Non-Euclidean Problems in Pattern Recognition

ACIT 2013, Khartoum, Sudan, 17-19 Dec 2013

Robert P.W. Duin, Delft University of Technology
(In cooperation with Elżbieta Pękalska, Univ. of Manchester)

Pattern Recognition Lab
Delft University of Technology, The Netherlands
http://rduin.nl
How to represent real-world objects, (with a size and a shape) given a set of examples such that we can generalize?

Real world objects and events

Images
Spectra
Time signals
Gestures

How to build a representation?
Features \leftrightarrow Structure

Blob Recognition

446 binary images, varying size, e.g.: 100 x 130

Shape classification by weighted-edit distances (Bunke)
**Colon Tissue Recognition**

- ???
- normal
- pathological

**Volcano / Seismic Signal Classification**

- 150,000 events (1994 – 2008)
- 5 volcanos
- 40 stations
- 15 classes

*J. Makario, INGEOMINAS, Manizales, Colombia*

*M. Orozco-Alzate, Nat. Univ. Colombia, Manizales*

*R. Duin, TUDelft*

*M. Bicego, Univ. of Verona, Italy*

*Cenatav, Havana, Cuba*

**Gesture Recognition**

- Is this gesture in the database?

**Pattern Recognition System**

- Sensor → Representation → Generalization

- Feature Representation
  - area
  - perimeter

* TU Delft *)
Feature Representation

Objects → points in a Euclidean Space
Features reduce → classes overlap
→ to be solved by statistics

Compactness

Representations of real world similar objects are close.
There is no ground for any generalization (induction) on representations that do not obey this demand.

True Representations

Similar objects are close
and
Dissimilar objects are distant.

→ no probabilities needed, domains are sufficient!

Dissimilarities → True Representation

Structural Representation

Strings

$X = (x_1, x_2, ...., x_k)$

$Y = (y_1, y_2, ...., y_n)$

Graphs

How to generalize? Distances!

Dissimilarities
Dissimilarities

Objects → Shape distances
Objects → Features → Euclidean distances
Objects → Graphs → Graph distances

Examples Dissimilarity Measures

Dist(A,B): 
\[ a \in A, \text{ points of } A \]
\[ b \in B, \text{ points of } B \]
\[ d(a,b): \text{ Euclidean distance} \]

\[ D(A,B) = \max_a\{\min_b\{d(a,b)\}\} \]
\[ D(B,A) = \max_b\{\min_a\{d(b,a)\}\} \]

Hausdorff Distance (metric):
\[ DH = \max\{\max_a\{\min_b\{d(a,b)\}\} , \max_b\{\min_a\{d(b,a)\}\}\} \]

Modified Hausdorff Distance (non-metric):
\[ DM = \max\{\text{mean}_a\{\min_b\{d(a,b)\}\} , \text{mean}_b\{\min_a\{d(b,a)\}\}\} \]

Dissimilarities – Possible Assumptions

1. Positivity: \[ d_{ij} \geq 0 \]
2. Reflexivity: \[ d_{ii} = 0 \]
3. Definiteness: \[ d_{ij} = 0 \text{ iff objects } i \text{ and } j \text{ are identical} \]
4. Symmetry: \[ d_{ij} = d_{ji} \]
5. Triangle inequality: \[ d_{ij} < d_{ik} + d_{kj} \]
6. Compactness: if the objects \( i \) and \( j \) are very similar then \[ d_{ij} < \delta \]
7. True representation: if \( d_{ij} < \delta \) then the objects \( i \) and \( j \) are very similar.
8. Continuity of \( d \).
**Non-metric distances**

- Weighted-edit distance for strings
  - object 78
  - object 419
  - object 425
  - Bunke's Chicken Dataset

- Single-linkage clustering
  - \[ D(A,C) > D(A,B) + D(B,C) \]

- Fisher criterion
  - \[ J(A,B) = \frac{\left| \mu_A - \mu_B \right|^2}{\sigma_A^2 + \sigma_B^2} \]
  - \[ J(A,C) = 0 \quad J(A,B) = \text{large} \]
  - \[ J(C,B) = \text{small} \neq J(A,B) \]

---

**Intrinsic Non-Euclidean Dissimilarity Measures**

**Single Linkage**

- Distance(Table, Book) = 0
- Distance(Table, Cup) = 0
- Distance(Book, Cup) = 1

- Single-linkage clustering
  - \[ D(A,C) > D(A,B) + D(B,C) \]

---

**Alternatives for the Nearest Neighbor Rule**

1. Dissimilarity Space
2. Embedding

Alternative 1: Dissimilarity Space

Dissimilarities

Given labeled training set

Unlabeled object to be classified

Selection of 3 objects for representation

Prototype Selection: Polygon Dataset

The classification error as a function of the number of selected prototypes. For 10-20 prototypes, results are already better than by using 1000 objects in the NN rules.


