CS4241: Lecture 1
Intro to MM Information Retrieval

Mohan S Kankanhalli

January 1 2002
School of Computing
National University of Singapore
Topics

These are the topics that we will cover in the course:

- Introduction
- MM Retrieval-Framework
- Color-based Retrieval
- Texture-based Retrieval
- Shape-based Retrieval
- Audio Retrieval
- Video Retrieval
- Multi-attribute query and knowledge-based retrieval
- MM Info Systems Trends
MM Info Retrieval

Concerns with:

- Basic concepts and techniques in retrieving (unstructured) information
- Indexing and similarity-based retrieval of multimedia data

What is an information retrieval system?

- A system used to process, store, search, retrieve and disseminate information items
- Examples: DBMS, Free-text Systems, Hypermedia Systems etc.
Information Needs

- Volume of Information is growing at an exponential rate
  - By 1800, the amount of scientific information was doubling every 50 years
  - From 1800 to 1966, the no. of scientific journals has increased from 100 to over 10,000
- Large amount of data is available in the electronic form
  - Newspaper and picture archives, video & music catalogs, satellite images, financial data etc.
  - Latest web estimates: 1 billion pages, 20 terabytes of information: see

http://www.neci.nj.nec.com/homepages/lawrence/websize.html

We are inundated with information!
Info Retrieval Needs

- Aids are needed to retrieve information:
  - friends, library card system, references, reviews etc.
- Difficult since the data is unstructured
  - it differs from the DBMS structured record:

<table>
<thead>
<tr>
<th>Name:&lt;s&gt;</th>
<th>Sex:&lt;s&gt;</th>
<th>Age:&lt;I&gt;</th>
<th>NRIC:&lt;s&gt;</th>
<th>⋯</th>
</tr>
</thead>
</table>

- Information must be analyzed, indexed (either automatically or manually) for retrieval purposes.

- Examples:
  - text retrieval systems (LINC, Information Banks, AltaVista)
  - image systems (hospitals, police mugshot systems, TCS)
  - others: video, music information bases
Properties of MM Data

1: Media

2: Visual in nature?

3: Ease of data entry

4: Abstract concepts

5: Spatial Dimensions

6: Temporal Dimension

7: Does it have a well-defined interaction unit?

8: Does it have a well-defined semantic unit?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Y</td>
<td></td>
<td>Easy</td>
<td>Y</td>
<td>1</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Graphics</td>
<td>Y</td>
<td></td>
<td>Moderate</td>
<td>N</td>
<td>2-3</td>
<td>Y/N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Image</td>
<td>Y</td>
<td></td>
<td>Difficult</td>
<td>N</td>
<td>2</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Video</td>
<td>Y</td>
<td></td>
<td>Very Difficult</td>
<td>N</td>
<td>2</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Audio</td>
<td>N</td>
<td></td>
<td>Moderate</td>
<td>Y</td>
<td>1</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>
Info Retrieval Process

1. Information Seeking Need

2. Formulation of Requirements (in terms of a Query)

3. Retrieve Information that meets the Requirements

4. Return (ranked) List of Relevant Information

5. Go back to 2 if necessary

Note:

- The process can be repeated with new, improved formulation of information requirements
System Overview

Overview of a typical query-based IR system

Information Problem

Representation

Multimedia Database

Representation

Indexed Representation

Query

Best Match

Ranked List of Items

Evaluation of Results
Attributes Considered

These are some of the typical attributes considered for MM (Image, Audio & Video) data:

- No inherent structure
  - just an array of pixels with text captions

- Global features:
  - text description
  - color, texture
  - color with spatial information
  - heuristic measures for dominant shape
  - can be automatically extracted and indexed

- Local features:
  - shapes: need to perform image/frame segmentation
  - manual in nature
Query Processing

Query processing for visual data needs special attention:

- query formulation:
  - sample image
  - colors, textures, shapes etc.

- queries are ambiguous and often incomplete

- representation is also inexact

- relevance can be determined easily
Architecture

The architecture of the retrieval model has four layers:

Concept Layer
\((names \ of \ objects \ and \ relationships)\)

Object Layer
\((blocks \ of \ attributes)\)

Feature Layer
\((colors, \ textures, \ extracted \ shapes)\)

Data Layer
\((images, \ video, \ text \ documents)\)
Similarity Measures

Given two multimedia objects:

\[
X = (x_1, x_2, \ldots, x_n), \quad \text{where } x_i \in \mathbb{R}
\]

\[
Y = (y_1, y_2, \ldots, y_n), \quad \text{where } y_i \in \mathbb{R}
\]

We need a measure to determine the closeness between these two objects.

There are three types of similarity measures:

- Distance Measures
- Correlation Coefficients
- Association Coefficients
Distance Measures

1. Mean Character Difference:

\[ \frac{1}{n} \sum_{i=1}^{n} |x_i - y_i| \]

2. Minkowski Metric:

\[ d(X,Y) = \left( \sum_{i=1}^{n} |x_i - y_i|^r \right)^{\frac{1}{r}} \]

(a) Manhattan Distance \((r = 1)\)

\[ d(X,Y) = \sum_{i=1}^{n} |x_i - y_i| \]

(b) Euclidean Distance \((r = 2)\)

\[ d(X,Y) = \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}} \]

Intuitive notion of distance!

