Hyperlink Regression via Bregman Divergence

RIKEN AIP — Akifumi Okuno
RIKEN AIP and Kyoto U. — Hidetoshi Shimodaira

Given data vectors $\{\boldsymbol{x}_i\}_{i=1}^n \subset \mathbb{R}^p$ and an index set $\mathcal{I}_n^{(U)} \subset \{1,2,\ldots,n\}^U$ for $p,n,U \in \mathbb{N}$, we consider an observed *hyperlink weight* $w_{\boldsymbol{i}} \in \mathcal{S}(\subset \mathbb{R})$ representing the association strength among U-*tuple* $\boldsymbol{X}_{\boldsymbol{i}}$, that is an unordered collection of U vectors $\boldsymbol{x}_{i_1}, \boldsymbol{x}_{i_2}, \ldots, \boldsymbol{x}_{i_U}$ indexed by $\boldsymbol{i} = (i_1, i_2, \ldots, i_U)$, for all $\boldsymbol{i} \in \mathcal{I}^{(U)}$.

An example consisting of such U-tuples and their hyperlink weights is co-authorship network, where $\boldsymbol{x}_i$ represents attributes of the researcher $i \in \{1,2,\ldots,n\}$ such as the number of publications in each of journals, and the hyperlink weight $w_{\boldsymbol{i}} \in \mathbb{N}_0$ represents the number of co-authored papers written by all the U researchers indexed by $\boldsymbol{i} \in \mathcal{I}_n^{(U)}$.

In this talk, we propose Bregman hyperlink regression (BHLR), that predicts tuple's hyperlink weight $w_{\boldsymbol{i}}$ through the corresponding tuple $\boldsymbol{X}_{\boldsymbol{i}}$, by learning a user-specified symmetric similarity function $\mu_{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{i}})$. The similarity function can employ non-linear functions such as neural networks, for fully enjoying the high expressive power. BHLR learns the similarity function by minimizing Bregman-divergence (BD)

$$D_\phi(\{w_{\boldsymbol{i}}\}, \{\mu_{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{i}})\}) := \frac{1}{|\mathcal{I}_n^{(U)}|} \sum_{\boldsymbol{i} \in \mathcal{I}_n^{(U)}} d_\phi(w_{\boldsymbol{i}}, \mu_{\boldsymbol{\theta}}(\boldsymbol{X}_{\boldsymbol{i}})),$$

where $d_\phi(a,b) := \phi(a) - \phi(b) - \phi'(b)(a-b)$ and $\phi : \mathrm{dom}(\phi) \to \mathbb{R}$ is a user-specified strictly convex function whose domain $\mathrm{dom}(\phi)$ includes the set $\mathcal{S}$. Then, BHLR encompasses many of existing methods, such as logistic regression ($U=1$), Poisson regression ($U=1$), graph embedding ($U=2$), matrix factorization ($U=2$), tensor factorization ($U \geq 3$), and their variants equipped with arbitrary BD, as special cases.

Regardless of the choice of ϕ and U, we theoretically show that the proposed BHLR is **(P-1) robust against distributional misspecification**, namely, it asymptotically recovers the underlying true conditional expectation of tuple's weight regardless of the conditional distribution of the weights, and **(P-2) computationally tractable**, namely, it is efficiently computed by stochastic algorithms using a proposed minibatch sampling procedure for hyper-relational data. These properties are obtained by generalizing Okuno and Shimodaira (2019). We also conduct some numerical experiments on a real-world dataset.

References

Okuno, A. and Shimodaira, H. (2019). Robust graph embedding with noisy link weights. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 664–673. PMLR.