6. Conclusions and Future Work

MapReduce works well on unstructured or semistructured data, since it is designed to interpret the data at processing time. In other words, the input keys and values for MapReduce are not an intrinsic property of the data, but they are chosen by the person analyzing the data. The problem of exact probabilistic inference in an arbitrary Bayes network is NP-hard. The process is time consuming and complex. To speed up the processing, we need to run parts of the subnetwork in parallel. In theory, this is straightforward: we could process different years in different processes, using all the available hardware threads on a machine. To date, a MapReduce can help us save computation time by processing the jobs in parallel. The solving of NP-hard problems can possibly be achieved. This work addresses the application of a MapReduce based distributed computing framework, Hadoop, to Bayesian network model to speed up the Bayesian update and inference processes. We present an analytical framework for understanding the transformation of Bayesian network model to Map and Reduce tasks.

As we plan for the future study, we need to conduct experiments to measure the performance of the proposed framework and compare the results obtained form different experimental settings.

References

[1] J. Pearl, "Probabilistic reasoning in intelligent systems," Networks of Plausible Inference, San Mateo, CA, Morgan Kaufmann Publishers, 1988.

[2] T. Bayes, "An essay towards solving a problem in the doctrine of chances," Philosophical Trans. Royal Soc. Of London, 1763.

[3] S. L. Lauritzen and D. J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems, " J. Royal Statistical Soc., 1988.

[4] D. Heckerman, "A tutorial on learning with bayesian networks," Technical Report MSR-TR-95-06, Microsoft, http://research.microsoft.com/research/pubs/view.aspx?msrtrid=MSR-TR-95-06, 1996.

[5] D. Heckerman, "Learning bayesian networks: The combination of knowledge and statistical data, "Proc. KDD Workshop, 1994.

[6] N. Friedman and M. Goldszmidt, "Learning bayesian networks with local structure, " Learning in Graphical Models, 1999.

[7] Cooper, Gregory F. (1990). The computational complexity of probabilistic inference using Bayesian belief networks. Artifical Intelligence, 42(2-3):393-405.

[8] Pearl, Judea (1986). Fusion, propagation, and structuring in belief networks. Artificial Intelligence, 29(3):241-288.

[9] Lauritzen, Steffen L. & David J. Spiegelhalter (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). Journal of the Royal Statistical Society, Series B (Methological), 50(2):157-224.

[10] J. Dean and S. Ghemawat. Mapreduce: Simplifed data processing on large clusters. Communications of the ACM, 51(1):107{113, 2008.

[11] A. Basak, I. Brinster, X. Ma, and O. J. Mengshoel, Accelerating Bayesian Network Parameter Learning Using Hadoop and MapReduce, BigMine'12 August 12, 2012 Beijing, China.

[12] N. Pansare, V. Borkar, C. Jermaine, T. Condie (2011), Online Aggregation for Large MapReduce Jobs, Proceedings of the VLDB Endowment, September 2011, Seattle, Washington, Vol. 4, No. 11, pp.1135-1145.

[13] K. Zhai et. al, Mr. LDA: a flexible large scale topic modeling package using variational inference in MapReduce, Proceeding WWW '12 Proceedings of the 21st international conference on World Wide Web, ACM New York, NY, USA ©2012, p. 879-888.

[14] Hadoop: The Definitive Guide. Tom White, Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, Printed in the United States of America (2009).

[15] Dagum, P. & Luby, M., 1993. Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artif. Intell. 60, 141–153.

[16] Pradhan et al., 1994, Malcolm Pradhan, Gregory Provan, Blackford Middleton, and Max Henrion, Knowledge engineering for large belief networks, In Proceedings of the Tenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-94), pages 484-490, San Francisco, CA, 1994. Morgan Kaufmann Publishers.

[17] Lauritzen, Steffen L. and David J. Spiegelhalter (1988) "Local computations with probabilities on graphical structures and their application to expert systems" in Journal Royal Statistics Society B, 50(2), 157-194.