Year 7

Final Project Report

2018-2019

Center for Visual & Decision Informatics

Project 7a.028-TAU -Co-Botics: Intelligent Cooperating Robots and Humans – Phase II

A National Science Foundation
University Cooperative Research Center

Project 7a.028.TAU – Co-Botics – Phase II

Contents

Personnel

- Moncef Gabbouj (PI)
- Alexandros Iosifidis (Co-PI)
- Jenni Raitoharju (Co-PI)
- Fahad Sohrab (Student/Researcher)
- Kateryna Chumachenko (Student/Assistant Researcher)

Sponsoring IAB member's first and last name (company name):

- **CA Software (Technologies)**
  - Steven Greenspan (CA Technologies)
  - Peter Matthews (CA Technologies)
- **Tieto**
  - Matti Vakkuri (Tieto)

Executive Summary/Abstract

In systems of collaborative robotics, the success of the decision-making process is often based on the ability to efficiently utilize the data coming from multiple sources. This is due to the fact that various sensors are usually utilized in such environment, resulting in the need of non-trivial combination of different signals (visual, audial, etc.). Such problems are referred to as multi-modal or multi-view problems, and the success of the solution generally relies on finding a common representation space for different data modalities. Besides, an important research direction in the context of human-machine interaction lies in the detection of unexpected (anomaly) events. Such problems can be solved by means of one-class classification. Another obstacle comes from the requirement of a fast speed of the developed algorithms.

In this project, we proposed several solutions to the problems of anomaly detection and fast analysis of multi-modal data in the context of collaborative robotics environment.

Goals and Objectives

The main objective of the project is the creation of novel methodologies and decision-making strategies for multi-view and multi-modal data analysis in the human-machine systems. Algorithms should be suitable for real-world data, that is not linear or unimodally distributed. In addition, methods for performing anomaly detection, detection of unexpected events in human-robot environment should be developed. The main limitation lies in the fact that since anomaly events are not common, there is generally a lack of data for representation of such events. Therefore, the methods where one class is of more interest than others and the lack of data is present should be developed. The developed algorithms should be efficient and have low computational requirements.

Differences from Current State of Art

Multiple multi-modal data analysis methods have been proposed in the past, such as Multi-view Discriminant Analysis [1] and Multi-view Canonical Correlation Analysis [2]. However, most of them rely on the assumption of unimodal distribution of data within each view, and availability of a large amount of data of each class.

Multiple approaches exist for solving anomaly detection problems by means of one-class classification, and SVDD is one of the widely used ones. However, significantly less effort has been put into solving one-class classification in multi-modal setup. To the best of our knowledge, there has not been performed any work on multi-modal subspace learning for one-class classification, except our newly proposed MS-SVDD.

Most multi-view methods rely on the assumption of unimodality of data within each view, while such scenario is oversimplified and uncommon in real-world data. MvSDA relaxes this limitation, assuming multi-modality of data in each separate view, hence targeting the real-world cases. Another difference is a proposed way to solve the problem that does not rely on eigen-decomposition, like most of the methods, but on a fast and efficient process.

Methods and Datasets

We developed two multi-modal methods: Multimodal Subspace Support Vector Data Description and Multi-view Subclass Discriminant Analysis.

Multimodal Subspace Support Vector Data Descriptions performs one-class classification on the data fused from multiple sensors. The proposed method iteratively transforms the data from original feature space of each modality to a new joint feature space, which is shared by all modalities and suitable for one-class classification. Different regularization terms and decision strategies are also proposed along with linear and non-linear formulation of the method. We show that the proposed Multimodal Subspace Support Vector Data Description outperforms all the methods using data from a single modality and performs better or equally well than the methods fusing data from all modalities.

Multi-view Subclass Discriminant Analysis allows to perform efficient analysis of data represented by multi-view data, where each view contains multi-modal data. The method is based on the representation of data of each class with multiple subclasses and subsequently finding a common projection space for all the views, where data of different classes lies far from each other, while remaining compact. Representation of data with several subclasses allows to efficiently apply the method to the problems, where one class contains subsets of data that are somewhat different. One example of such situation can be the face recognition problem, where part of photos of one person were collected in the dark environment, and part – in a light environment. Besides, we have developed an extension that allows to significantly improve the training speed compared to other methods.

