

Research on Distributed TBL Support Vector Machine Based on MapReduce

Jianwei Hu, Aiping Sheng, Zhuge Huixiang

College of Medicine, Jinhua Polytechnic, Jinhua, 321001, China

Jianwei_Hu2014@yeah.net

Keywords: MapReduce frame, TBL method, Cluster model, M-tree, Internal medicine.

Abstract. In the research process of internal medicine learning method, chunk cognitive processing is not easy to explore because of the complexity of having overall cognitive understanding and structure semantic understanding. Through the attempt of TBL method, this paper carries out chunk cognitive evaluation under the MapReduce framework, the method is validated by using clustering mathematical model, and then using cluster graph visualization function carries out analysis each cluster for the chunk cognitive process. Under the continuously increasing of the number of learning cognitive clusters, the method can display the individual chunk cognitive level of data set, finally using M- tree node overflow can get the contribution subtree model of the TBL method for the chunk perception. Based on the TBL method and the control group experiment, we discover that the contribution of TBL method is larger for the chunk perception, which can effectively improve the chunk cognitive performance, but also provide a new computer method for the inquiry of internal medicine learning method.

Introduction

The emergence of cloud computing brings new opportunity and challenge for the development of data mining technology, Hadoop is one of the many cloud computing platform, now it is the Apache open source project [1,2]. Hadoop is composed of distributed file system and MapReduce programming framework, we run the searching and sorting task of the benchmark performance under the MapReduce framework, to analyze the effect of different TBL method. This paper uses clustering principle to improve the clustering algorithm by using the M- tree node overflow model, and then we can obtain a new chunk cognitive evaluation method, in which the overall frame is shown in Figure 1.

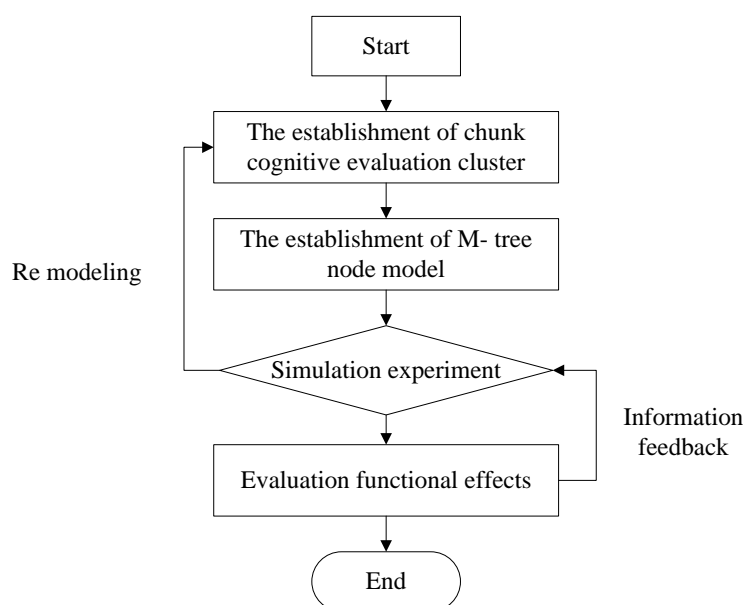


Fig.1: The chunk cognitive clustering method framework

Figure 1 shows the overall framework of chunk cognitive evaluation, and the main process is to establish chunk cognitive clustering evaluation [4-6]. Through the spillover effect of M-tree nodes,

we can get the degree of chunk cognitive contributions function, and then the effect of TBL is evaluated by simulation results, using the information feedback function goes through constant re modeling, finally we get the TBL best plan.

Design of Chunk Cognitive Clusters Mathematics Model and Algorithm

In order to study the effect of TBL method for cognitive, this method is verified by using cluster mathematical model [7]. The cluster diagram is a visualization technique, it is used to analyze each cluster in the chunk cognitive process, and then it will display the individual data set to be assigned to clusters, when the number of clusters is increasing continuously as shown in Figure 2.

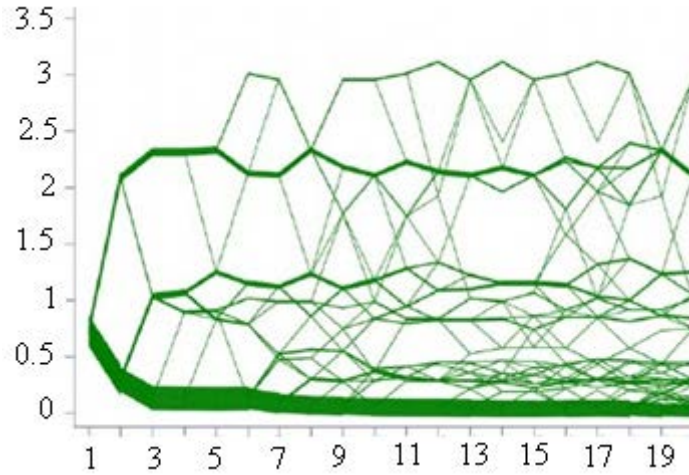


Fig.2: Cluster diagram graphic

Selecting the number of clusters is an important parameter in each cluster analysis, and this technology can analyze each cluster change results to achieve a better understanding in the number of different clusters [8,9]. For the chunk cognitive cluster, the linear correlation coefficient can be expressed as

