Efficient Online Bandit Multiclass Learning

Ang Cao, Hangting Cao, Kedong He

1 Problem Statement

Bandit multiclass learning problem is a special case of the general contextual bandit learning, where exactly one of the losses is 0 and all others are 1 in every round [1]. The learner repeatedly classify data into one of the k classes and at each step t, the learner observes an example $\mathbf{x}_t$ and predicts its label $\hat{y}_t$ from k classes. With the deepening of research, one open problem arises: Is there an efficient bandit multiclass learning algorithm that achieves expected regret of $\tilde{O}(\sqrt{T})$[1] [2]? And we are going to study to answer this question.

2 Learning Methods

The machine learning method we are going to study and implement is an efficient online bandit muticlass learning algorithm proposed by Beygelzimer et al. [3]. It's a second-order algorithm with $\tilde{O}(\sqrt{T})$ regret, called Second Order Banditron Algorithm (SOBA), solving the bandit online multiclass problem, especially answering the open question [2] mentioned above.

SOBA makes decisions mainly using γ-greedy strategy. At each iteration t, the classifier has the form $\hat{y}_t = \arg\max_{i\in[k]}(\mathbf{W}_t\mathbf{x}_t)_i$. And during the learning process, it mainly update two model parameters $\boldsymbol{\theta}_t$ and $\mathbf{A}_t$ and $\mathbf{W}_t$ is computed as $\mathbf{A}_{t-1}^{-1}\boldsymbol{\theta}_{t-1}$. The paper [3] also offers an conceptually simpler SOBA algorithm in Appendix. So if having the time, we will also examine that algorithm as a comparison.

3 Data Sets and Evaluations

To evaluate the performance of this theoretical algorithms, we plan to use three popular public datasets from Kakade [4]: SynSep, SynNonSep and Reuters4. SynSep and SynNonSep dataset are synthetic datasets consisting of 10^6 samples in 400 dimensions, which are labeled in 9 classes. Generated from the RCV1 dataset Lewis [5], the Reuters4 dataset has 665,265 labeled examples which contains 47,236 features. Besides, we are also gonna to explore the performance of this algorithm on the Covtype dataset from LibSVM repository[2].

To validate this algorithms, we would implement other previous methods for comparison, including the Banditron [4], the PNewtron as well as the diagonal version of the Newtron algorithm in Hazan [6].

We will study the method together. Hangting Cao and Kedong He will focus more on the implementation of the method and Ang Cao will focus more on the evaluation of the method and the data sets and other methods needed in evaluations. Finally, we will participate in writing the project report together.

[1] $\tilde{O}(\cdot)$ hides logarithmic factors

[2] https://www.csie.ntu.edu.tw/cjlin/libsvmtools/datasets/

References

[1] J. Langford and T. Zhang, *The epoch-greedy algorithm for multi-armed bandits with side information.* In NIPS, pp.817-824, 2008.

[2] J. Abernethy and A. Rakhlin, *An Efficient Bandit Algorithm for $\sqrt{T}$-Regret in Online Multiclass Prediction?.* In COLT, 2009.

[3] A. Beygelzimer, F. Orabona, and C. Zhang, *Efficient Online Bandit Multiclass Learning with $\tilde{O}(\sqrt{T})$ Regret.* Proc. of the International Conference on Machine Learning (ICML), 2017.

[4] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari, *Efficient bandit algorithms for online multiclass prediction.* In ICML, pp.440-447, 2008.

[5] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, *RCV1: A new benchmark collection for text categorization research.* In JMLR, pp.361-397, 2004.

[6] E.Hazan, A.Agarwal, and S.Kale, *Logarithmic regret algorithms for online convex optimization.* Machine Learning, pp.169-192, 2007.