

Structured Data and Representations in Probabilistic Models for Categorization

Saturday December 18, 2004

Josh Tenenbaum, *MIT*

David Danks, *Carnegie Mellon University*

Chris Williams, *University of Edinburgh*

<http://web.mit.edu/cocosci/nips04.html>

Probabilistic models for categorization are by now widespread in both cognitive science and machine learning, but they have traditionally exploited only minimal structure in either the data or the model. Standard approaches have focused on categorization of minimally structured data (usually flat vectors of feature/attribute values) using relatively flat models (such as mixture models or sets of conditional probability tables in machine learning, and various measures of distance to prototypes or exemplars in computational psychological models). There has been a recent surge of interest in cognitive science and machine learning in developing models and techniques that exploit additional structure either in the relations among the categories (e.g., hierarchical models), or in the data being categorized (e.g., relational data). The advances in learning and applying more structured data and models have led to improved performance for a range of domains and problems, such as inference from very small samples.

This workshop will present state-of-the-art approaches to probabilistic categorization that incorporate more structure, and more sophisticated kinds of structure, than traditional approaches do. Talks will focus on three broad research problems. The morning session presents several different probabilistic models for hierarchical categorization. Different tasks are addressed, including visual categorization of objects and categorization by humans. The afternoon session begins with research on structured models for learning multiple or overlapping categories, as often occurs in text classification. These approaches attempt to learn relationships among the classes in addition to learning each class. The second part of the afternoon session centers on categorization with relational data, and the discovery of relational roles in social, communication, or ecological networks.

The talks in this workshop all approach these problems from machine learning points-of-view, but the workshop is intended for a broader audience than just computer scientists interested in categorization. Many of the problems to be discussed, as well as some of the solutions, are inspired by human cognition or perception, and so the workshop should appeal to computational cognitive scientists, psychologists, and neuroscientists. A broad range of application domains will also be discussed, including visual datasets, text and email corpora, human behavioral judgments, social networks and anthropological data. The discussion sessions will include proposals for “challenge problems”, problems that participants feel are important and cut across these research fields, and “crucial datasets” – perhaps already existing, perhaps yet-to-be-collected – that should be a focus of further research and experimental tests.

Structured Data and Representations in Probabilistic Models for Categorization

Saturday December 18, 2004

Organizers: Josh Tenenbaum, David Danks & Chris Williams

Friday morning session: 7:30am–10:30am

- 7:30am **Introduction**, *Workshop organizers*
- 7:40am **Models for Probabilistic Hierarchical Clustering**, *C. Williams*
- 8:00am **Bayesian Methods for Hierarchical Clustering**, *Z. Ghahramani*
- 8:30am **Modeling Topical Structure in Text**, *D. Blei and J. Lafferty*
- 9:00am *coffee break*
- 9:10am **Hierarchical Constraints in Visual Object Categorization**,
G. Shakhnarovich and T. Darrell
- 9:30am **Chain Graph Models of Psychological Categorization**, *D. Danks*
- 10:00am *Discussion*

Friday afternoon session: 4:00pm–7:00pm

- 4:00pm **The couch potato project: learning about objects from looking at images**,
J. Sivic, B. Russell, A. Efros, A. Zisserman, B. Freeman
- 4:30pm **Infinite latent feature models and the Indian Buffet Process**,
T. Griffiths and Z. Ghahramani
- 4:50pm **Semiparametric Latent Factor Models for Learning Multiple Related Categories**,
Y. W. Teh, M. Seeger, and M. Jordan
- 5:10pm *coffee break*
- 5:20pm **Discovering Latent Classes in Relational Data**,
C. Kemp, T. Griffiths, and J. Tenenbaum
- 5:50pm **The Author-Recipient-Topic Model for Topic and Role Discovery in Social
Networks**, *A. McCallum, A. Corrada-Emmanuel, and X. Wang*
- 6:10pm *General discussion*