DISCRIMINATIVE MANIFOLD EMBEDDING WITH IMPRECISE, UNCERTAIN, AND AMBIGUOUS DATA

By

CONNOR H. MCCURLEY

A ORAL QUALIFYING EXAM PROPOSAL PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2019

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

1.0.1 Causes of ambiguity/ imprecision in groundtruth

Target detection is a paramount problem in many remote sensing applications. Target detection can be formulated as a two-class classification problem where samples belonging to a class of interest are discriminated from a global background distribution. For example, Alternatively, the problem can be posed as distinguishing a single class from all other classes presented in the training data. Traditional supervised learning approaches require highly precise groundtruth to guide algorithmic training. However, acquiring large quantities of accurately labeled training data can be expensive both in terms of time and resources, and in some cases, may even be infeasible to obtain. Furthermore, label imprecision and ambiguity are often presented from multiple avenues. Additionally, many applications such as medical diagnosis or wildlife identification require domain experts to provide accurate data labels. However, label subjectivity is even introduced in this case, as there might not always be agreement between expert annotators. . For example, GPS

Techniques which can address these forms of label ambiguity while achieving comparable or better target detection than standard supervised methods can greatly ease the burdens associated with many remote sensing tasks.

Images showing examples of weak labels (point, scribble, in-image, region/ weak bounding box)

1.0.2 What is Multiple Instance Learning/ How it Addresses Uncertainty

Learning from uncertain, imprecise and ambiguous data has been an active area of research since the late 1990s and is known as multiple instance learning (MIL) [1]. Supervised learning assumes that each training sample is paired with a corresponding classification label. In multiple instance learning, however, the label of each sample is not necessarily known. Instead, MIL approaches learn from groups of data points called *bags*, and each bag concept is paired with a label [2]. Under the two-class classification scenario the bags are labeled as *negative* if all data points (or *instances*) are known to belong to the background class (not the class of interest). While the actual number of positive and negative instances may be unknown, bags are labeled *positive* if *at least one* instance is known to belong to the target class (also called a "true positive") [3]. The goal of learning under the MIL framework is to train a model which can classify a bag as positive or negative (bag-level classification) or to predict the class labels of individual instances (instance-level classification). This problem formulation fits many remote sensing scenarios and is thus an important area of investigation [4].

1.0.3 Statement of Problems

Multiple instance learning approaches in the literature can be broadly generalized into two categories: learning a single or set of concepts which effectively describe the positive and/or negative class or training a classifier to discriminate bags or individual instances. Existing approaches make two assumptions: (1) MIL methods assume that target instances are already separable in the feature space or that a signature(s) can be learned which represents the target class well, while poorly representing the entirety of the negative class, regardless of potential distribution overlap. (2) These methods assume high data dimensionality as a non-issue, or fail to address high-dimensionality.

1.0.4 What is Dimensionality Reduction/ Manifold Learning/ Feature Embedding? Better feature representations facilitate classification.

Effects of high-dimensionality on classification and relation to obtaining training data (We need more, but it is a cascading effect since obtaining accurate labels is difficult.)

- What is manifold learning
- Why reduce dimensionality?
- Examples of uninformative and redundant (correlated) features (crabs, etc)

1.0.5 Existing Approaches

Recently, methods for representation learning have gained in popularity because they typically result in high levels of accuracy. Some of these methods learn features in a supervised manner to obtain a more discriminative representation. However, this learning cannot be done directly in MIL because of the uncertainty on the labels. Thus, *adaptation of discriminative feature learning methods would be beneficial to MIL* [5].