$$F(X) = (f_1(X), f_2(X), \dots, f_p(X))^T$$

$$s.t. \ f_i(X) \geq 0$$
(1)

Among them, $f(x)$ shows chunk cognitive function, and TBL method for the degree of chunk cognitive contribution can be represented by a utility function. Assuming the contribution rate of target utility function is

$$U(x) = U(f_1, f_2, \dots, f_p)$$
(2)

When the degree of chunk cognitive contribution carries out similarity search, in order to simplify the search path, we will first compare the search coverage and the representative elements coverage in the node to determine the semantic correlation subtree, then to further search in subtree [10]. The schematic diagram of M-tree node is shown in Figure 3.

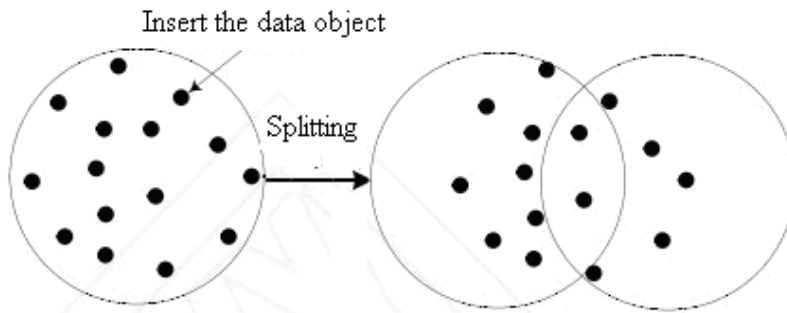


Fig.3: Schematic diagram of M-tree node overflow

As shown in Figure 3, the representative element of selecting nearest distance is corresponding to the subtrees insertion when inserting data; when nodes lead to overflow, we need to perform split

operation [11,12]. Assuming splitting function is $a_{ij} = f_i(x_j)$, the degree of contribution can be further written

$$U(x_j) = U(a_{1j}, a_{2j}, \dots, a_{pj}) \quad (3)$$

Assumption $f_1(X_1), f_2(X_2), \dots, f_p(X_p)$ has the same dimension, f_i respectively gives same weight coefficients ω_i according to certain rules. The linear weighted and contribution degree function of the chunk cognitive degree is

$$U(X) = \sum_{i=1}^p \omega_i f_i(X) \quad (4)$$

In order to realize the contribution of TBL chunk algorithm, this paper adopts the programming way to design algorithm, the main algorithm is as follows [13]:

```
#/usr/bin/perl
While(<STDIN>){
s/(\s)/\t1/;
print STDOUT;}
#/usr/bin/perl
While(<STDIN>){
s/\t//;
print STDOUT;
public class ProtoCluster implements Writable {
public ProtoCluster() {
super();}
private IntWritable Id;
private DoubleArrayWritable Center;
private DoubleArrayWritable[] ClusterMembership ;
private IntWritable Size;
.....
```

Analysis of the TBL Learning Effect based on MapReduce Programming Framework

In the development process of MapReduce programming framework, it needs to install the Eclipse development environment. This paper studies the corresponding Hadoop developing plug-in JAR, and the Hadoop plug-in should put into the plugin of Eclipse directory as shown in Figure 4.

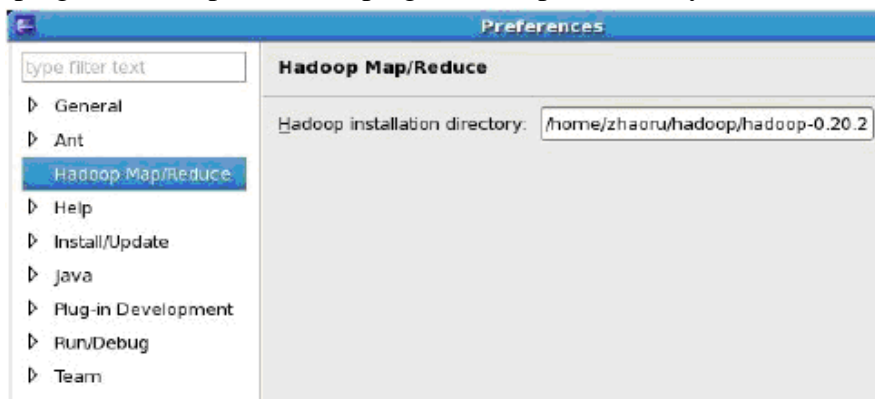


Fig. 4 The Hadoop introduction sample chart

Figure 4 shows the sample graph of Hadoop introduction process. Through a new MapReduce program created by Eclipse, this paper develops cluster functions in Eclipse, and then the evaluation model is derived, to evaluate the contribution of TBL as shown in Figure 5.

