



MACHINE LEARNING LAB

Department of Computer Science and Automation
Indian Institute of Science



People

• Faculty Members

Dr. Chiranjib Bhattacharyya, Dr. Indrajit Bhattacharya and Dr. Shivani Agarwal

• Former doctoral students

Saketha Nath, Sourangshu Bhattacharya

• Current Students

Ph.D : Achintya Kundu, Adway Mitra, Vikram Tankasali, Mrinal Das, Sahely Bhadra, Goutham Tholpadi, Raman Sankaran, Yogananda A P, Arunachalam E V, Arun R, Ranganath B N

M.E. : Harish Guruprasad, Dyut Kumar Sil

Project Assistants : Neeraja Yadwadkar

• Academic Collaborators

Dr. A Ben-tal (Technion, Israel), Dr. Soumen Chakrabarti (IIT Bombay), Dr. K R Ramakrishnan (EE, IISc), Dr. Shirish Shevade (CSA, IISc), Dr. Saketha Nath (IIT Bombay), Dr. Suresh Manandhar (Univ. of York)

• Industry partners

NetApp, Yahoo!, SAP, IBM IRL, Samsung, Infosys

Areas of Research

- We are interested in both theoretical and applied aspects of Machine Learning. At this point, the 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
 - Natural Language Processing
 - Computer Vision
 - Computational Biology
 - Computer Systems
 - Text Mining

Achievements

- Publishing regularly in **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-Up Best Paper Award** in PAKDD 2009
- Alumni placed in leading positions of industry and academia.

Artificial intelligence is the study of how to make real computers act like the ones in the movies.
Anonymous

Computational Biology

Chance-constrained Classification

- Real world data is noisy due to measurement error
- Constructing maximum margin classifiers and non-linear kernels which are robust to uncertainties in data
- Application in bioinformatics: **tumor classification**, **classification of drugs**
- Papers published in **Mathematical Programming Series B, 2010** and **ICML 2010**

Learning from multiple indefinite similarity matrices

- Kernel functions can be viewed as measures of similarity
- A Mirror Descent algorithm for learning kernels from multiple indefinite similarity matrices, for any general convex loss function
- Application to **classification of proteins into folds**
- Paper published in **NIPS 2010**

Computer Vision

Covariance Profiles for Video Tracks

- Short video tracks have associated "entities": an object or a person
- **Represent** and **compare** tracks to classify, cluster or retrieve videos based on associated entities
- **Covariance Profiles**: to summarize tracks and define kernels on tracks
- Can also be used for face, object and facial gesture recognition

