Meta-Tsallis-Entropy Minimization: A New Self-Training Approach for Domain Adaptation on Text Classification

Menglong Lu[1] , **Zhen Huang**[1] , **Zhiliang Tian**[1] , **Yunxiang Zhao**[2] ,
Xuanyu Fei[3] and **Dongsheng Li**[1]
[1]National Key Laboratory of Parallel and Distributed Computing, National University of Defense Technology, China
[2]Beijing Institute of Biotechnology, China
[3]Suiren Information,China
{lumenglong, huangzhen, tianzhiliang, dsli}@nudt.edu.cn, {zhaoyx1993, xuanyufelix}@163.com

Abstract

Text classification is a fundamental task for natural language processing, and adapting text classification models across domains has broad applications. Self-training generates pseudo-examples from the model's predictions and iteratively train on the pseudo-examples, i.e., mininizes the loss on the source domain and the Gibbs entropy on the target domain. However, Gibbs entropy is sensitive to prediction errors, and thus, self-training tends to fail when the domain shift is large. In this paper, we propose Meta-Tsallis Entropy minimization (MTEM), which applies meta-learning algorithm to optimize the instance adaptive Tsallis entropy on the target domain. To reduce the computation cost of MTEM, we propose an approximation technique to approximate the Second-order derivation involved in the meta-learning. To efficiently generate pseudo labels, we propose an annealing sampling mechanism for exploring the model's prediction probability. Theoretically, we prove the convergence of the meta-learning algorithm in MTEM and analyze the effectiveness of MTEM in achieving domain adaptation. Experimentally, MTEM improves the adaptation performance of BERT with an average of 4 percent on the benchmark dataset.

1 Introduction

Text classification plays a crucial role in language understanding and anomaly detection for social media text. With the recent advance of deep learningf [Kipf and Welling, 2017; Devlin *et al.*, 2019], text classification has experienced remarkable progress. Despite the success, existing text classification approaches are vulnerable to domain shift. When transferred to a new domain, a well-performed model undergoes severe performance deterioration. To address such deterioration, domain adaptation, which aims to adapt a model trained on one domain to a new domain, has attracted much attention [Du *et al.*, 2020; Lu *et al.*, 2022].

A direct way to achieve domain adaptation is to build a training set that approximates the distribution of the target domain. For this purpose, self-training [Zou *et al.*, 2019; Liu *et al.*, 2021] uses the unlabeled data from the target domain to bootstrap the model. In specific, self-training first uses the model's prediction to generate pseudo-labels and then uses the pseudo-labeled data to re-train the model. In this process, self-training forces the model to increase its confidence in the confident class, which is a Gibbs entropy minimization process in essence [Lee and others, 2013].

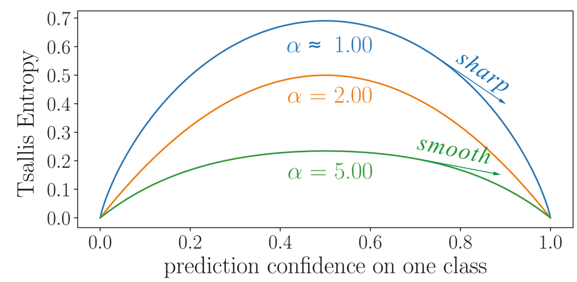

Figure 1: Tsallis entropy curve with respect to different entropy index (i.e., α below the curve).

However, Gibbs entropy minimization is sensitive to prediction errors [Mukherjee and Awadallah, 2020]. To handle the intractable label noise (i.e., prediction errors), data selection strategies are designed to select reliable pseudo labels [McClosky *et al.*, 2006; Reichart and Rappoport, 2007; Rotman and Reichart, 2019]. Among them, many qualified achievements [RoyChowdhury *et al.*, 2019; Shin *et al.*, 2020] are grounded on prior knowledge about the tasks (e.g., the temporal consistency on video [RoyChowdhury *et al.*, 2019]), and thus hard to be applied in text classification tasks. Since the Gibbs entropy minimization process in self-training is to minimize the model's uncertainty on the new domain, [Liu *et al.*, 2021] recently proposes to replace the Gibbs entropy with the Tsallis entropy, which is another effective metric for measuring uncertainty.

Tsallis entropy is a generalization of Gibbs entropy, referring to a set of entropy types controlled by the entropy index. Fig. 1 shows the change of Tsallis entropy with different entropy indexes for binary problems. When the entropy index is small (the resultant entropy curve is sharp),