



## People

- **Convener** : Prof. Chiranjib Bhattacharyya
- **Former doctoral students** : Saketha Nath, J, Sourangshu Bhattacharyya
- **Current Students**
  - Ph.D** : Achintya Kundu, Mrinal Das, Sahely Bhadra, Goutham Tholpadi, Yogananda A.P.
  - M.Sc(Engg)** : Raman Sankaran, Neeraja Yadwadkar, Vishal Patel, Vipul Agrawal.
  - M.E.** : Adway Mitra, Himabindu, Vikram Tankasali, Rahul Saherawat, Shunmugakrishnan.S
- **Academic Collaborators** : Prof. A.Ben-tal (Technion, Israel), Prof. Soumen Chakrabarti (IIT, Bombay), Prof Hari Kalva (Univ. Florida), Prof K.R.Ramakrishnan (EE, IISc), Dr. Shirish Shevade (CSA, IISc), Dr. Saketha Nath (IIT, Bombay)
- **Industry partners** : NetApp, Yahoo!, SAP, IBM IRL, Samsung.

## Areas of Research

- We are interested in both theoretical and applied aspects of Machine Learning. At this point theoretical interests are
- Convex Optimization for machine learning
  - Multiple Kernel Learning
  - Covariance profiles
  - Approximate inference in Graphical Models
- We are pursuing applications of the above in several domains including
- Natural Language Processing
  - Computer Vision
  - Computational Biology
  - Computer Systems

## Computational Biology

### Chance-constrained Classification

- Project carried out by **Sahely Bhadra** along with **J.Saketha Nath**
- A novel methodology for constructing maximum margin classifiers which are robust to uncertainties in data by employing partial statistical information of the uncertainties.
- Uncertain data classification problem posed as CCP and relaxed as a convex SOCP problem using Bernstein bounding schemes.
- This method of efficiently handling data uncertainties result in very good classification margin, which reduces generalization error.
- Application in bioinformatics- like **tumor classification**, based on features calculated from the tumor cell and **classification of drugs**, based on gene-expression data, in presence of measurement uncertainties.
- Paper accepted for publication in **Mathematical Programming Series B, 2010**

## NETGEM

- Project carried out by **Vinay Jethava**
- A tractable model rooted in Markov dynamics, for analyzing temporal profiles of genetic expressions arising out of known protein interaction networks evolving with unknown dynamics.
- It is able to **identify temporal interactions between gene expressions** and determine their strength
- It can also identify functional categories of the actively interacting partners and dynamics of interactions in perturbed networks.

### Sub-Pointset pattern matching using Graphical Models

- Project carried out by **Vikram Tankasali** in collaboration with **Sourangshu Bhattacharyya, Yahoo!**
- 3-Dim pointset pattern matching in presence of jitter, and missing points is extremely difficult.
- Problem is formulated as MAP estimation over graphical models.
- The learnt mapping is optimal if the mapped subgraph is rigid.
- Applied on the task of **binding site comparisons of ligands**.
- Can also be applied for **robust shape matching in computer vision**.

### Learning from multiple indefinite similarity matrices

- Project carried out by **Vikram Tankasali** and **Achintya Kundu**.
- Kernel functions are useful tools for measuring similarity. We consider the broad problem of combining multiple similarity matrices.
- Proposed a Mirror-Descent based algorithm for learning from multiple indefinite similarity matrices,
- The algorithm can handle any general convex loss function.
- Experimental results on **classification of proteins into folds demonstrate the efficiency of the algorithm**.

## Computer Vision

### Covariance Profiles for Video Track Clustering

- Project carried out by **Adway Mitra** in collaboration with **Computer Vision Lab, EE Department, IISc and Sreedal Menon**
- Covariance Profiles are a summary of covariance matrices obtained by Joint Diagonalization
- Application to computer vision tasks by representing images with covariance descriptors
- Efficient Representation of a set of images having similar covariance structure, as in successive frames of a video track
- Can also be used for class representation in classification problems such as object recognition and face recognition.
- Impressive results for **clustering low-resolution noisy videos e.g. Youtube**. which may help in retrieving videos of any popular personalities.