Feature embedding and dimensionality reduction under the multiple instance learning framework has scarcely been explored. The first true experimentation was performed in [6]. In this work, Sun et al. showed that Principal Component Analysis (PCA) failed to incorporate bag-level label information and thus provided poor separation between positive and negative bags. Additionally, Linear Discriminant Analysis (LDA) was used to project bags into a latent space which maximized between-bag separation, while minimizing within-bag dissimilarity. However, LDA often mixed the latent bag representations due to the uncertainty of negative sample distributions in the positive bags. To address these issues, Sun proposed Multiple Instance Dimensionality Reduction (MIDR) which optimized an objective through gradient descent to discover sparse, orthogonal projection vectors into the latent space. Their approach relied on fitting a distribution of negative instances to which each instances' probability was evaluated. This approach was later extended in [7] in attempt to improve sparsity. Both of these methods rely on fitting a distribution of negative instances and determining positive instances as samples which have low likelihood of belonging to the distribution. However, these approaches fail when the the target and background instances are similar. It would, therefore, be beneficial to first transform the instance representations into more discriminable forms. Most existing approaches in the literature extend LDA to distinguish between positive and negative bags [8, 7]. These methods typically rely on costly optimization procedures to maximize an objective for orthogonal and sparse projection vectors. The work in [9] investigated metric learning on sets of data (where each bag was a set) to learn an appropriate similarity metric to compare bags. They do not go as far as to discriminate positive instances within the positive bags, nor do they propose positive target concepts. In fact, virtually all MIL DR methods in the literature are applied, solely, to predict bag-level labels. While bag-level label prediction is useful in many remote sensing applications, such as region of interest (ROI) proposal for anomaly detection, they limit the the ability of learners to classify on the pixel or sample level. Xu et al. proposed an importance term to weight samples believed to be true target exemplars, as those instances are more important in determining the bag-level label [9]. However, to author's knowledge, no work has been done to investigate instance-level discrimination through manifold learning/ dimensionality reduction. *With this in mind, the goal of this work is, using only bag-level labels, to find discriminative instance embeddings which allow for accurate sample-level class discrimination.*

1.0.6 Proposition and Research Questions

To address these points, I propose the following. During this project, techniques will be explored for use in instance-level classification given uncertain and imprecise groundtruth. These methods will be developed as universal approaches for discriminative manifold/feature representation learning and dimensionality reduction and will be evaluated on a variety of sensor modalities, including: mid-wave IR, visible, hyperspectral and multispectral imagery, LiDAR and more. *The aim of this project is to develop dimensionality reduction methods which promote class discriminability and are simultaneously capable of addressing uncertainty and imprecision in training data groundtruth.* The motivating idea is to facilitate instance/concept proposition by increasing instance discriminability under the constraints of multiple instance learning. Roughly, the following research questions will be addressed during the scope of this project:

1. Supervised and semi-supervised manifold learning has proven effective at discovering low-dimensional data representations which provide adequate class separation in the latent space. However, only a handful of manifold learning procedures consider data which is weakly or ambiguously labeled. To address this gap in the literature, a method for weakly-supervised manifold learning will be developed. How does this method of manifold construction compare to state-of-the-art manifold learning techniques as well as alternative MI dimensionality reduction methodologies for instance-level label prediction?

2. Metric embedding has been shown to promote class separability through dimensionality reduction. Without instance-level labels, a method for ranking loss embedding will be developed in conjunction with Objective 1 to improve class separation of individual instances. Additionally, a procedure to select the most influential examples for training will be developed.

3. Do the proposed methods facilitate concept learning/selection? Using alternative, SOA MIL approaches, are the selected target instances/concepts more discriminable with the proposed methods than without? How do the proposed methods compare to the alternatives in terms of representation dimensionality, computational complexity, and promotion of discriminability?

1.0.7 Datasets

LFW Faces in the Wild

CHAPTER 2
BACKGROUND

This chapter provides a literature review on Manifold Learning, including classic approaches, supervised and semi-supervised methods and uses of manifolds for functional regularization. A review of the Multiple Instance Learning framework for learning from weak and ambiguous annotations is provided. Additionally, this chapter reviews the existing literature in classification over graphs, focusing heavily on the utilization of graph convolutional neural networks. A brief overview of competency aware machine learning methods is also included. Reviews describe basic terminology and definitions. Foundational approaches are described and advances are addressed.

2.1 Multiple Instance Learning

2.1.1 Overview and short description

2.1.2 Multiple Instance Concept Learning

2.1.3 Multiple Instance Learning via Embedded Instance Selection

2.1.4 Multiple Instance Dictionary Learning

2.1.5 Multiple Instance Learning on Graphs/Manifolds

2.2 Metric Embedding

2.2.1 Overview of Metric Embedding

2.2.1.1 Metric Learning

2.2.2 Point-wise Loss

2.2.3 Contrastive Loss

Definition of Contrastive Loss:

2.2.3.1 Siamese Networks

2.2.4 Triplet Loss

Definition of Triplet Loss:

Triplet loss was extended in [10] to simultaneously optimize against N negative classes.