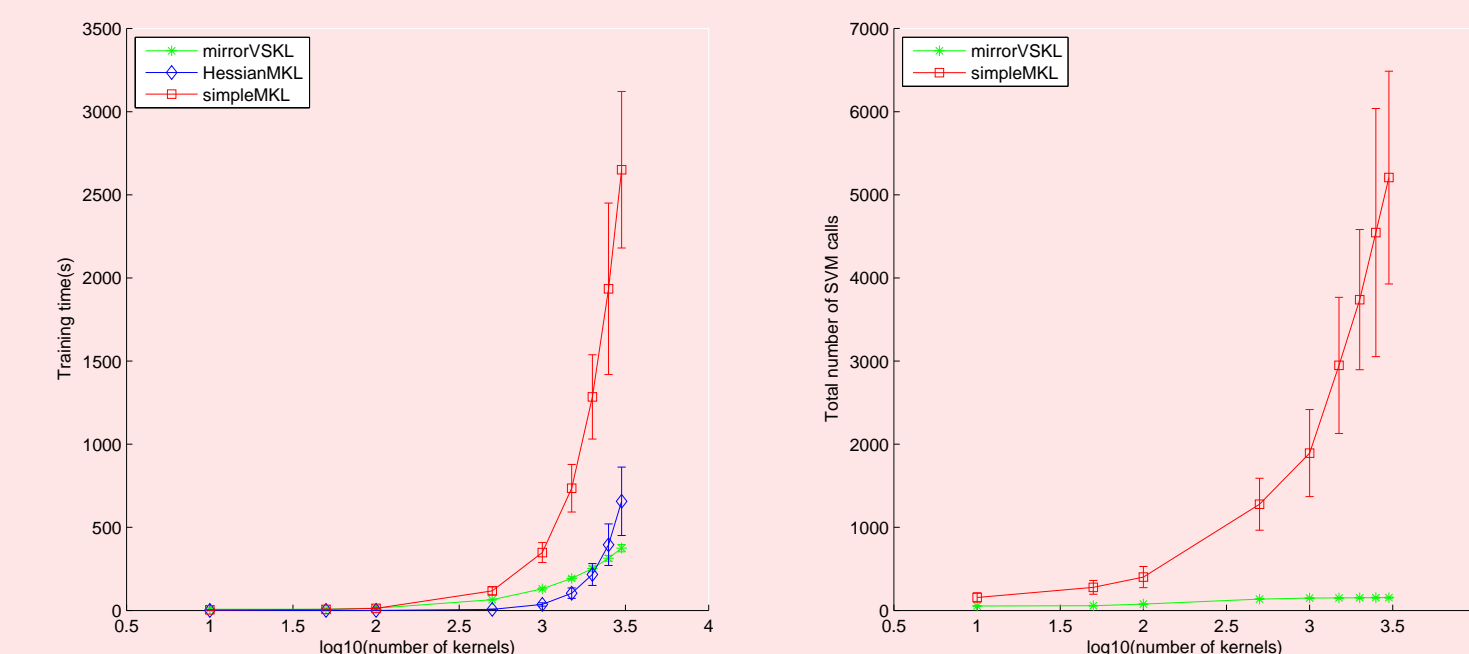
EigenProfiles for Object Tracking

- **Object Tracking**: Track a moving object across a video
- Object can change in appearance due to change in viewing angle and illumination
- Need efficient and robust incremental object representation
- **Eigen Profiles**: online estimate of object descriptor from recent samples
- Also works well on videos taken in poor illumination