(c) Chebyshev Distance \((r = \infty)\)

\[ d(X,Y) = \max_{1 \leq i \leq n} |x_i - y_i| \]
Correlation Coefficients

- Unnormalized Correlation Coefficient:

\[ C = \sum_{i=1}^{n} x_i y_i \]

Inner product of vectors!

- Cosine Measure:

\[ \cos \theta = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} x_i^2}} \]

- Pearson Product Moment Measure:

\[ r_{XY} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x}_i)^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2}} \]

- measures moments around mean
- similar to cosine measure
Association Coefficients I

- For binary or multi-state features
- Measures the amount of agreement

Assume: 2 binary valued feature vectors:

\[ X = (x_1, x_2, \ldots, x_n) \quad Y = (y_1, y_2, \ldots, y_n) \]

For the two objects \( X \) and \( Y \), let

\[ \alpha = \text{number of features which are 1 for both} \]
\[ \beta = \text{number of features which is 1 for } X \text{ and 0 for } Y \]
\[ \gamma = \text{number of features which is 0 for } X \text{ and 1 for } Y \]
\[ \delta = \text{number of features which are 0 for both} \]


Then we have the following measures:

\[
[Russel \ & \ Rao] \ d(X, Y) = \frac{\alpha}{\alpha + \beta + \gamma + \delta}
\]

\[
[Jaccard \ & \ Needham] \ d(X, Y) = \frac{\alpha}{\alpha + \beta + \gamma}
\]

\[
[Kulzinski] \ d(X, Y) = \frac{\alpha}{\beta + \gamma}
\]

\[
[Sokal \ & \ Mitchener] \ d(X, Y) = \frac{\alpha + \delta}{\alpha + \beta + \gamma + \delta}
\]

\[
[Rogers \ & \ Tanimoto] \ d(X, Y) = \frac{\alpha + \delta}{\alpha + \delta + 2(\beta + \gamma)}
\]

\[
[Yule] \ d(X, Y) = \frac{\alpha \delta - \beta \gamma}{\alpha \delta + \beta \gamma}
\]

Gower’s general similarity coefficient, \(S_G\),

\[
S_G = \frac{\sum_{i=1}^{n} \omega_i s_i}{\sum_{i=1}^{n} \omega_i}
\]

where \(s_i = 1\) if \(x_i\) matches \(y_i\) and \(s_i = 0\) if \(x_i\) mismatches \(y_i\). \(\omega_i = 1\) if the comparison is valid and is 0 if comparison is not valid or if the state is unknown.
Complex Sim. Measures

- Multiple Features:
  - like color, texture and shape for images

  If each feature $i$ has measure $\sigma_i$
  then the overall similarity measure is:

  $$\sigma = \sum_{i=1}^{k} \lambda_i \sigma_i$$

  The weights $\lambda_i$ are usually chosen empirically.

- Compound Retrieval:
  - mix of continuous & binary features
  - $k$ features with similarity measure $\sigma_i$
  - $j$ binary features

  Then the overall similarity measure is:

  $$\sigma = \prod_{j} \rho_j(X, Y) \sum_{k} \lambda_k \sigma_k(X, Y).$$
Data Considerations I

Normalization

1. Ranging:

\[ x'_i = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \]

Features now range from 0 to 1
No lopsided effect of any feature!

2. Zero Mean: Assume \( N \) objects:

\[ m_i = \frac{1}{N} \sum_N x_i \]

Then the normalization can be done in this manner:

\[ x'_i = x_i - m_i \]

3. Zero Mean and Unit Variance:

\[ s_i = \frac{1}{N} \sum_N (x_i - m_i)^2 \]

and the normalization can be applied as:

\[ x'_i = \frac{x_i - m_i}{s_i}. \]
Data Considerations II

Missing Data

Suppose a feature is missing from the query object or a database object. Then how to compute the distance? Let

\[
d_i = \begin{cases} 
0 & \text{if } x_i \text{ or } y_i \text{ is missing} \\
 x_i - y_i & \text{otherwise} 
\end{cases}
\]

Then the distance between X and Y is defined as:

\[
d(X, Y) = \frac{n}{n - n_0} \sqrt{\sum_i d_i^2}
\]

where \( n_0 \) is the number of features missing in X or Y or both.
Summary

- There is an information explosion
- Need to process, store and retrieve this data effectively (by content) and efficiently (using indexes)
- We will focus on unstructured text, image and video data
- These data have local and global attributes
- Features are extracted for these attributes which uniquely capture the attributes
- Data objects are retrieved by feature similarity
- Various types of similarity measures are employed:
  - Distance Measures
  - Correlation Coefficients
  - Association Coefficients