Foundations of Large-Scale Multimedia Information Management and Retrieval

Lecture #1 Introduction

Edward Y. Chang
Foundations of Large-Scale Multimedia Information Management and Retrieval Mathematics of Perception covers knowledge representation and semantic analysis in multimedia data and scalability in signal extraction, data mining, and indexing. The book is divided into two parts: Part I - Knowledge Representation and Semantic Analysis focuses on the key components of mathematics of perception as it applies to data management and retrieval. These include feature selection/reduction, knowledge representation, semantic analysis, distance function formulation for measuring similarity, and multimodal fusion. Part II - Scalability Issues presents indexing and distributed methods for scaling up these components for high-dimensional data and Web-scale datasets. The book presents some real-world applications and remarks on future research and development directions.

The book is designed for researchers, graduate students, and practitioners in the fields of Computer Vision, Machine Learning, Large-scale Data Mining, Database, and Multimedia Information Retrieval.

Dr. Edward Y. Chang was professor at the Department of Electrical & Computer Engineering, University of California at Santa Barbara, before he joined Google as a research director in 2006. Dr. Edward Y. Chang received his M.S. degree in Computer Science and Ph.D degree in Electrical Engineering, both from Stanford University.
Datasets

• YouTube
• Facebook Photos
• Netflix Movies
• Apple Music
• Archive.org
  – http://www.archive.org/details/audio
  – 400k recordings
Technical Challenges

• Volume, both too small and too large
• Variety, text, video, image, music, social, etc.
• Velocity
## Prefixes

<table>
<thead>
<tr>
<th>SI prefix</th>
<th>Name</th>
<th>Power of 10 or 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>kilo</td>
<td>$10^3$ $2^{10}$</td>
</tr>
<tr>
<td>M</td>
<td>mega</td>
<td>$10^6$ $2^{20}$</td>
</tr>
<tr>
<td>G</td>
<td>giga</td>
<td>$10^9$ $2^{30}$</td>
</tr>
<tr>
<td>T</td>
<td>tera</td>
<td>$10^{12}$ $2^{40}$</td>
</tr>
<tr>
<td>P</td>
<td>peta</td>
<td>$10^{15}$ $2^{50}$</td>
</tr>
<tr>
<td>E</td>
<td>exa</td>
<td>$10^{18}$ $2^{60}$</td>
</tr>
<tr>
<td>Z</td>
<td>zetta</td>
<td>$10^{21}$ $2^{70}$</td>
</tr>
<tr>
<td>Y</td>
<td>yotta</td>
<td>$10^{24}$ $2^{80}$</td>
</tr>
</tbody>
</table>
## Prefixes

<table>
<thead>
<tr>
<th>SI prefix</th>
<th>Name</th>
<th>Power of 10 or 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>kilo</td>
<td>$10^3$</td>
</tr>
<tr>
<td>M</td>
<td>mega</td>
<td>$10^6$</td>
</tr>
<tr>
<td>G</td>
<td>giga</td>
<td>$10^9$</td>
</tr>
<tr>
<td>T</td>
<td>tera</td>
<td>$10^{12}$</td>
</tr>
<tr>
<td>P</td>
<td>peta</td>
<td>$10^{15}$</td>
</tr>
<tr>
<td>E</td>
<td>exa</td>
<td>$10^{18}$</td>
</tr>
<tr>
<td>Z</td>
<td>zetta</td>
<td>$10^{21}$</td>
</tr>
<tr>
<td>Y</td>
<td>yotta</td>
<td>$10^{24}$</td>
</tr>
</tbody>
</table>
Key Applications

Use Image as an Example

• Content-based Retrieval
  – Given an image, find perceptually similar ones

• Annotation
  – Given an image, identify what (objects, events), where (location), when (time), and who (people)

• Do the above both quickly and accurately

• Applicable to any multimedia data types
Key Subroutines

- Feature Extraction
- Machine Learning
  - Small sample pool
  - Large sample pool
- Similarity
- Multimodal Fusion
- Indexing
- Scalability
  - Parallel algorithms
  - Online algorithms
Feature Extraction

• Representative Methods
  – Color, texture, shape
  – SIFT
  – Bio-motivated HMAX
  – Deep Learning

• Model-Based vs. Data-Driven
Classification

• Query-Concept Learning
  – Online
  – Concept-dependent learning
  – Imbalanced data
  – Large D
  – D >> N
  – N⁻ >> N⁺

• Annotation
  – Offline
  – Feature-to-semantics mapping
  – Large N
Similarity

• Traditional Distance Function
• Dynamic Partial Function
• Learning Similarity from Data
Multimodal Fusion

• Feature Fusion (Cross-Modality)
  – Weighted sum
  – PCA
  – Dimensionality curse

• Personalization (Cross-Domain)
  – Latent behavior

• Context + Content
  – Location, time, et al.
Storage and Indexing

• Dimensionality Curse
• High-Dimensional Indexing
Large-Scale Learning

3V Problem

• Huge Volume of Data
• High Velocity
• High Variety
Key Parallel Algorithmic Work


- **PFP**: **Parallel** FP-Growth for Query Recommendation, H. Li, Y. Wang, D. Zhang, M. Zhang, and E. Y. Chang, ACM Recommendation Systems, Lausanne, October 2008
More Data vs. Better Algorithms

Banko & Brill, 2001

Figure 2. Learning Curves for Confusable Disambiguation
More Data vs. Better Algorithms

Banko & Brill, 2001

Figure 2. Learning Curves for Confusable Disambiguation
More Data vs. Better Algorithms

Banko & Brill, 2001

Figure 2. Learning Curves for Confusible Disambiguation
More Data vs. Better Algorithms

Figure 2. Learning Curves for Confusable Disambiguation

Banko & Brill, 2001
Comparative Study on Small Dataset

Fig. 7: Accuracy for ImageNet dataset. X-axis shows the number of training instances for each category. Y-axis shows the average accuracy.
Large Dataset

10,000 instances per class
Fig. 14: Accuracy for PicasaWeb dataset. X-axis shows the number of training instances for each category. Y-axis shows the average accuracy.
Data Center (Cloud)

- Servers
  - CPUs
  - DRAM
  - Disks

- Racks
  - 40-80 servers
  - Ethernet switch

Warehouse-scale Computer (WSC)
Sample Hierarchy

• Server
  – 16GB DRAM; 160 GB SSD; 5 x 1TB disk
• Rack
  – 40 servers
  – 48 port Gigabit Ethernet switch
• Warehouse
  – 10,000 servers (250 racks)
  – 2K port Gigabit Ethernet switch
Syllabus

• Feature Extraction (Lecture #2)
• Machine Learning (Lecture #3)
  – Small sample pool
  – Large sample pool
• Similarity (Lecture #4)
• Multimodal Fusion (Lecture #5)
• Indexing (Lecture #6)
• Scalability (Lectures #7 & #8)
  – Parallel algorithms
  – Online algorithms
Reading

  - Chapter #1