Breaking the Closed World Assumption in Text Classification

Geli Fei and **Bing Liu**
Department of Computer Science
University of Illinois at Chicago
gfei2@uic.edu, liub@cs.uic.edu

Abstract

Existing research on multiclass text classification mostly makes the *closed world* assumption, which focuses on designing accurate classifiers under the assumption that all test classes are known at training time. A more realistic scenario is to expect unseen classes during testing (*open world*). In this case, the goal is to design a learning system that classifies documents of the known classes into their respective classes and also to reject documents from unknown classes. This problem is called *open* (*world*) *classification*. This paper approaches the problem by reducing the open space risk while balancing the empirical risk. It proposes to use a new learning strategy, called *center-based similarity* (CBS) *space learning* (or *CBS learning*), to provide a novel solution to the problem. Extensive experiments across two datasets show that CBS learning gives promising results on multiclass open text classification compared to state-of-the-art baselines.

1 Introduction

With the rapid growth of online information, text classifiers have become one of the most important tools for people to track and organize information. And the emergence of social media platforms has brought increasing diversity and dynamics to the Web. Many social science researchers rely on the collected online user generated content to carry out research on different social phenomenon. In this case, multiclass text classifiers are widely used to gather information of several topics of interest. However, most existing research on multiclass text classification makes the *closed world* assumption, meaning that all the test classes have been seen in training. However, in a more realistic scenario where people use a multiclass classifier to collect information of several topics from a data source that covers a much broader range of topics, it is normal to break the closed world assumption and to see the arrival of documents from unknown classes that have never been seen in training. In this case, a multiclass classifier should not always assign a document to one of the known classes. Instead, it should identify unknown classes of documents and label them as unknown or reject. This is called *open* (*world*) classification.

More precisely, in the traditional multiclass classification setting, the learner assumes a fixed set of classes $Y = \{C_1, C_2, \ldots, C_m\}$, and the task is to construct a m-class classifier using the training data. The resulting classifier is tested/applied on the data from only the m classes. While in *open* classification, we allow the classifier to predict labels/classes from the set of $C_1, C_2, \ldots, C_m, C_{m+1}$ classes, where the $(m+1)^{th}$ class C_{m+1} represents the unknown which covers documents of all unknown or unseen classes or topics. In other words, every test instance may be predicted to belong to either one of the known classes $y_i \in Y$, or C_{m+1} (unknown).

It is thus not sufficient for a classifier to just return the most likely class label among the m known classes. An option to reject must be provided. An obvious approach to predicting the class label $y \in Y \cup \{C_{m+1}\}$ for an n-dimensional data point $x \in R^n$ is to incorporate a posterior probability estimator $p(y|x)$ and a decision threshold into an existing multiclass learning algorithm (Kwok, 1999; Fumera and Roli, 2002; Huang et al., 2006; Bravo et al., 2008). There are many reasons this technique would not achieve good results in open classification. As we will discuss in the following sections, one of the most important reasons is that the underlying classifier is not robust or is not in-

Proceedings of NAACL-HLT 2016, pages 506–514,
San Diego, California, June 12-17, 2016.