## Variable Sparsity Kernel Learning

- Project carried out by **Raman Sankaran** along with **G.Dinesh** and **J.Saketha Nath**
- Mixed norm MKL formulation, and an efficient algorithm to solve the dual, which uses Mirror Descent optimization.
- Well suited for **Object categorization** and other multi modal tasks, where the data have **multiple feature descriptors**.
- Outperforms state-of-the-art MKL formulations and solvers in terms of accuracy and speed.
- Relevant literature: Published in **NIPS 2009**

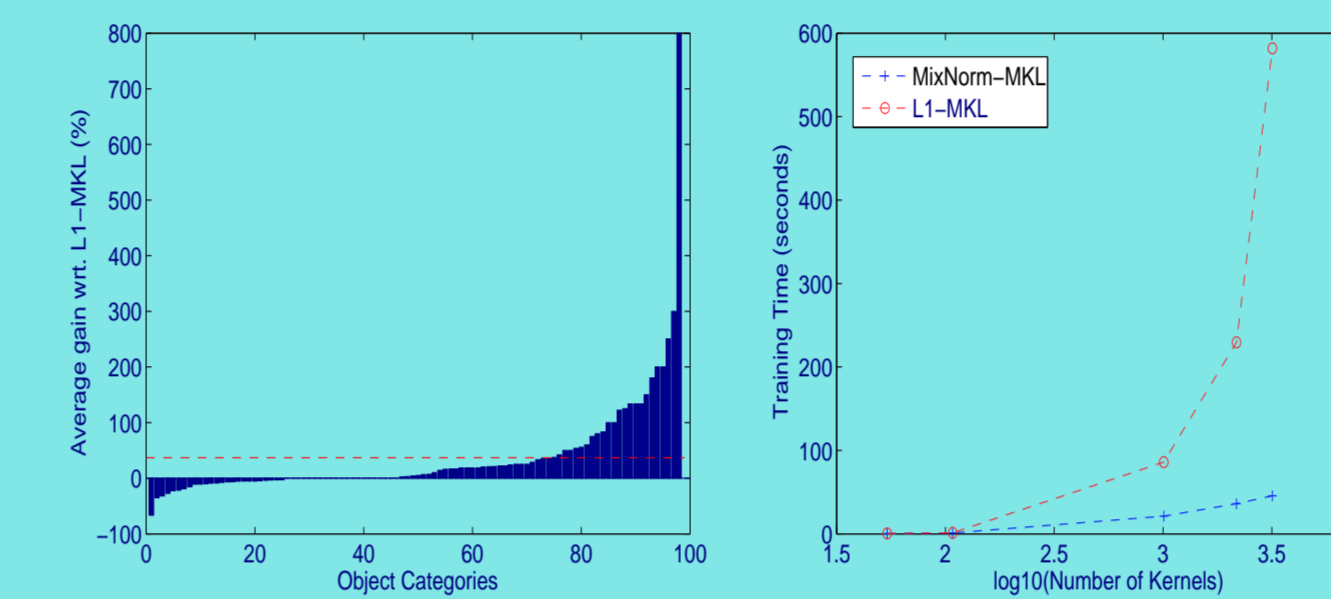


Figure 2: Accuracy Gain and speedup of VSKL compared to SimpleMKL.

## Computer Systems

### Application workload discovery

- Project carried out by **Neeraja Yadwadkar**
- **Workload identification** based on NFS traces, which can be obtained non-intrusively and analyzed either online or offline.
- Design of a trace analysis methodology with Powerful Discriminatory Capabilities as well as **Annotation Capabilities**.
- Figure 4 shows a Profile Hidden Markov Models based trace analysis methodology.
- Knowledge of applications running can enable **autonomic systems, Provenance Detection** along with building highly efficient storage systems
- Relevant literature: Published in **FAST 2010**