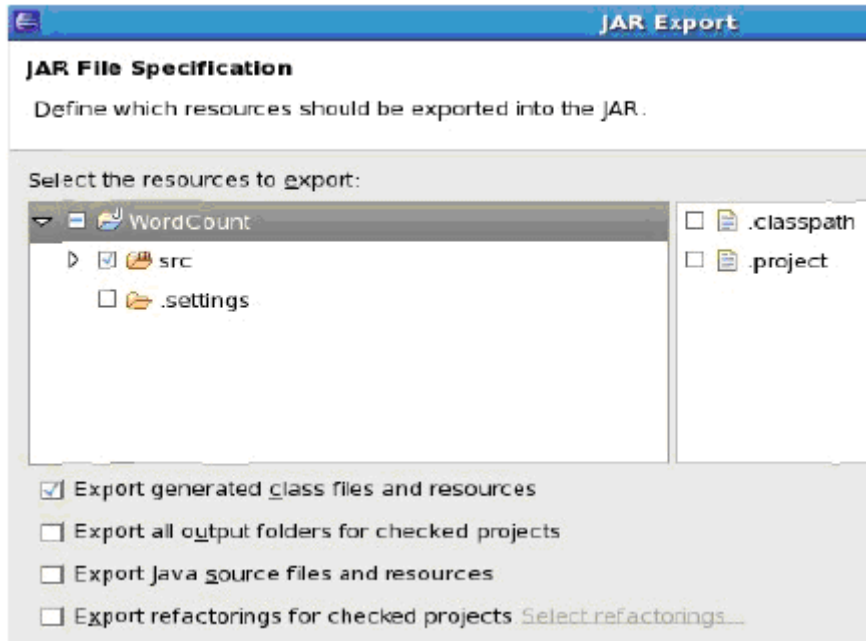


Fig.5: Export package interface diagram

Figure 5 shows the Export package interface diagram of the evaluation results, the packaged copies to the Hadoop installation folder, and then it is executed on the command line [14]. Before putting into a text file in HDFS, the WordCount is as input file. As shown in Table 1, the results are obtained by calculation.

Table 1: TBL chunk cognitive results

Question number	A	B	C
1	34 (65.4)	7 (13.4)	11 (21.2)
2	29 (55.8)	23 (44.2)	
3	33 (63.5)	13 (25.0)	6 (11.5)
4	30 (57.7)	14 (26.9)	7 (13.4)
5	24 (46.1)	20 (38.5)	8 (13.4)
6	30 (57.7)	14 (26.9)	8 (13.4)
7	22 (42.3)	21 (40.4)	9 (17.3)
8	20 (38.5)	21 (40.4)	9 (17.3)

Note: the figure in () is the percentage (%).

Table 1 shows the chunk cognitive result table obtained by using TBL method, it can be seen from the table that each problem passes the different numbers, and it is related to the degree of TBL acceptance [15]. In order to further reveal the advantages of TBL method, control group carries out comparison with the experimental group, the results are obtained as shown in Table 2.

Table 2: The comparison of the experimental group and the control group internal medicine summative exam

Group	The number of people	Performance ($\bar{X} \pm S$)	<i>t</i> value	<i>p</i> value
Experimental group	52	79.33 ± 6.85	7.281	< 0.001
Control group	50	67.68 ± 8.02		

The comparisons of scores between two groups have significant difference ($p < 0.001$).

In Table 2, $X \pm S$ represents measurement data, it uses t test; count data uses X_2 test; $p < 0.05$ or $p < 0.01$ indicates the difference. Through the calculation, we found that the scores of TBL combination control group is significantly different, the different testing time results are plotted curves, the result curve is obtained shown in Figure 6.