Variable Sparsity Kernel Learning

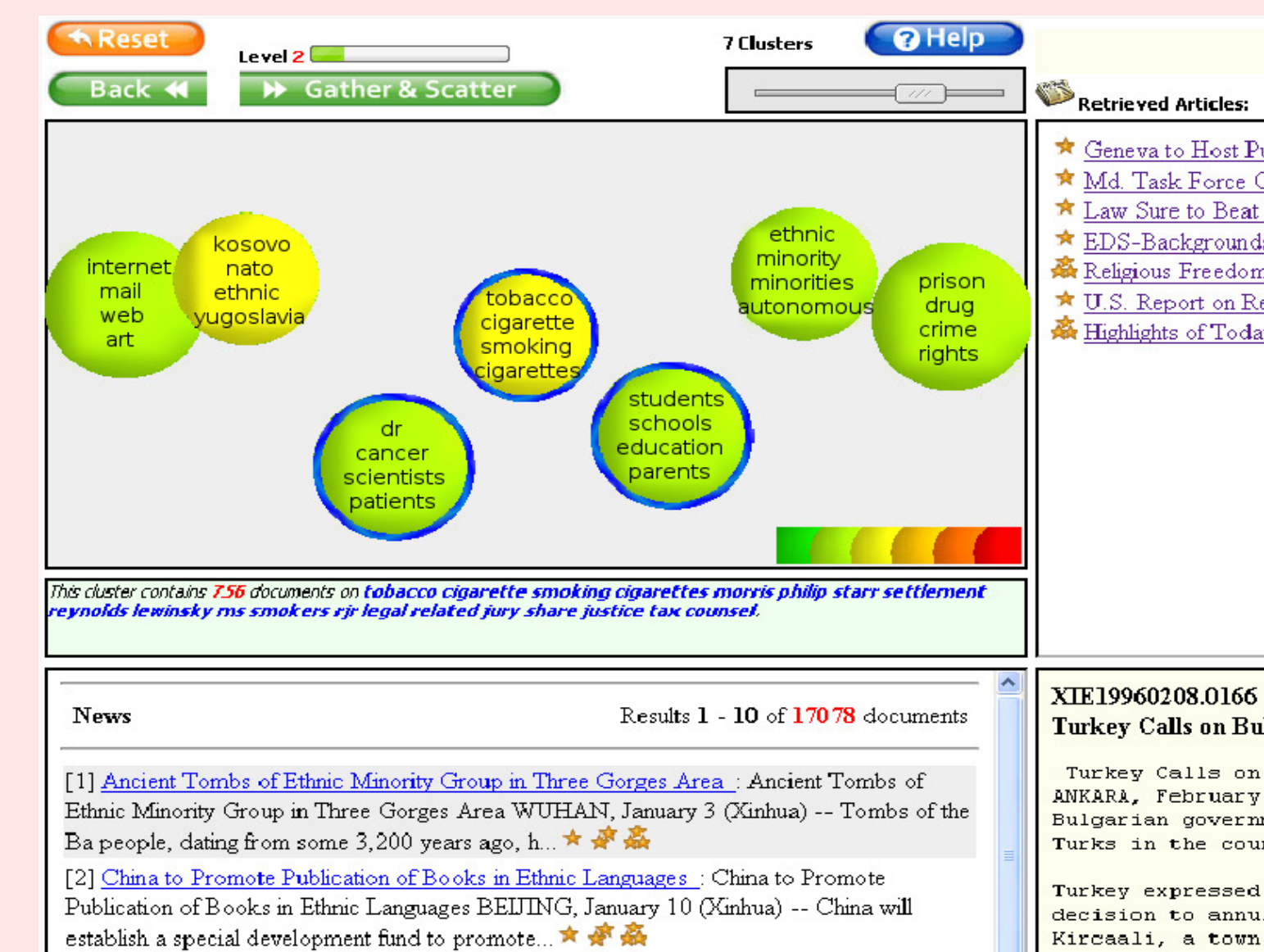
- Mixed-norm MKL formulation, and an efficient algorithm to solve the dual, which uses Mirror Descent optimization
- Applications: **Object categorization**, and other multi-modal tasks, where the data have **multiple feature descriptors**
- Outperforms state-of-the-art MKL formulations and solvers in terms of accuracy and speed
- Published in **NIPS 2009**; accepted for publication **JMLR 2011**



Natural Language Processing

Multilingual Exploratory Search

- Clustering multilingual document collections and labeling these 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



Text Mining

Facet and Sentiment Extraction from Customer Reviews

- Automated summarization of customer reviews of products
- Facet-specific opinion summaries: e.g. "What are people saying about the picture quality of a camera X?"
- Probabilistic models to jointly discover latent facets and sentiments independent of language and domain
- Accepted for publication in **SDM 2011**; a patent is being filed

ReadAlong: Article Segments and Comments

- Match user comments(c_i) to article segments(s_j)
- Similarity(c_i, s_j) is taken using topic distribution (from **LDA**) and cosine similarity (**tf-idf** vectors)
- Accepted for publication (poster) in **WWW 2011**

Karate Kid
The latest version of *Karate Kid* has two stars – Jaden Smith and Jackie Chan. Jaden Smith looks great as an expat while Jackie Chan is his usual charming self as a martial arts master.

Comments
1. Jackie Chan teaches Kung Fu in this film.
2. Jaden Smith's initial experiences in China are well-portrayed.
3. Did you know he combines wushu, hung gar and wing chun?

Jaden Smith's initial experiences in China are well-portrayed.