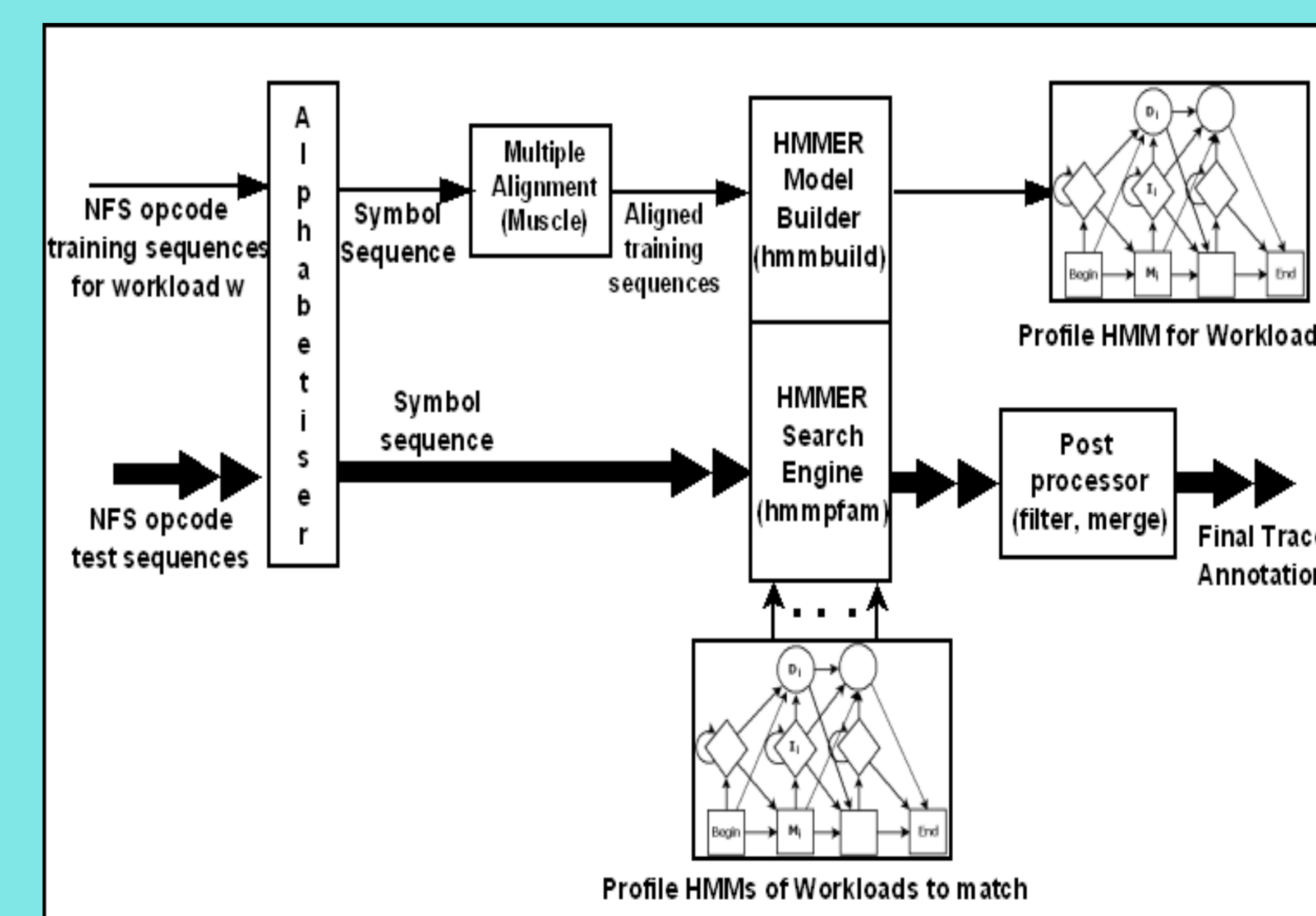


Figure 4: Profile HMMs Training and Usage Workflow: An Overview

## Natural Language Processing

### ShoBha

- Project carried out by **Goutam Tholpadi** and **Mrinal Das**
- ShoBha: a system that enables exploratory search over multilingual document collections
- **Labeling multilingual document clusters** using generative topic models
- Uses no language resources other than a comparable multilingual corpus
- Has relevance for resource scarce languages and suited for Indian languages.