Alternative 2: Embedding

Training set

Dissimilarity matrix D → X

Is there a feature space for which Dist(X,X) = D?

Position points in a vector space such that their Euclidean distances → D

Embedding

13 x 13 Distance Matrix
(Pseudo-)Euclidean Embedding

\[ m \times m \text{D} \text{ is a given, imperfect dissimilarity matrix of training objects.} \]

Construct inner-product matrix:
\[ B = -\frac{1}{2} JD^{(2)} J \]
\[ J = I - \frac{1}{m} 11 \]

Eigenvalue Decomposition,
\[ B = Q \Lambda Q^T \]

Select \( k \) eigenvectors:
\[ X = Q_k \Lambda_k^+ \]
(problem: \( \Lambda_k < 0 \))

Let \( S_k \) be a \( k \times k \) diag. matrix,
\[ S_k(i,i) = \text{sign} (\Lambda_k(i,i)) \]
\[ \Lambda_k(i,i) < 0 \rightarrow \text{Pseudo-Euclidean} \]

\( m \times m \text{D} \) is the dissimilarity matrix between new objects and the training set.

The inner-product matrix:
\[ Z = B_k Q_k \Lambda_k^+ S_k \]

Distances in PES

\[ d^2(O, A) > 0 \]
\[ d^2(O, E) > 0 \]
\[ d^2(O, B) = 0 \]
\[ d^2(O, D) < 0 \]

All points in the grey area are closer to \( O \) than \( O \) itself!?

Any point has a negative square distance to some points on the line \( v^T x = 0 \).

Can it be used as a classifier? Can we define a margin as in the SVM?

PES: Pseudo-Euclidean Space (Krein Space)

If \( D \) is non-Euclidean, \( B \) has \( p \) positive and \( q \) negative eigenvalues.

A pseudo-Euclidean space \( \mathcal{E} \) with signature \((p,q)\), \( k = p+q \), is a non-degenerate inner product space \( \mathbb{R}_k = \mathbb{R}_p \oplus \mathbb{R}_q \) such that:
\[ \langle x, y \rangle_\epsilon = x^T S_p q y = \sum_{i=1}^p x_i y_i - \sum_{j=p+1}^q x_j y_j \]
\[ S_p q = \begin{bmatrix} I_{p \times p} & 0 \\ 0 & -I_{q \times q} \end{bmatrix} \]
\[ d^2_p(x, y) = (x - y, x - y)_\epsilon = d^2_p(x, y) - d^2_q(x, y) \]

Pseudo Euclidean Space

Euclidean embedding \( D \rightarrow X \)
\[ d^2_{ij} = \| x_i - x_j \|^2 \]

Pseudo Euclidean embedding \( D \rightarrow \{ X^p, X^q \} \)
\[ d^2_{ij} = \| x_i^p - x_j^p \|^2 - \| x_i^q - x_j^q \|^2 \]

'Positive' and 'negative' space,
Compare Minkowsky space in relativity theory.
**PE-Space classifiers**

- kNN, Parzen, Nearest Mean
  As object distances can be computed (are known)

- LDA, QDA
  As PE inner possibly product definitions cancel they can be computed, interpretation ... ?

- SVM
  May get a result (indefinite kernel), possibly not optimal

- Others ??

**Examples Dissimilarity Measures**

Matching new objects to various templates:

\[
\text{class}(x) = \text{class}(\arg \min_y (D(x,y)))
\]

Dissimilarity measure appears to be non-metric.

