Feature Sets Based Similarity Measures for Image Retrieval

Abebe Rorissa

(arorissa@albany.edu)

College of Computing and Information, University at Albany, State University of New York (SUNY)

ASIS&T Annual Meeting, Charlotte, NC
October 28 – November 2, 2005
Outline

- Background
  - Similarity and similarity measures in information retrieval
- Feature sets based similarity measures
  - Boolean & Extended Boolean
  - Fuzzy sets model
  - A similarity measure based on Tversky’s Contrast Model
- Concluding remarks
Background

Similarity is a construct that plays significant roles in:
- Cognitive psychology
- Linguistics & language studies
- Sociology & anthropology
- Ecology & biology
- Library & information science
Background: Similarity Measures and Information Retrieval

- Information items content
- Feature extraction
- Information need
- Query formulation
- Structured document representation
- Structured query representation
- Retrieval model: relevance
- Matching
- Retrieval system: efficiency
- Ranked/binary result
Background: Similarity Measures and Image Retrieval

Zachary & Iyengar (2001)
Background: Similarity Measures based on Geometric models

The general form of the distance (called the Minkowski metric) between two points representing stimuli \( a \) and \( b \) on an \( n \)-dimensional space is given by:

\[
D(a, b) = r \sqrt[\bar{r}]{\sum_{i=1}^{n} |a_i - b_i|^r}
\]

Where \( a_i \) and \( b_i \) are corresponding coordinates of stimuli \( a \) and \( b \) on the \( i^{th} \) dimension and \( r \) is a parameter that determines the type of metric.
Background: Similarity Measures and Image Retrieval

Rubner (1999)
Background: Geometric models and Metric Space

Symmetry: \(D(a,b) = D(b,a)\)

Triangle inequality: \(D(a,b) + D(b,c) \geq D(a,c)\)
Background: The Contrast Model

\[ S(a,b) = \theta f(A \cap B) - \alpha f(A-B) - \beta f(B-A) \]

- Distinctive features of \(a\)
- Distinctive features of \(b\)
- Common features of \(a\) & \(b\)
### Background: The Contrast Model

#### Results of a Study

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>B</th>
<th>Beta</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common Features (method1)</td>
<td>1.196</td>
<td>.486</td>
<td>9.505**</td>
</tr>
<tr>
<td>Distinctive Features of a (method1)</td>
<td>-.033</td>
<td>-.197</td>
<td>-4.311**</td>
</tr>
<tr>
<td>Distinctive Features of b (method1)</td>
<td>-.016</td>
<td>-.090</td>
<td>-2.181*</td>
</tr>
<tr>
<td><strong>R² = .45, F(3,431) = 116.101</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Features (method2)</td>
<td>.252</td>
<td>.469</td>
<td>11.136**</td>
</tr>
<tr>
<td>Distinctive Features of a (method2)</td>
<td>-.042</td>
<td>-.163</td>
<td>-3.865**</td>
</tr>
<tr>
<td>Distinctive Features of b (method2)</td>
<td>-.020</td>
<td>-.089</td>
<td>-2.210*</td>
</tr>
<tr>
<td><strong>R² = .30, F(3,431) = 60.423</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01
Feature Sets based Similarity Measures

- **Applied in the following retrieval models:**
  - Boolean model
  - Extended Boolean model
  - Fuzzy sets model
Feature Sets based Similarity Measures – Generalized Contrast Model

\[ S(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)} \]

Where \( \alpha \) & \( \beta \) are non-negative (\( \geq 0 \))

The values of \( S(a, b) \) range from 0 to 1
Summary & Concluding Remarks

- Similarity matching/measure is a key component of information retrieval.
- A similarity measure based on the Contrast Model offers an alternative.
- It gives more weight to common features than distinctive features.
- It takes into account any definition of a feature set and could be applied to any type of document (text, 2D, 3D, moving, etc.).
- It matches perceptual similarity, which is important if we are to bridge the gap between IR systems and users.
Looking Ahead

- Image organization and retrieval research not coordinated
- Need for a common test collection and a retrieval evaluation conference along the lines of the TREC video track
- ASIS&T is an appropriate forum
- Current initiatives http://ir.shef.ac.uk/imageclef/


Thank You!

Questions?