

数学与系统科学研究院

计算数学所学术报告

报告人: **Prof. Xuecheng Tai**

(*Department of Mathematics, University of Bergen, Norway*)

报告题目:

**Nonlocal graph TV for variational
semi-supervised learning and
clustering and fast algorithms**

邀请人: 谢和虎 研究员

报告时间: 2016 年 7 月 27 日 (周三)

下午 16:00~17:00

报告地点: 数学院南楼二层

202 会议室

Abstract:

In this talk, we first show the essential ideas to use graph cut for data clustering. Then we present some recent developed continuous max-flow model and algorithms. Afterwards, we show the application of these models and algorithms for machine learning and data clustering.

One essential ingredient is to use nonlocal total variation and Laplacian to formulate a the problem as a graph cut. We propose an algorithm for semi-supervised clustering of high-dimensional data. The data points are modeled as vertices of a weighted graph, and the labeling function defined on each vertex takes values from the unit simplex, which can be interpreted as the probability of belonging to each class. The algorithm is proposed as a minimization of a convex functional of the labeling function. There are two versions of the models. The first one combines the Rayleigh quotient for the graph Laplacian and a region-force term, and the second one only replaces the Rayleigh quotient with the total variation of the labeling function. The region-force term is calculated by the affinity between each vertex and the training samples, characterizing the conditional probability of each vertex belonging to each class. The numerical methods for solving these two versions of the proposed algorithm are presented, and both are tested on several benchmark data sets such as handwritten digits (MNIST) and moons data. Experiments indicate that the correction rates and the computational speed are competitive with the state-of-the-art in multi-class semi-supervised clustering algorithms. Especially, the new models produces substantial improvements of the classification accuracy in comparison with the corresponding models without the regional force in cases that the sampling rate is relatively low.

This is joint work with Ke Yin and Stanley Osher.

欢迎大家参加！