DISCRIMINATIVE MANIFOLD EMBEDDING WITH IMPRECISE 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

page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Feature embedding and dimensionality reduction under the multiple instance learning framework has scarcely been explored. The first true experimentation was performed in [1]. 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. The authors 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 by [2] in attempt to improve sparsity. Most existing approaches in the literature extend LDA to distinguish between positive and negative bags [3, 2]. These methods typically rely on costly optimization procedures to maximize an objective for orthogonal and sparse projection vectors. Additionally, the objective of these approaches is, when given a new bag at test, to provide a bag-level label. The work in [4] 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. 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. However, to other'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 to find discriminative instance embeddings which allow for accurate sample-level class discrimination from weakly-labeled data.*

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 learning and dimensionality reduction and will be evaluated on a variety of sensor modalities, including: mid-wave IR, visible, hyper-spectral and multi-spectral 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. 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 bag-level label prediction?

2. Can metric embedding or metric learning be used to improve discriminability among the bags/instances? Should ranking loss by incorporate within individual bags or across them? How is this loss incorporated to update the embedding function(s)?

3. Multiple instance classification procedures such as Multiple Instance Adaptive Cosine/Coherence Estimator, Multiple Instance Random Forest and Multiple Instance Support Vector Machines have shown considerable success in anomaly detection and positive instance classification/concept representation. Do the methods developed in the previous objectives obtain comparable sample-level detection/segmentation results to these alternative approaches? Or does the discriminative dimensionality reduction technique developed facilitate classification in standard, supervised approaches?

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 Metric Embedding

2.1.1 Overview of Metric Embedding

2.1.1.1 Metric Learning

2.1.2 Point-wise Loss

2.1.3 Contrastive Loss

2.1.3.1 Siamese Networks

2.1.4 Triplet Loss

2.1.4.1 Large-Margin K-Nearest Neighbors (LMNN)

2.1.4.2 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 [5].

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

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

2.1.5 Manifold Regularization

2.2 Multiple Instance Learning

2.2.1 Overview and short description

2.2.2 Multiple Instance Concept Learning

2.2.3 Multiple Instance Classification

2.2.4 Multiple Instance Regression

2.2.5 Multiple Instance Learning on Graphs/Manifolds

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

2.4 Competency Aware Overview

CHAPTER 3
PROBLEM DESCRIPTION

CHAPTER 4
EXPERIMENTAL DESIGN

CHAPTER 5
PRELIMINARY WORK

CHAPTER 6
FUTURE TASKS

CHAPTER 7
CONCLUSIONS

REFERENCES

[1] 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

[2] 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

[3] 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

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

[5] 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