論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	One-class Text Classification with Neural Networks
著者(和文)	HuChenlong
Author(English)	Chenlong Hu
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12070号, 授与年月日:2021年9月24日, 学位の種別:課程博士, 審査員:奥村 学,熊澤 逸夫,中山 実,篠﨑 隆宏,船越 孝太郎
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12070号, Conferred date:2021/9/24, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

専攻： Department of	Information Processing 専攻	申請学位（専攻分野）： Academic Degree Requested	博士 Doctor of (Engineering)
学生氏名： Student's Name	Hu Chenlong	指導教員（主）： Academic Supervisor(main)	Prof. Manabu Okumura
		指導教員（副）： Academic Supervisor(sub)	

要旨（英文 800 語程度）
Thesis Summary（approx.800 English Words ）

In this study, we apply neural networks to the problem of one-class text classification. One-class (text) classification is a special classification problem that aims to learn a model on the basis of training samples only from one class. In this problem, the model is expected to make an accurate description of the "one" class, called *target, positive* or *normal*, and to distinguish the *target* from samples for negative classes during testing. As for one-class text classification, it can be applied to the scenario where a large number of labelled negative samples (e.g., web pages, spam emails) are hard or time-consuming to obtain, while positive training data samples (e.g., ham emails) are available.

Traditionally, document-to-word co-occurrence matrices or hand-crafted features have been commonly used for representing text data in most of the previous work. In recent years, *neural networks* (*deep learning*) have been applied to learn the vector representations of words and documents. Recent studies have developed neural networks for the problem of one-class classification. The approaches include *reconstruction*-based methods such as AuoEncoder and its variants as well as *boundary*-based methods including deep support vector data description (deep SVDD) and one-class neural network (OC-NN).

This dissertation studies one-class text classification with neural networks in two successive parts. The first part addresses the task of acquiring semantic lexicons by using a *boot-strapping* technique. Bootstrapping methods are efficient tools for exploring knowledge, such as semantic lexicon, from unannotated corpora with increasing text size. We focus on *bootstrapping approach to semantic lexicon induction* (*Basilisk*), an effective and commonly used bootstrapping approach for constructing semantic lexicon from text corpora. To start the bootstrapping process of Basilisk, all you need is to input an initial *seed list* with a limited number of seed words. Our motivation is that, since we have only these seed instances, which can be considered *positive* samples from *one* semantic category (e.g., *food* domain), we can try to solve it within the one-class learning framework. Accordingly, we address the following research issues in the first part. 1) Can we formulate the framework of Basilisk as a one-class problem? 2) How do we improve its bootstrapping process with neural networks?

The second part of the thesis investigates a wider range of *boundary*-based methods with neural networks for a more general one-class text classification task. Recently, by incorporating neural feature maps, two boundary-based methods, one-class neural network (OC-NN) and deep support vector data description (deep SVDD), are proposed. Deep SVDD learns a hypersphere model with a center while OC-NN utilizes a linear model with margin. Deep SVDD and OC-NN outperform their kernel-based versions. Nevertheless, target data samples may have distinctive distributions that are located in different regions. Therefore, uni-modal deep SVDD or OC-NN may not be enough to describe the target samples. In the second part, we try to extend the deep SVDD to a version with multiple modes, a *multi-modal deep support vector data description* (mSVDD), where each mode is expected to describe the target samples from a distinctive aspect. Consequently, we address three research issues in the second part: 1) How do we develop uni-modal deep support vector data description into a multi-modal version? 2) What are the relationships with other related models? 3) Due to the unavailability of training samples from negative classes, it is hard for the one-class models to learn effective discrimination information, especially for mSVDD with a multi-layer neural structure. Can we incorporate *negative supervision* to improve mSVDD in some way?

The outline of this thesis is organized as follows.
First, Chapter 1 introduces the background, motivation, and contributions of this thesis. Then, Chapter 2 introduces methods for one-class classification, the relationships between one-class classification and other tasks like clustering, bootstrapping methods, and so on. In Chapter 3, we first describe *Basilisk*, one commonly used framework of bootstrapping lexicon. Then, we show how to formulate this framework as one special one-class problem and describe the proposed methods with AutoEncoder to guide its learning iterations. Finally, the experimental results demonstrate that our proposed methods for guiding the bootstrapping of a semantic lexicon with AutoEncoder can boost overall performance. In Chapter 4, we describe the framework of multi-modal deep support vector data description (mSVDD). We show how to describe the target data with multiple hyperspheres. We also prove a number of propositions to elaborate mSVDD's relationships with other models. In addition, methods for incorporating negative supervision for mSVDD are also introduced. To complete the negative supervision, we also propose a word-level data augmentation method that can construct *pseudo*-negative samples from only given positive target training dataset. Experimental results demonstrate that mSVDD outperforms uni-modal SVDD and that it can get further improvements when negative supervision is incorporated. Finally, Chapter 5 summarizes this dissertation, provides a retrospective discussion on the methods used, and discusses some future directions for future work.