2.2.4.1 Large-Margin K-Nearest Neighbors (LMNN)

2.2.4.2 FaceNet

FaceNet is a convolutional neural network which learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity [11].

$$||f(x_i^a) - f(x_i^p)||_2^2 + \alpha < ||f(x_i^a) - f(x_i^n)||_2^2, \quad \forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in \mathcal{T} \tag{2-1}$$

$$\mathcal{L} = ||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) - f(x_i^n)||_2^2 + \alpha \tag{2-2}$$

2.2.4.3 Siamese Neural Networks

2.2.5 Manifold Regularization

2.2.6 Multiple Instance Metric Learning

2.3 Manifold Learning

2.3.1 Overview and short description

2.3.2 Linear Methods

2.3.2.1 Principal Component Analysis (PCA) (Kernel PCA)

2.3.2.2 Multidimensional Scaling (MDS)

2.3.3 Nonlinear Methods

2.3.3.1 Graph-based Methods

Nonlinear dimensionality reduction methods typically rely on the use of adjacency graphs. These graphs represent data structure pooled from local neighborhoods of samples. An overview of computational graphs, as well as the two most prominent methods for graph construction in manifold learning are presented.

2.3.3.2 Graphs

2.3.3.3 K-Nearest Neighbor Graph

2.3.3.4 ϵ Neighborhood Graph

2.3.3.5 Geodesic Distance Approximation

2.3.3.6 Isomap

2.3.3.7 Locally Linear Embedding (LLE)

2.3.3.8 Laplacian Eigenmaps

2.3.3.9 Hessian Eigenmaps

2.3.3.10 Diffusion Maps

2.3.3.11 Sammon Mapping

2.3.3.12 Latent Variable Models

2.3.3.13 Competitive Hebbian Learning

2.3.3.14 Deep Learning

2.3.4 Supervised and Semi-Supervised Approaches

CHAPTER 3
PROBLEM DESCRIPTION

CHAPTER 4
EXPERIMENTAL DESIGN

CHAPTER 5
PRELIMINARY WORK

CHAPTER 6
FUTURE TASKS

CHAPTER 7
CONCLUSIONS

REFERENCES

[1] J. Bocinsky, "Learning multiple target concepts from uncertain, ambiguous data using the adaptive cosine estimator and spectral match filter," Master's thesis, Univ. of Florida, Gainesville, FL, May 2019.

[2] M. Cook, "Task driven extended functions of multiple instances (td-efumi)," Master's thesis, Univ. of Missouri, Columbia, MO, 2015.

[3] A. Zare, C. Jiao, and T. C. Glenn, "Multiple instance hyperspectral target characterization," *CoRR*, vol. abs/1606.06354, 2016. [Online]. Available: http://arxiv.org/abs/1606.06354

[4] X. Du, "Multiple instance choquet integral for multiresolution sensor fusion," Ph.D. dissertation, Univ. of Missouri, Columbia, MO, Dec. 2017.

[5] M. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *CoRR*, vol. abs/1612.03365, 2016. [Online]. Available: http://arxiv.org/abs/1612.03365

[6] Y.-Y. Sun, M. K. Ng, and Z.-H. Shou, "Multi-instance dimensionality reduction," in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, ser. AAAI'10. AAAI Press, 2010, pp. 587–592. [Online]. Available: http://dl.acm.org/citation.cfm?id=2898607.2898702

[7] H. Zhu, L.-Z. liao, and M. K. Ng, "Multi-instance dimensionality reduction via sparsity and orthogonality," *Neural Comput.*, vol. 30, no. 12, pp. 3281–3308, dec 2018. [Online]. Available: https://doi.org/10.1162/neco_a_01140

[8] J. Chai, X. Ding, H. Chen, and T. Li, "Multiple-instance discriminant analysis," *Pattern Recognition*, vol. 47, no. 7, pp. 2517 – 2531, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320314000387

[9] Y. Xu, W. Ping, and A. T. Campbell.

[10] K. Sohn, "Improved deep metric learning with multi-class n-pair loss objective," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 1857–1865.

[11] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *CoRR*, vol. abs/1503.03832, 2015. [Online]. Available: http://arxiv.org/abs/1503.03832