

the difference in the speed of the two methods is gradually expanding. Until the number of nodes increases to a certain degree, the computing time of the classification becomes very short, and the gap between them is gradually reduced.

Classification effect comparison

From the existing research results [2] Pusu Bias classification recognition classification of degree 86. 1%, here no longer. The total recognition rate of the improved classification method was 91. 2%. The results of classification are listed in Table 1. It shows that the improved method can improve the accuracy of classification to a certain extent.

Table 1. Improved classification experiment results

Classification prediction	car	Sports	healthy	Tourism	IT	recruit	education	Culture	Economics	military	recall/%
automobile	7843	3	10	21	43	2	14	7	54	3	98.0
Sports	7	7723	15	29	37	56	28	80	11	8	96.5
healthy	27	16	7185	77	124	56	325	117	6	67	89.8
Tourism	31	24	19	7417	66	78	68	147	114	36	92.7
IT	5	12	77	147	7389	54	214	4	76	22	92.4
recruit	7	5	141	22	76	6997	445	128	142	37	87.5
education	2	15	263	7	101	24	7551	9	14	14	94.4
Culture	44	16	347	112	76	54	976	6123	185	67	76.5
Economics	201	3	17	137	303	38	179	44	7012	66	87.7
military	18	5	14	28	45	22	54	93	24	7697	96.2
recall/%	95.8	98.7	88.8	92.8	89.5	94.8	76.6	90.7	91.8	96.0	

5. CONCLUSIONS

In this paper, the algorithm based on the naive Bias classification algorithm based on the characteristics of distributed computing has been improved, and deployed in the Hadoop cloud computing platform, the corresponding testing and improvement. The experiments show that the improved method can effectively improve the accuracy and the speed of the classification algorithm in dealing with the mass data, which compared with the naive Bayes method.

ACKNOWLEDGMENT

This work was supported by the Ningbo Natural Science Foundation under grant No. 2015A610141, and by the Zhejiang Science and Technology Program under grant No. 2016C33195, and by the National Undergraduate Training Programs for Innovation and Entrepreneurship under grant No. 201510876021.

REFERENCES

[1] Jing Y. S., Pavlovic V., Rehg J. M., "Boosted Bayesian network classifiers," *Machine Learning*, 2008, vol. 73, no. 2, pp. 155-184.

[2] Webb G. I., Boughton J. R., Zheng F., et al. "Learning by extrapolation from marginal to full-multivariate probability distributions: Decreasingly naive Bayesian classification," *Machine Learning*, 2012, vol. 86, no. 2, pp. 233-272

[3] Tillman R. E., "Structure learning with independent non-identically distributed data," *Proceedings of the*

26th Annual International Conference on Machine Learning, New York, 2009, pp. 1041-1048.

[4] Su J., Zhang H., Ling C. X., et al., "Discriminative parameter learning for Bayesian networks," *Proceeding of the 25th International Conference on Machine Learning Helsinki*, Finland, 2008, pp. 1014-1023.

[5] Ekanayake J., Li H., Zhang B., et al. "Twister: A runtime for interactive MapReduce," *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, Chicago, Illinois, USA, 2010, pp. 810-818.

[6] Dean J., Ghemawat S. Mapreduce, "Simplified data processing on large clusters," *Proceedings of the 6th Symposium on Operating System Design and Implementation*, SanFrancisco, California, USA: USENIX Association, 2004, pp. 137-150.

[7] Thusoo A., Sarms J. S., Jain N., et al., "Hive: A warehousing solution over a map-reduce framework," *Proceedings of the Conference on Very Large Databases*, Ly-on, France, 2009, pp. 1626-1629.

[8] Dean J., Ghemawat S., "Map/Reduce advantages over parallel databases include storage-system independence and fine-grain fault tolerance for large jobs," *Communications of the ACM*, vol. 53, no. 1, pp. 72-77.

[9] Dittrich J., Quiane-Ruiz J. A., Jindal A., et al., "Hadoop++: Making a yellow elephant run like a cheetah(without it evennoticing)," *Proceedings of the VLDB Endowment*, vol. 3, no. 1, pp. 518-529, 2010.

[10] Bu Y., Howe B., Balazinska M., et al., "HaLoop: Efficient iterative data processing on large clusters," *Proceedings of the VLDB Endowment*, vol. 3, no. 1, pp. 285-296, 2010.