**SIKS-DKE Symposium 2018**

Speaker: Dr. Siamak Mehrkanoon (KU Leuven)

 Date: Tuesday, February 20th, 2018

Time: 15:30-16:30

Room: 0.015

Location: Maastricht University, Department of Data Science & Knowledge Engineering, Bouillonstraat 8-10

**Abstract:**

In many applications ranging from machine learning to data mining, obtaining the labeled samples is costly and time consuming. On the other hand with the recent development of information technologies one can easily encounter a huge amount of unlabeled data coming from the web, smartphones, satellites etc. In these situations, one may consider to design an algorithm that can learn from both labeled and unlabeled data. Starting from the Kernel Spectral Clustering (KSC)core formulation, which is an unsupervised algorithm, extensions towards integration of the available side information and devising a semi-supervised algorithm are a scope of the first part of the talk. In particular, the multi-class semi-supervised learning model(MSS-KSC) will be introduced that can address both semi-supervised classification and clustering. The labeled data points are incorporated into the KSC formulation at the primal level via adding a regularization term. This converts the solution of KSC from an eigenvalue problem to a system of linear equations in the dual. The algorithm realizes a low dimensional embedding for discovering micro clusters.

 Though the portion of the labeled instances is small, one can easily encounter a Huge amount of the unlabelled data points. Therefore, in order to make the model scalable to large scale data two approaches are proposed, Fixed-size and reduced kernel MSS-KSC (FS-MSS-KSC and RD-MSS-KSC). The former relies on the Nyström method for approximating the feature map and solves the problem in the primal whereas the latter uses a reduced kernel technique and solves the problem in the dual. Both approaches possess the out-of-sample extension property to unseen data points.

 In today’s applications, evolving data streams are ubiquitous. Due to the complex underlying dynamics and non-stationary behavior of real-life data, the demand for adaptive learning mechanisms is increasing. An incremental multi-class semi- supervised kernel spectral clustering (I-MSS-KSC) algorithm is proposed for an on-line clustering/classification of time-evolving data. It uses the available side information to continuously adapt the initial MSS-KSC model and learn the underlying complex dynamics of the data stream. The performance of the proposed method is demonstrated on synthetic data sets and real-life videos. Furthermore, for the video segmentation tasks, Kalman filtering is used to provide the labels for the objects in motion and thereby regularizing the solution of I-MSS-KSC.

Manual labeling of sufficient training data for diverse application domains is a costly, laborious task and often prohibitive. Therefore, designing models that can leverage rich labeled data in one domain and be applicable to a different but related domain is highly desirable. In particular, domain adaptation or transfer learning algorithms seek to generalize a model trained in a source domain (training data) to a new target domain (test data). The most common underlying assumption of  many machine learning models is that both training and test data exhibit the same distribution or the same feature domains. However, in many real life problems, there is a distributional, feature space and/or dimension mismatch between the two domains or the statistical properties of the data evolve in time. Here a brief overview of the Regularized Semi-Paired Kernel Canonical Correlation Analysis (RSP-KCCA) formulation for learning a latent space for the domain adaptation problem will be provided. The optimization problem is formulated in the primal dual LS-SVM setting where side information can be readily incorporated through regularization terms. The proposed model learns a joint representation of the data set across different domains by solving a generalized eigenvalue problem or linear system of equations in the dual. The approach is naturally equipped with out-of-sample extension property which plays an important role for model selection.

 See also [https://project.dke.maastrichtuniversity.nl/colloquia/](https://solismail.uu.nl/owa/redir.aspx?C=slrvPsiq7gXsTVby_9xutl1YgZtvt3yiVAuZ3LxDu_qAWBcQTHjVCA..&URL=https%3a%2f%2fproject.dke.maastrichtuniversity.nl%2fcolloquia%2f)   for information on coming and previous talks.