

Zero-Shot Feature Selection via Transferring Supervised Knowledge

Zheng Wang, Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing, China

Qiao Wang, School of Software, Tsinghua University, Beijing, China

Tingzhang Zhao, Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing, China

Chaokun Wang, School of Software, Tsinghua University, Beijing, China

Xiaojun Ye, School of Software, Tsinghua University, Beijing, China

ABSTRACT

Feature selection, an effective technique for dimensionality reduction, plays an important role in many machine learning systems. Supervised knowledge can significantly improve the performance. However, faced with the rapid growth of newly emerging concepts, existing supervised methods might easily suffer from the scarcity and validity of labeled data for training. In this paper, the authors study the problem of zero-shot feature selection (i.e., building a feature selection model that generalizes well to “unseen” concepts with limited training data of “seen” concepts). Specifically, they adopt class-semantic descriptions (i.e., attributes) as supervision for feature selection, so as to utilize the supervised knowledge transferred from the seen concepts. For more reliable discriminative features, they further propose the center-characteristic loss which encourages the selected features to capture the central characteristics of seen concepts. Extensive experiments conducted on various real-world datasets demonstrate the effectiveness of the method.

KEYWORDS

Data Mining, Feature Selection

1. INTRODUCTION

The problem of feature selection (Guyon & Elisseeff, 2003; Wang, Ye, Wang, & Yu, 2019) has been widely investigated due to its importance for pattern recognition and image processing systems. This problem can be formulated as follows: identify an optimal feature subset which provides the best tradeoff between its size and relevance for a given task. The identified features not only provide an effective solution for the task, but also provide a dimensionally-reduced view of the underlying data.

Supervised knowledge (e.g., labels or pair-wise relationships) associated to data is capable of significantly improving the performance of feature selection methods (Chandrashekar, Sahin, & Engineering, 2014). However, it should be noted that existing supervised feature selection methods are facing an enormous challenge — the generation of reliable supervised knowledge cannot catch up with the rapid growth of newly-emerging concepts and multimedia data. In practice, it is costly to annotate sufficient training data for new concepts timely, and meanwhile, impractical to retrain the feature selection model whenever a new concept emerges. As illustrated in Figure 1, traditional methods perform well on the seen concepts which have correct guidance, but they may easily fail on

DOI: 10.4018/IJDWM.2021040101

the unseen concepts which have never been observed, like the newly invented product “quadrotor”. Therefore, the problem of *Zero-shot Feature Selection (ZSFS)*, i.e., building a feature selection model that generalizes well to unseen concepts with limited training data of seen concepts, deserves great attention. However, few studies have considered this problem.

The major challenge in the ZSFS problem is how to deduce the knowledge of unseen concepts from seen concepts. In fact, the primary reason why existing studies fail to handle unseen concepts is that they only consider the discrimination among seen concepts (like the 0/1-form class labels illustrated in Figure 1), such that little knowledge could be deduced for unseen concepts. To address this, as illustrated in Figure 2, we adopt the class-semantic descriptions (i.e., attributes) as supervision for feature selection. This idea is inspired by the recent development of *Zero-shot Learning (ZSL)* (Farhadi, Endres, Hoiem, & Forsyth, 2009) (Guo, Ding, Han, & Gao, 2017) which has demonstrated that the capacity of inferring attributes allows us to describe, compare, or even categorize unseen objects.

An attendant problem is how to identify reliable discriminative features with attributes which might be inaccurate and noisy (Jayaraman & Grauman, 2014). To alleviate this, we further propose a novel loss function (named *center-characteristic loss*) which encourages the selected features to capture the central characteristics of seen concepts. Theoretically, this loss function is a variant of the *center loss* (Wen, Zhang, Li, & Qiao, 2016) which has shown its effectiveness to learn discriminative and generalized features for categorizing unseen objects.

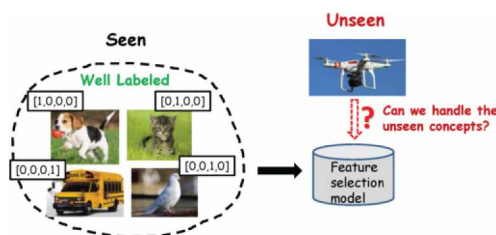
We evaluate the performance of the proposed method on several real-world datasets, including SUN, aPY and CIFAR10. One point should be noted is that the attributes of CIFAR10 are automatically generated from a public Wikipedia text-corpus (Shaoul, 2010) by a well-known NLP tool (Huang, Socher, Manning, & Ng, 2012). The experimental evidence shows that no matter with manually or automatically generated attributes, our method generalizes well to unseen concepts.

We summarize our main contributions as follows:

- We study the problem of ZSFS, i.e., building a feature selection model which generalizes well to unseen concepts with limited training data of seen concepts. To our best knowledge, little work has addressed this problem.
- We propose an efficient strategy to reuse the supervised knowledge of seen concepts. Concretely, feature selection is guided by the seen concept attributes which provide discriminative information about unseen concepts.
- To select more reliable discriminative features, we further encourage the selected features to capture the central characteristics of seen concepts. We formulate this by a novel loss function named center-characteristic loss.
- We conduct extensive experiments on three benchmark datasets to demonstrate the effectiveness of our method.

The rest of this paper is organized as follows. In Section 2, we briefly review some related work on feature selection and zero-shot learning. In Section 3, we formally define the problem studied

Figure 1. The Zero-shot Feature Selection Problem



18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/zero-shot-feature-selection-via-transferring-supervised-knowledge/276762

Related Content

Creating Strategic Information for Organizations with Structured Text

Roberto Penteado and Eric Boutin (2008). *Emerging Technologies of Text Mining: Techniques and Applications* (pp. 34-53).

www.irma-international.org/chapter/creating-strategic-information-organizations-structured/10175

Electronic Records Management - An Old Solution to a New Problem: Governments Providing Usable Information to Stakeholders

Chinh Nguyen, Rosemary Stockdale, Helana Scheepers and Jason Sargent (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 2249-2274).

www.irma-international.org/chapter/electronic-records-management---an-old-solution-to-a-new-problem/150264

Structuring and Facilitating Online Learning through Learning/Course Management Systems

Shalin Hai-Jew (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 1358-1375).

www.irma-international.org/chapter/structuring-facilitating-online-learning-through/73500

Ranking of Evaluation Targets Based on Complex Sequential Data

Shigeaki Sakurai (2017). *International Journal of Data Warehousing and Mining* (pp. 19-32).

www.irma-international.org/article/ranking-of-evaluation-targets-based-on-complex-sequential-data/188488

The Application of Data Mining for Drought Monitoring and Prediction

Tsegaye Tadesse, Brian Wardlaw and Michael J. Hayes (2009). *Data Mining Applications for Empowering Knowledge Societies* (pp. 278-289).

www.irma-international.org/chapter/application-data-mining-drought-monitoring/7557