demonstrate via using MATLAB R2014 [5]which has introduced new feature for Map Reduce & big data processing features[6] and Apache Hadoop with YARN.[7]

IV. CONCLUSION & FUTURE WORK

This dissertation include study of current popular map reduce framework paradigm with Apache Hadoop. Here Fuzzy neural network is introduce with Map reduce framework. This lead us to new modified approach of map reduce with ANFIS. ANFIS has proven its efficiency with its various applications in image processing, control theory and other related field. So the portion after map function which shuffle and reducer are replace by ANFIS. So it will lead to an efficiently and faster calculation of that part of current map reduce framework.

REFERENCES

- [1] Jeffrey Dean and Sanjay Ghemawat." MapReduce: Simplified Data processing on large clusters. 2004
- [2] Anam Alam, Jamil Ahmed. "Hadoop Architecture and its issues", 2014 International Conference on Computational Science and Computational Intelligence @2014 IEEE ISBN – 978-1-4799-3010-4
- [3] Vasiliki Kalavri , Vladimir Vlassov. "MapReduce: Limitation , Optimizations and Open Issues" 12th International Conference on Trust , Security and Privacy in Computing and Communications 2013
- [4] Jang, Jyh-Shing R (1991). "Fuzzy Modeling Using Generalized Neural Networks and Kalman Filter Algorithm". Proceedings of the 9th National Conference on Artificial Intelligence, Anaheim, CA, USA, July 14–19 2. pp. 762–767

- [5] Archana C S, Janani K, NALINI B S, Pooja Krishnan, Savita K Shetty. "M2M Automation: Matlab-TO-Map Reduce Automation", IRF International Conference ,April 2014, Chennai, India, ISBN: 978-93-84209-05-6
- [6] <http://in.mathworks.com/help/matlab/mapreduce.html> (last accessed on 10th December 2014)
- [7] http://en.wikipedia.org/wiki/Apache_Hadoop (last accessed on 10th December 2014)