**人工智能前沿研讨会**

**内容简介**

1. **Gerard de Melo：Towards Deep Learning with Heterogeneous Supervision**

Although unprecedented amounts of Big Data are now available, in most machine learning problems it is still not yet clear how to exploit it to achieve additional gains. In this talk, I will discuss the need to move towards learning with heterogeneous supervision instead of the classic setting of regular supervised learning. In other words, the new goal should be to learn from multiple, possibly quite heterogeneous kinds of data. This is particularly true in deep learning and representation learning. While there is still no general one-size-fits-all solution to this, I will give a few examples of how heterogeneous supervision can help in tasks related to natural language semantics and Web Mining.

1. **Kaizhu Huang: A Unified Gradient Regularization theory for Adversarial Examples**

Adversarial examples are augmented data points generated by imperceptible perturbation of input samples. They have recently drawn much attention with the machine learning and data mining community. Being difficult to distinguish from real examples, such adversarial examples could change the prediction of many of the best learning models including the state-of-the-art deep learning models. Recent attempts have been made to build robust models that take into account adversarial examples. However, these methods can either lead to performance drops or lack mathematical motivations. In this paper, we propose a unified framework to build robust machine learning models against adversarial examples. More specifically, using the unified framework, we develop a family of gradient regularization methods that effectively penalize the gradient of loss function w.r.t. inputs. Our proposed framework is appealing in that it offers a unified view to deal with adversarial examples. It incorporates another recently-proposed perturbation based approach as a special case. In addition, we present some visual effects that reveals semantic meaning in those perturbations, and thus support our regularization method and provide another explanation for generalizability of adversarial examples. By applying this technique to Maxout networks, we conduct a series of experiments and achieve encouraging results on two benchmark datasets. In particular,we attain the best accuracy on MNIST data (without data augmentation) and competitive performance on CIFAR-10 data.

1. **Zenglin Xu: Large Scale Nonparametric Tensor Analysis**

Tensor factorization is an important approach to multiway data analysis. Many popular tensor factorization approaches—such as the Tucker decomposition and CANDECOMP/PARAFAC (CP)—amount to multi-linear factorization. They are insufficient to model (i) complex interactions between data entities, (ii) various data types (e.g., missing data and binary data), and (iii) noisy observations and outliers. In this talk, I will introduce tensor-variate latent nonparametric Bayesian models for multiway data analysis models. We name these models InfTucker, which essentially conduct Tucker decomposition in an infinite feature space. To further make these models scalable to large data, we will also introduce various extensions, which take advantages of distributed computing techniques such as MapReduce (e.g., Hadoop and Spark), online learning, and data sampling. Finally, I will show some experimental results in real world applications, such as network modeling, access log analysis, and click through rate prediction.

1. **Yingming Li: Learning with Marginalized Corrupted Features and Labels Together**

Tagging has become increasingly important in many real-world applications noticeably including web appli- cations, such as web blogs and resource sharing sys- tems. Despite this importance, tagging methods often face difficult challenges such as limited training sam- ples and incomplete labels, which usually lead to de- generated performance on tag prediction. To improve the generalization performance, in this paper, we pro- pose Regularized Marginalized Cross-View learning (RMCV) by jointly modeling on attribute noise and label noise. In more details, the proposed model con- structs infinite training examples with attribute noises from known exponential-family distributions and ex- ploits label noise via marginalized denoising autoen- coder. Therefore, the model benefits from its robustness and alleviates the problem of tag sparsity. While RMCV is a general method for learning tagging, in the evalua- tions we focus on the specific application of multi-label text tagging. Extensive evaluations on three benchmark data sets demonstrate that RMCV outstands with a su- perior performance in comparison with state-of-the-art methods.

1. **Yafang Wang: Methods and Tools for Temporal Knowledge Harvesting**

The world is dynamic: periodic events like sports competitions need to be interpreted with their respective time points, and facts such as coaching a sports team, holding political or business positions, and even marriages do not hold forever and should be augmented by their respective timespans. We describe how we gather temporal facts from semi-structured and free-text sources.