T4 (Full day)

Media Semantics and the Statistical Foundations for Understanding It

AM: *Marc Davis (UC Berkeley),*
Chitra Dorai (IBM Research),
Frank Nack (CWI, Netherlands)

PM: *Edward Chang (UC Santa Barbara)*

Part 2 (PM):

Focus: Statistical Foundation for Multimodal Multimedia Data Analysis and Fusion
Presenter: *Edward Chang*, University of California, Santa Barbara

Intended Audience: Intermediate

Overview:

Multimedia consists of metadata of multiple modalities. For example, a video consists of visual and audio tracks, and captions. An image can be depicted by perceptual features and keywords. Effective analysis and fusion of multimodal metadata plays a key role in the success of annotating, organizing, and indexing multimedia data.

This tutorial aims to provide an overview on statistical methods for conducting multimodal data analysis and fusion. The workshop will be divided into four segments as follows:

1. Overview and Examples
   We start by introducing two challenging research problems in Multimedia: bridging the semantic gap, and metadata fusion. We use examples to illustrate that the current statistical or machine-learning models cannot address the technical challenges arouse from these two problems. Specifically, when the training data is scarce, when the data dimension is high, or when the distribution of training data is imbalanced, no traditional method can conduct reliable statistical inference for even single modality. Furthermore, it is not clear what is the best model for fusing metadata of multiple modalities.

2. Supervised Learning
   We survey key algorithms of the generative and discriminant approaches. For the generative approach, we present Bayesian Networks. We discuss the shortcomings of this model in terms of computational complexity,

and its inability to infer causality in certain conditions. For the discriminant approach, we present kernel methods.

3. Manifold Learning
   We discuss recently developed non-linear feature-reduction algorithms. We will survey several manifold learning algorithms including LLE, Isomap, MDS, kernel PCA, and advanced methods.

4. Multimodal Fusion Techniques
   Building upon the above materials, we will introduce three fusion models: super-kernel fusion, Bayesian fusion, and inference fusion. We will detail their similarities and differences, and point out their pros and cons.

   If time permits, we will discuss learning methods in a non-metric space. This is a critical topic in Multimedia, since many effective distance functions or fusion techniques are non-metric. SVMs cannot deal with a non-metric function.

In summary, this tutorial will not only pinpoint some key research problems in Multimedia, it will also provide an overview on solutions and potential research directions for solving these challenging problems.

Outline:

Overview and Examples
- Problem statement
- Primal & dual models

Supervised Learning
- Kernel Methods
- Bayesian Networks
- (Dirichlet Process if time permits)

Manifold Learning
- Isomap, LLE, kernel PCA, etc.

Fusion Techniques
- Graph model

Super kernel model
Inference model

Advanced Topics

Materials:

Slides and papers will be made available two weeks before the tutorial.

Short Bio:

Professor **Edward Chang** received his M.S. in Computer Science and PhD in Electrical Engineering at Stanford University in 1994 and 1999, respectively. Since 2003, he is an Associate Professor of Electrical and Computer Engineering at the University of California, Santa Barbara. His recent research activities are in the areas of machine learning, data mining, high-dimensional data indexing, and their applications to image databases and video surveillance. Recent research contributions of his group include methods for learning image/video query concepts via active learning, formulating distance functions via dynamic associations and kernel alignment, managing and fusing distributed video-sensor data, and categorizing and indexing high-dimensional image/video information. Professor Chang has served on several ACM, IEEE, and SIAM conference program committees. He co-founded the annual ACM Video Sensor Network Workshop and co-chaired it in 2003 and 2004, and will co-chair three major conferences (ACM MM 06, MMM 06, SIPE 06) in the next two years. He serves as an Associate Editor for IEEE Transactions on Knowledge and Data Engineering, and ACM Multimedia Systems Journal. Professor Chang is a recipient of the IBM Faculty Partnership Award and the NSF Career Award. He is a co-founder of VIMA Technologies, which provides image searching and filtering solutions.

Contact Information:

Presenter: *Edward Chang*, University of California, Santa Barbara
Department of Electrical & Computer Engineering
University of California
Santa Barbara, CA, 93106
echang@ece.ucsb.edu