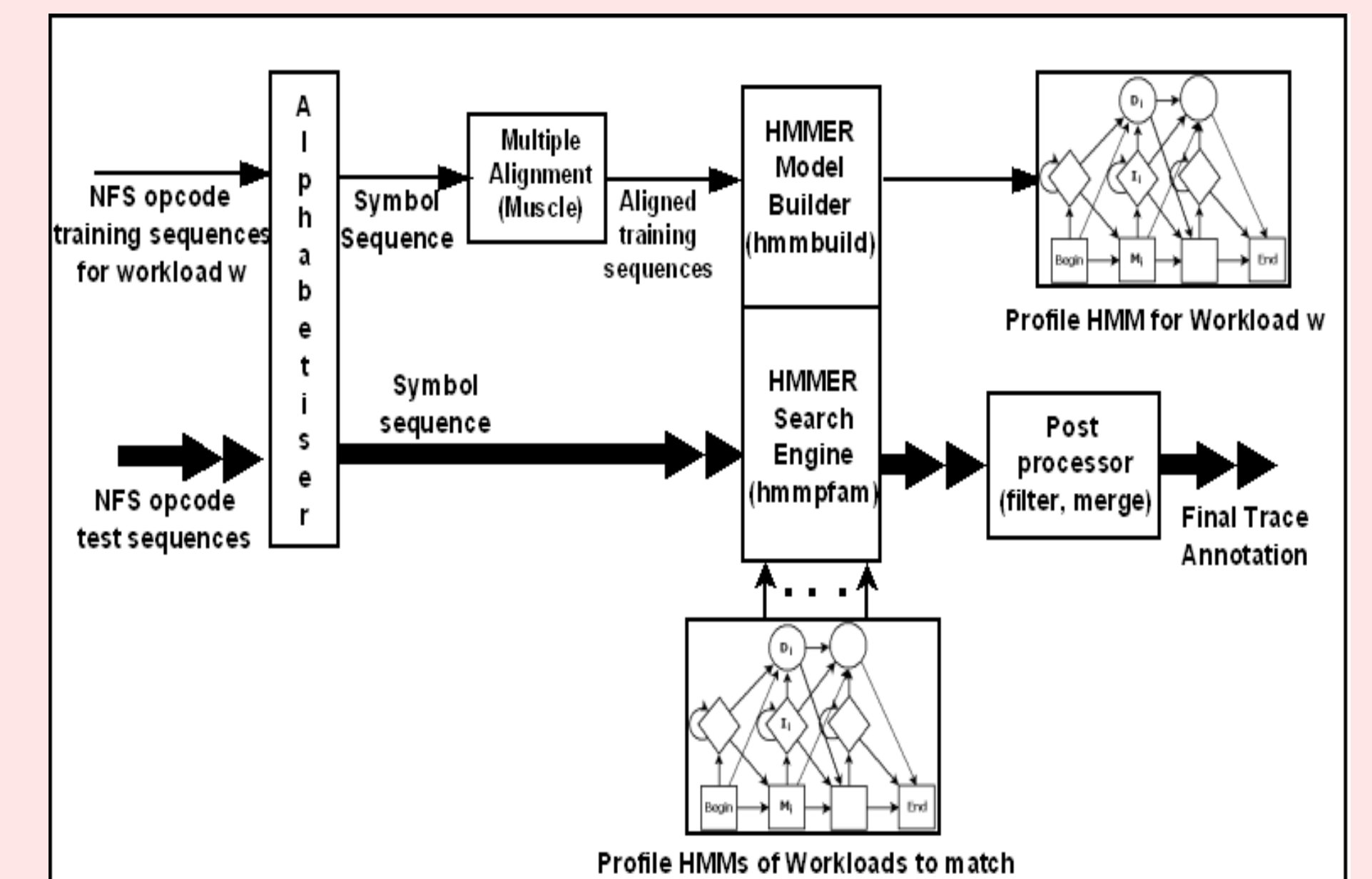
Karate Kid
The latest version of *Karate Kid* has two stars – Jaden Smith and Jackie Chan. Jaden Smith looks great as an expat while Jackie Chan is his usual charming self as a martial arts master.

1. Jackie Chan teaches Kung Fu in this film.
2. Did you know he combines wushu, hung gar and wing chun?

Computer Systems

Workload discovery

- **Workload identification** based on file system traces
- Trace analysis methodology with powerful discriminatory and annotation capabilities
- Enables **autonomic systems** and **provenance detection**
- Published in **FAST 2010**



Learning Application Environment from Traces

- Identifying execution environment (operating system, file system, application type etc.) using network file traces
- Unified trace processing engine based on SVMs

The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard.
Steven Pinker