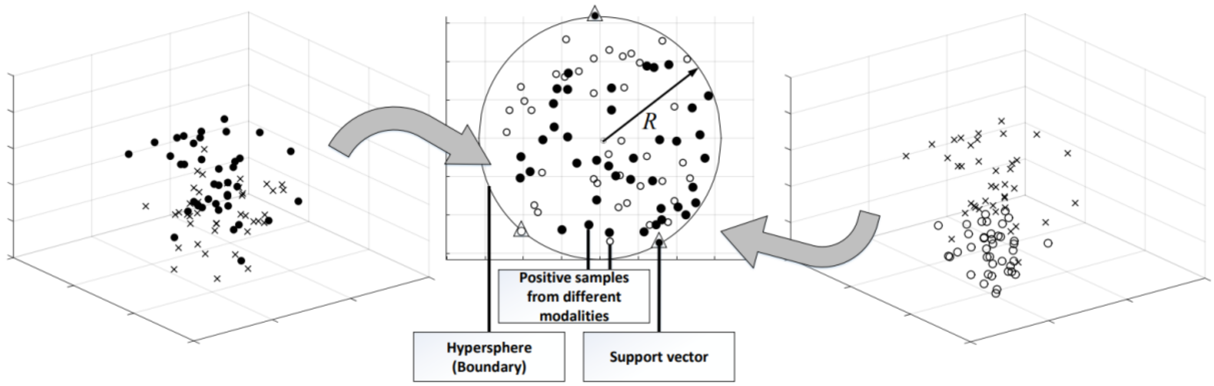

Figure 1 : Depiction of proposed MS-SVDD: Data from two modalities in their corresponding feature space is mapped to a common subspace, where positive class data is enclosed inside a (hyper)sphere.

The MS-SVDD method is evaluated on 2 datasets, namely, robot execution failures [3] and SPECTF heart dataset [4]. The second method is evaluated on nine single-view and multi-view datasets, among which are the robot execution failures dataset [3], human action recognition datasets [5, 6], and object recognition datasets [7, 8].

Results

Comparing to other one-class classification methods, the linear Multimodal Subspace Support Vector Data Description outperforms all the methods using data from a single modality and performs better or equally well than the methods fusing data from all modalities.

Experiments on an extensive range of datasets show that Multi-view Subclass Discriminant Analysis outperforms the existing methods in terms of accuracy in both kernel and linear cases, while always taking significantly less time to train. Extensive quantitative result can be found in [9, 10].

Functionality of Innovation(s)

The non-linear MS-SVDD using the kernel trick requires computing the eigen-decomposition at every iteration. This is computationally expensive and, therefore, we propose an alternative non-linear approach using Non-linear Projection Trick (NPT). In NPT, a non-linear mapping is applied only at the beginning of the process, while the optimization follows the linear MS-SVDD.

Conclusions and Recommendations

In this project, we proposed two methods for analysis of multi-view and multi-modal data. The proposed methods allow performing analysis on data coming from different sources and not following unimodal distributions. Both methods show competitive performance, often outperforming the existing methods, and efficient solutions have been proposed as well. The proposed one-class classification method can be applied to a wide range of tasks relevant to the human-machine environment, where a specific class is of more interest than others, or when data of one class only is available. These include anomaly detection, face recognition, speaker verification, etc. The proposed MvSDA method can be applied to any classification task, where the data is not unimodal, and the result can be obtained in fast and efficient manner. In the future, we plan to explore the effects of different boundary shapes in MS-SVDD and evaluate different model-based decision strategies.

Impact and Uses/Benefits

We have submitted two papers: "Multimodal Subspace Support Vector Data Description" [9] to "Speed-up and multi-view extensions to Subclass Discriminant Analysis" [10] to Pattern Recognition for this project.

List of References

[1] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, 2015. Multi-view discriminant analysis. IEEE transactions on pattern analysis and machine intelligence, 38(1), pp.188-194.

[2] J. Rupnik, and J. Shawe-Taylor, 2010, October. Multi-view canonical correlation analysis. In Conference on Data Mining and Data Warehouses (SiKDD 2010) (pp. 1-4).

[3] M. Afifi and A. Abdelhamed. Integration and learning in supervision of flexible assembly systems. IEEE Transactions on Robotics and Automation 12 (1996), 202–219

[4] L. A. Kurgan, K. J. Cios, R. Tadeusiewicz, M. Ogiela, and L. S. Goodenday, "Knowledge discovery approach to automated cardiac spect diagnosis," Artificial intelligence in medicine, vol. 23, no. 2, pp. 149– 169, 2001.

[5] D. Anguita, A. Ghio, L. Oneto, X. Parra and J. Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges, Belgium, 2013.

[6] S. Torres, D. Ranasinghe, Q. Shi and A. Sample. Sensor enabled wearable RFID technology for mitigating the risk of falls near beds. IEEE International Conference on RFID. 2013, 191–198

[7] L. Fei-Fei, R. Fergus and P. Perona. One-Shot learning of object categories. IEEE Transactions on Pattern Recognition and Machine Intelligence 8 (2006), 594–611

[8] T. Chua, J. Tang, R. Hong, H. Li, Z. Luo and Y. Zheng. NUS-WIDE: A real-world web image database from National University of Singapore. ACM International Conference on Image and Video Retrieval. Greece, 2009.

[9] F. Sohrab, J. Raitoharju, A. Iosifidis, and M. Gabbouj, 2019. Multimodal Subspace Support Vector Data Description. arXiv preprint arXiv:1904.07698.

[10] K. Chumachenko, J. Raitoharju, A. Iosifidis, and M. Gabbouj, 2019. Speed-up and multi-view extensions to Subclass Discriminant Analysis. arXiv preprint arXiv:1905.00794.