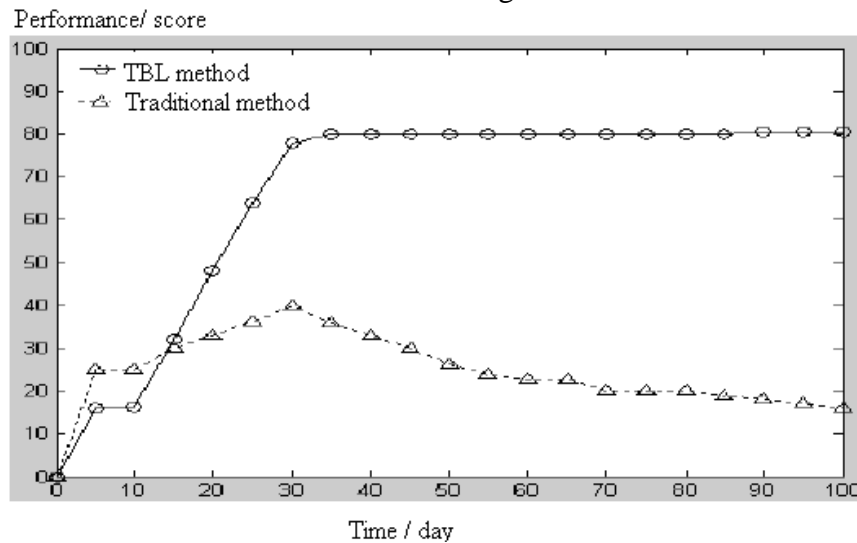


Fig.6: Different time TBL contribution curve

Figure 6 shows the chunk cognitive contribution degree curve of TBL method and traditional method in different test time, wherein the contribution mainly goes through the grades to reflect. It can be seen from the chart that the TBL method has obvious effect on the department of internal medicine chunk cognitive learning, which is a very good cognitive method.

Summary

- (1) Using cluster mathematical model and combined with the cluster graph visualization function, this paper proposes a new chunk cognitive evaluation method, the method is based on M- tree node overflow model as the basis, and then through the MapReduce framework for algorithm programming, we can obtain the degree model of a single chunk cognitive contributions.
- (2) The use of Eclipse development environment designs the MapReduce programming framework, and using JAR text file evaluates on the effect of TBL method. By comparing the experimental group and control group, we found that the TBL method has larger contribution for the chunk perception, and it is a fast and effective method for improving the chunk cognitive level, which improves the theoretical reference for the study of internal medicine learning method.

Acknowledgments

The work was supported by the Education Department of Zhejiang Province in 2014 with the project number 2014YB010 and the project name Design and Application of TBL Teaching Method in Internal Medicine.

References

- [1] H.Z. Wang, A.Q. Peng. Data mining research status and its development trend. Industrial automation, 2011(2): 78-80.
- [2] H. Li, S. Liu, M. Li. Study on data mining theory and its application. Fault block oil and gas field, 2011(1): 112-113.
- [3] Y.J. Xia, J.J. Zhou, C.S. Xiang. The research progress of modern data mining technology. Jiangxi Journal of Agricultural Sciences, 2012(4): 136-137.
- [4] F. Zhou, X.W. Li. An improved MapReduce parallel programming model. Science Forum, 2011(2): 98-99.

- [5] F.L. Shan. Data mining technology research and its application. Modern enterprise education, 2011(6): 23-26.
- [6] H.W. Zhan. Study on the psychological reality of L2 chunks from speech processing. Language and foreign language teaching, 2012(6): 34-37.
- [7] H.Q. Yu, Y. Liu, P. Li. Application of robust recursive kernel learning modeling method in blast furnace process. Journal of Zhejiang University, 2012(4): 45-48.
- [8] J.H. Yang. Discussion on the chunk characteristics, nature determination and role. Journal of Jinan University College of Chinese language and culture, 2011(2): 49-50.
- [9] G.L. Yan, L.L. Zhang, S.S. Sun, X.J. Bai. Chinese "subjective word" characterization and processing. Journal of psychology, 2013(4): 41-42.
- [10] L.J. Ma, J.H. Hu, J.J. Zhang. Study of the Chinese idiom semantic properties and its relationship. Language application, 2013(1): 91-95.
- [11] Y. Liu, H.Q. Wang, P. Li. The support vector machine regression method of adaptive partial least squares in fermentation process online modeling. Journal of chemical industry and engineering, 2012(8): 12-15.
- [12] H.W. Zhan. The tools and methods of chunk extraction in the corpus. Foreign language teaching, 2011(2): 81-84.
- [13] T. Liu, H.W. Zhan. The effects of strange chunks highlights on its incidental acquisition and text understanding. Journal of Hubei University of Education, 2011(3): 34-38.
- [14] S.H. Huang, H.W. Zhan. The latest progress of chunk cognitive processing research. Foreign language (Journal of Shanghai International Studies University), 2011(2): 78-81.
- [15] C.X. Tao. The discussion of data mining and its development status. Technology science information, 2012(4): 24-28.