A Computational Model of Similarity

Ichiro Fujinaga

Schulich School of Music

McGill University
A model for similarity

❖ Based on an exemplar-based categorization (classification / identification) strategy; used successfully for:

❖ Optical music recognition
❖ Instrument identification
❖ Genre classification
❖ Modification necessary
Exemplar-based categorization

• The exemplar-based learning model is based on the idea that objects are categorized by their proximity to one or more stored examples.

• There is much evidence from psychological studies to support exemplar-based categorization by humans.

• This model differs both from rule-based or prototype-based (neural nets) models of concept formation in that it assumes no abstraction or generalizations of concepts.

• This model can be implemented using k-nearest neighbour classifiers.
K-nearest neighbour classifier

• Determine the class of a given sample by its feature vector:
  • Distances between feature vectors of an unclassified sample and previously classified samples are calculated
  • The class represented by the majority of k-nearest neighbours (k-NN) is then assigned to the unclassified sample
Examples of similar symbols

<table>
<thead>
<tr>
<th>Symbols</th>
<th>16th</th>
<th>32nd</th>
<th>8th</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>α</td>
<td>α</td>
<td>a</td>
</tr>
<tr>
<td>action</td>
<td>β</td>
<td>β</td>
<td>β</td>
</tr>
<tr>
<td>barline</td>
<td>γ</td>
<td>γ</td>
<td>γ</td>
</tr>
<tr>
<td>c</td>
<td>δ</td>
<td>δ</td>
<td>δ</td>
</tr>
<tr>
<td>caps</td>
<td>ε</td>
<td>ε</td>
<td>ε</td>
</tr>
<tr>
<td>d</td>
<td>ζ</td>
<td>ζ</td>
<td>ζ</td>
</tr>
<tr>
<td>e</td>
<td>η</td>
<td>η</td>
<td>η</td>
</tr>
<tr>
<td>f</td>
<td>θ</td>
<td>θ</td>
<td>θ</td>
</tr>
</tbody>
</table>
Examples of similar symbols
Examples of similar symbols.

III. PERSONAL NARRATIVES

29. The Story of a Navaho Woman Captured by the Utes

Examples of similar symbols
Examples of similar symbols

farmers are not incommodious, in which they enter,
becoming hospitality. The near neighbourhood of
renders larger towns less necessary. The lower ra
ber and decent in their manners, intelligent and tra
with their wealthier neighbours, abundantly ready
bute, according to their ability, to the relief of the
Examples of similar symbols
Example of k-NN classifier
Basketball players and Sumo wrestlers

https://www.flickr.com/photos/29650319@N06/3172412470
Example of k-NN classifier

Classification of athletes by height and weight
(Sumo wrestlers vs NBA basketball players)

- **Height (cm):**
  - 170
  - 179
  - 188
  - 197
  - 206
  - 215

- **Weight (kg):**
  - 75
  - 100
  - 125
  - 150
  - 175
  - 200

- **Sumo**
- **Chicago Bulls**
Example of k-NN classifier
Classifying Michael Jordan

Classification of athletes by height and weight
(Sumo wrestlers vs NBA basketball players)
Example of k-NN classifier
Classifying David Wesley

Classification of athletes by height and weight
(Sumo wrestlers vs NBA basketball players)

http://www.stat-nba.com/image/playerImage/3930.jpg
Example of k-NN classifier
Reshaping the Feature Space

Classification of athletes by height and weight
(Sumo wrestlers vs NBA basketball players)

- **Sumo**
- **Chicago Bulls**
- **Michael Jordan**
- **David Wesley**

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>Weight (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>180</td>
<td>0</td>
</tr>
<tr>
<td>184</td>
<td>140</td>
</tr>
<tr>
<td>188</td>
<td>280</td>
</tr>
<tr>
<td>192</td>
<td>420</td>
</tr>
<tr>
<td>196</td>
<td>560</td>
</tr>
<tr>
<td>200</td>
<td>700</td>
</tr>
</tbody>
</table>
Distance measures

The distance in a $N$-dimensional feature space between two vectors $X$ and $Y$ can be defined as:

$$d = \sum_{i=0}^{N-1} |x_i - y_i|$$

A weighted distance can be defined as

$$d = \sum_{i=0}^{N-1} w_i |x_i - y_i|$$
Features used for instrument recognition

- **Static features (per window)**
  - pitch
  - mass or the integral of the curve (zeroth-order moment)
  - centroid (first-order moment)
  - variance (second-order central moment)
  - skewness (third-order central moment)
  - amplitudes of the harmonic partials
  - number of strong harmonic partials
  - spectral irregularity
  - tristimulus

- **Dynamic features**
  - means and velocities of static features over time
Distance metric

1. **Equal self-similarity**
   \[ d(A, A) = d(B, B) \] for all points A and B

2. **Minimality**
   \[ d(A, B) > d(A, A) \] for all points \( A \neq B \)

3. **Symmetry**
   \[ d(A, B) = d(B, A) \] for all points A and B

4. **Triangle Inequality**
   \[ d(A, B) + d(B, C) \geq d(A, C) \] for all points A, B, and C
Similarity distance?

1. **Equal self-similarity** ×
   \[ d(A, A) = d(B, B) \] for all points A and B

2. **Minimality** ✓
   \[ d(A, B) > d(A, A) \] for all points \( A \neq B \)

3. **Symmetry** ×
   \[ d(A, B) = d(B, A) \] for all points A and B

4. **Triangle Inequality** ×
   \[ d(A, B) + d(B, C) \geq d(A, C) \] for all points A, B, and C
Unequal self-similarity and asymmetry of similarity

• Examples

  • Unequal self-similarity
    • Octave intervals are more similar to other octave intervals than major 6th (Balzano 1977)

  • Asymmetry of similarity
    • North Korea is more similar to China than China is to North Korea (Tversky 1977)
    • Preferred “103 is virtually 100” than “100 is virtually 103” (Rosch 1973)

• Solutions exist

  • Krumhansl (1978) and Nosofsky (1991): by considering the densities of exemplars of each category
Triangle Inequality?

Triangle Inequality
\[ d(A, B) + d(B, C) \geq d(A, C) \]
for all points A, B, and C

(Between two points, a straight line is always the shortest way)

- An example
  - “A flame is similar to the moon because they both appear luminous, and the moon is similar to a ball because they are both round. But a flame and a ball are not similar.” (James 1890)
  
  \[ d(\text{flame, moon}) + d(\text{moon, ball}) \geq d(\text{flame, ball})? \]

- Solution?
A computational model for similarity is possible based on k-nearest neighbour classifier with special distance functions.

It is basically a geometrical model with variable feature weights.

"Toward a Unified Theory of Similarity and Recognition" (Ashby & Perrin 1988)

Can emulate Tversky’s model (using binary weights) and the transformational model (Imai 1977) (using appropriate distance function, such as the Earth-mover’s distance).

Can deal with Global vs Dimensional (Implicit vs Explicit) similarity.

- Bag of features vs pre-selected set of features

But, what use is it?

- Extremely context dependent
- A model without predictive powers
- Maybe useful for specific tasks