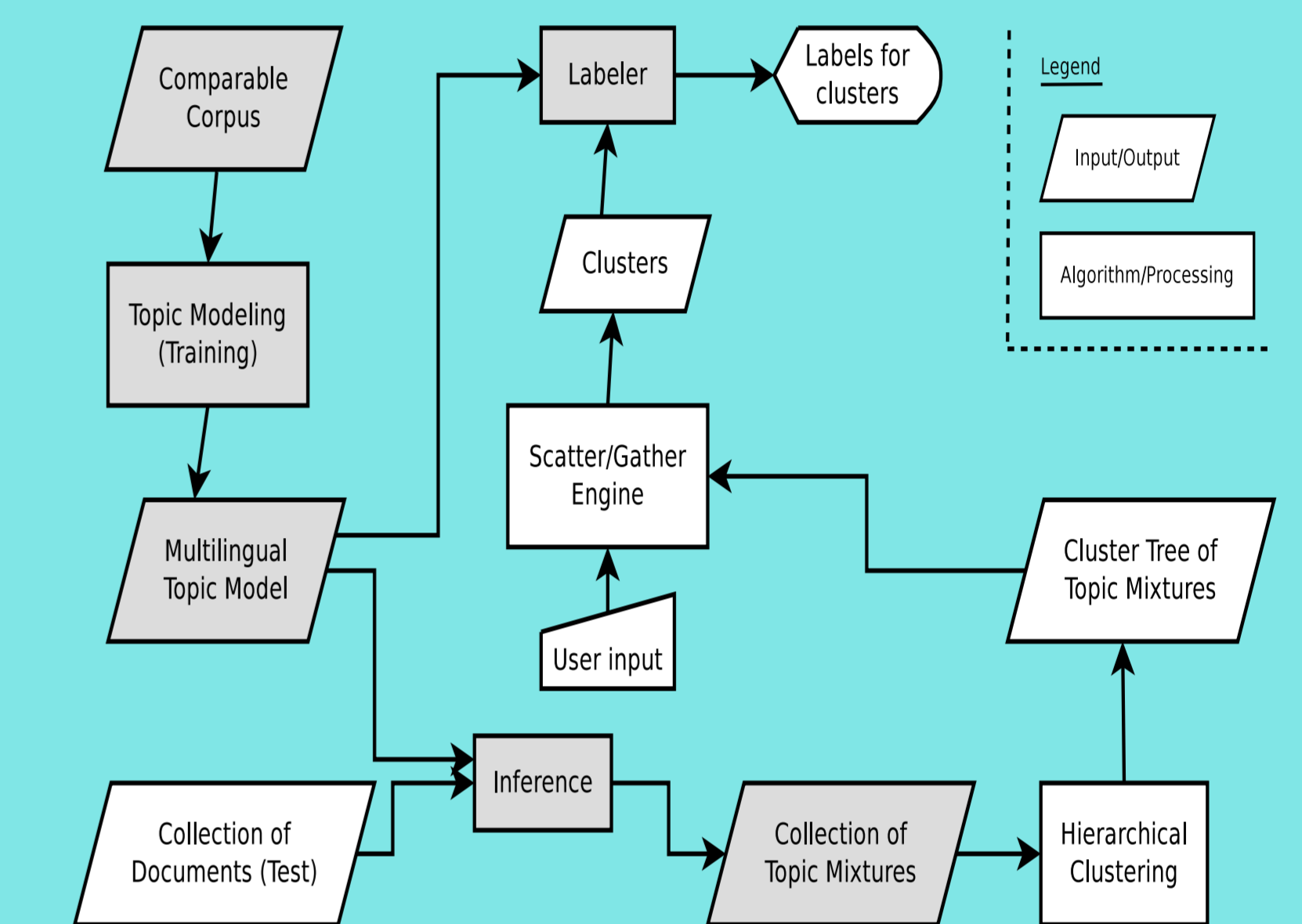


Figure 6: Flow diagram of ShoBha

## Achievements

- Publishing regularly to **top Machine Learning conferences** like NIPS, ICML, KDD etc
- Won the **Best Student Submission** award in KDD Cup 2006
- Sahely Bhadra's paper won the **Runner's-Up Best Paper Award** in PAKDD 2009
- Alumni placed in leading positions of industry and academia.