Statistical Approximation of Likelihood Functions of Restricted Boltzmann Machines

Muneki Yasuda
Graduate School of Information Sciences, Tohoku University

Generally, it is difficult to learn the parameters of graphical models by using maximum likelihood (ML) estimation because of the intractability of computing the normalizing constant and its gradients. Maximum pseudo-likelihood (PL) estimation [1] is a statistical approximation of the ML estimation. Unlike the ML estimation, the maximum PL estimation is computationally fast; however, the estimates obtained by this method are not very accurate.

Composite likelihoods (CLs) [2] are higher-order generalizations of the PLs. Asymptotic analysis has shown that maximum CL estimation is statistically more efficient than the maximum PL estimation [3]. It is known that the maximum PL estimation is asymptotically consistent [1]. Similarly, the maximum CL estimation is also asymptotically consistent [2]. Furthermore, the maximum CL estimation has an asymptotic variance that is smaller than that of the maximum PL estimation but larger than that of ML estimation [3]. Recently, it has been found that the maximum CL estimation corresponds to a block wise contrastive divergence learning [5].

In the maximum CL estimation, one can freely choose the size of "blocks" that contain several variables, and it is widely believed that by increasing the size of blocks, one can extract more dependence relations in a model and increase the accuracy of the estimates [5]. We present a systematic choice of blocks in the maximum CL estimation. Increasing the size of these blocks, one is guaranteed to obtain values that are close to the true likelihood. We apply our maximum CL estimation to restricted Boltzmann machines (RBMs) [6] which are important components of the deep learning, and we present results of numerical experiments performed using synthetic data.

References

[1] J. Besag. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society D (The Statistician)*, Vol. 24, No. 3, pp. 179–195, 1975.

[2] B. G. Lindsay. Composite likelihood methods. *Contemporary Mathematics*, Vol. 80, No. 1, pp. 221–239, 1988.

[3] P. Liang and M. I. Jordan. An asymptotic analysis of generative, discriminative, and pseudo-likelihood estimators. *Proceedings of the 25th International Conference on Machine Learning*, pp. 584–591, 2008.

[4] A. Asuncion, Q. Liu, A. T. Ihler, and P. Smyth. Learning with blocks: composite likelihood and contrastive divergence. *Proceedings of the 13th International Conference on AI and Statistics (AISTAT)*, Vol. 9, pp. 33–40, 2010.

[5] G. E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, Vol. 8, No. 14, pp. 1771–1800, 2002.