An Estimation Method for Bradley-Terry and its Related Models based on the Bregman Divergence

Yu Fujimoto[1] **Hideitsu Hino**[2] **Noboru Murata**[3]

[1]**Aoyama Gakuin University,**
5-10-1 Fuchinobe, Sagamihara,
Kanagawa 229-8558, Japan
yu.fujimoto@it.aoyama.ac.jp

[2,3]**Waseda University,**
3-4-1 Ohkubo, Shinjuku,
Tokyo, 169-8555, Japan
hideitsu.hino@toki.waseda.jp[2]
noboru.murata@eb.waseda.ac.jp[3]

The Bradley-Terry (BT) model [3] is a basic probability model for representing user preference or ranking data, and it is also used for formulating classification problems [6, 10] in recent years. Estimation methods of this model have been discussed from several contexts [6, 8], and most of the proposed methods are based on the sum of weighted Kullback-Leibler (KL) divergences.

The purpose of our work is to interpret the estimation mechanism of the BT model and its related models from a viewpoint of information geometry [2]. Based on this interpretation, we generalize the BT model and its estimation method by using other divergences to improve robustness and accuracy of estimation in practical scenes.

We first reformulate the estimation method in a framework of the *em* algorithm [1], which is an information geometrical interpretation of the EM algorithm [9]. Observations for estimating some kinds of probabilistic models including the BT model can be regarded to form "m-flat" data manifolds; those are useful subsets in the space of probability models from a geometrical perspective. For example, an m-flat manifold for the BT model is constructed from a set of comparison data between specific two objects (items, players, etc.), and from comparison data between various pairs, the corresponding number of m-flat manifolds are constructed. Intuitively speaking, an estimator for the BT model is obtained as "the nearest point from all the m-flat data manifolds" and an objective function for the estimation is naturally defined as the sum of weighted KL divergences. Such an optimization problem is effectively exploited with the *em* algorithm because of the m-flatness. With this notion, we can geometrically interpret the estimation process as a sequence of projections on a probability simplex. Since our *em* process does not need any types of well-tuned numerical optimizers, it is as simple as previously proposed estimation methods for the BT model. Note that our method is applicable not only to the BT model but also to the other models which have similar m-flat data manifolds [7].

In addition, the m-flatness is compatible with a class of divergence called Bregman divergence [4], which shows robustness against small sample sets and outliers in usual statistical inference and includes the KL divergence as a special case. By using the m-flatness, we generalize the estimation method based on the *um* framework [5], which is an *em*-like algorithm based on the Bregman divergence. Unfortunately, our proposed method is not always tractable in calculation because it needs non-linear optimizers. However, a type of Bregman divergence that we call η-divergence has a good property which allows us to estimate model parameters without any numerical optimizers in the same manner as the KL divergence. Our current experimental results show that the sum of weighted η-divergences is flexible enough to improve accuracy of the conventional estimation methods.

In our presentation, we visualize a geometrical interpretation of our *em* estimation procedure based on the weighted KL divergences. Then, we generalize it based on the weighted η-divergences, and experimentally show improvement of estimation results.

References

[1] S. Amari. Information geometry of the EM and *em* algorithms for neural networks. *Newral Networks*, 8(9):1379–1408, 1995.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry.* Oxford University Press, 2000.

[3] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, Dec. 1952.

Topic: learning algorithms
Preference: poster

[4] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[5] Y. Fujimoto and N. Murata. A modified EM algorithm for mixture models based on Bregman divergence. *Annals of the Institute of Statistical Mathematics*, 59(1):3–25, 2007.

[6] T. Hastie and R. Tibshirani. Classification by pairwise coupling. *The Annals of Statistics*, 26(2):451–471, 1998.

[7] H. Hino, Y. Fujimoto, and N. Murata. Item preference parameters from grouped ranking observations. In *Proceedings of the 13-th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2009.

[8] T.-K. Huang, R. C. Weng, and C.-J. Lin. Generalized Bradley-Terry models and multi-class probability estimates. *Journal of Machine Learning Research*, 7:85–115, 2006.

[9] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley-Interscience, 1996.

[10] B. Zadrozny. Reducing multiclass to binary by coupling probability estimates. *Advances in Neural Information Processing Systems*, 14:1041–1048, 2002.