Announcing the Final Examination of Yinjie Huang for the degree of Doctor of Philosophy

Time & Location: June 30, 2016 at 10:00 AM in HEC 113

Title: Content-based Information Retrieval via Nearest Neighbor Search

Content based information retrieval (CBIR) has attracted lots of interest in the past few years. When given a search query, the search engine will compare the query with all the stored information in database through nearest neighbor search. Finally, the system will return the most similar items. We contribute to CBIR research community in the following two aspects: firstly, Distance metric learning (DML) is studied to improve retrieval accuracy of nearest neighbor search. Additionally, Hash Function Learning (HFL) is considered to accelerate the retrieval process.

On one hand, we proposed a new local metric learning approach - Reduced-Rank Local Metric Learning (R2LML). By considering a conical combination of Mahalanobis metrics, the proposed method is able to better capture information like location and similarity of the data. A regularization to suppress the noise and avoid overfitting is also incorporated into the formulation. Based on the different methods to infer the weights for the local metric, we considered two frameworks: Transductive Reduced-Rank Local Metric Learning (T-R2LML) utilizes transductive learning while Efficient Reduced-Rank Local Metric Learning (E-R2LML) employs a simple trick. Besides, we study the convergence property of the proposed block coordinate descent algorithms for both our frameworks. The extensive experiments show the superiority of our approaches.

On the other hand, *Supervised Hash Learning (*SHL), which could incorporate supervised, semisupervised and unsupervised learning scenarios, was proposed in the dissertation. By considering several codewords which could be learned from the data, the proposed method naturally leads to a set of Support Vector Machine (SVM) problems. After providing an efficient training algorithm, we also study the theoretical generalization bound of the new hashing framework. In the final experiments, *SHL outperforms many other state-of-art hash function learning approaches. Additionally, in order to cope with large data sets, we also conducted experiments running on big data using a parallel computing software package LIBSKYLARK.

Major: Electrical Engineering

Educational Career:
Bachelor's of Electronic Engineering, BS, 2010, Dalian University of Technology
Master's of Electrical Engineering, MS, 2014, University of Central Florida

Committee in Charge:
Michael Georgiopoulos, Chair, Electrical And Computer Engineering
Georgios C. Anagnostopoulos, Co-Chair, Electrical & Computer Engineering
Haiyan Hu, Computer Science
Gita Sukthankar, Computer Science
Liqiang Ni, Statistics

Approved for distribution by Michael Georgiopoulos, Committee Chair, on June 13, 2016.

The public is welcome to attend.