**Three Approaches Compared for the Zongker Data**

Dissimilarity Space equivalent to Embedding better than Nearest Neighbour Rule

**Representation Strategies**

Avoiding the PE space

Dissimilarity Space: \( X = D \)

Correcting

Associated space \( X = \{[X_p, X_q], \emptyset\} \)

\[
\tilde{d}_{ij}^2 = d_p^2(x_i, x_j) + d_q^2(x_i, x_j)
\]

Positive space \( X = X_p \)

\[
\tilde{d}_i^2 = d_i^2(x_i, x_i)
\]

Negative space \( X = X_q \)

\[
\tilde{d}_i^2 = d_i^2(x_i, x_i)
\]

Additive Correction \( \tilde{d}_i^2 = d_i^2 + c, i \neq j \) \( X = \text{Embedding}(\tilde{D}) \)

As it is

Pseudo Euclidean Space \( X = \{X_p, X_q\} \)

\[
\tilde{d}_i^2 = d_p^2(x_i, x_i) - d_q^2(x_i, x_i)
\]

Classifiers to be developed further


Ball Distances

- Generate sets of balls (classes) uniformly, in a (hyper)cube; not intersecting.
- Balls of the same class have the same size.
- Compute all distances between the ball surfaces.

$\rightarrow$ Dissimilarity matrix $D$

---

**Balls3D**

<table>
<thead>
<tr>
<th>Classifier</th>
<th>PE Sp</th>
<th>Ass Sp</th>
<th>Pos Sp</th>
<th>Neg Sp</th>
<th>Cor Sp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN</td>
<td>47.4 (2.0)</td>
<td>47.4 (2.0)</td>
<td>47.4 (2.0)</td>
<td>44.2 (1.5)</td>
<td>47.4 (2.0)</td>
</tr>
<tr>
<td>Parzen</td>
<td>45.7 (1.7)</td>
<td>45.5 (1.6)</td>
<td>45.6 (1.7)</td>
<td>35.5 (1.7)</td>
<td>45.7 (1.7)</td>
</tr>
<tr>
<td>NM</td>
<td>47.5 (2.0)</td>
<td>47.7 (2.0)</td>
<td>47.6 (1.9)</td>
<td>49.6 (0.2)</td>
<td>48.1 (1.8)</td>
</tr>
<tr>
<td>SVM-1</td>
<td>50.7 (2.2)</td>
<td>50.0 (2.7)</td>
<td>50.0 (2.5)</td>
<td>62.1 (1.7)</td>
<td>50.1 (2.0)</td>
</tr>
</tbody>
</table>

10 x (2-fold crossvalidation of 50 objects per class)

---

Is the PE Space Informative?

- Chickenpieces35: 4.6 5 0 0.156 0.791
- Chickenpieces50: 4.6 5 0 0.162 0.791
- Chickenpieces90: 4.6 5 0 0.152 0.791
- Chickenpieces120: 4.6 5 0 0.130 0.791

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Original $D_2$</th>
<th>Positive $D_P$</th>
<th>Negative $D_N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-NN</td>
<td>0.022 0.132 0.178</td>
<td>0.020 0.067 0.173</td>
<td>0.022 0.052 0.148</td>
</tr>
<tr>
<td>Parzen</td>
<td>0.034 0.108 0.148</td>
<td>0.036 0.073 0.142</td>
<td>0.036 0.073 0.142</td>
</tr>
<tr>
<td>NM</td>
<td>0.025 0.123 0.171</td>
<td>0.023 0.078 0.171</td>
<td>0.023 0.078 0.171</td>
</tr>
<tr>
<td>SVM-1</td>
<td>0.025 0.123 0.171</td>
<td>0.023 0.078 0.171</td>
<td>0.023 0.078 0.171</td>
</tr>
</tbody>
</table>

Examples
Example: Chickenpieces (H. Bunke, Bern)

BACK
BREAST
DRUMSTICK
THIGH-AND-BACK
WING

446 binary images, varying size, e.g.: 100 x 130


Shape classification by weighted-edit distances (Bunke)

Best classification result is for a very non-Euclidean dissimilarity measure!

Flow Cytometry

612 histograms in 3 classes

Nap & van Rodijnen, Atrium Hospital, Heerlen

Flow Cytometry: classification errors

Pairwise, horizontal (intensity calibration):
\[ D(\text{hist}_1,\text{hist}_2) = \min_{i} L_1(\text{hist}_1,\text{hist}_2(x)) \]

<table>
<thead>
<tr>
<th>Data Source</th>
<th>NEF</th>
<th>1-NN</th>
<th>1-NND</th>
<th>SVM-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tube 1</td>
<td>0.27</td>
<td>0.38</td>
<td>0.38</td>
<td>0.30</td>
</tr>
<tr>
<td>Tube 2</td>
<td>0.27</td>
<td>0.37</td>
<td>0.37</td>
<td>0.29</td>
</tr>
<tr>
<td>Tube 3</td>
<td>0.27</td>
<td>0.38</td>
<td>0.40</td>
<td>0.27</td>
</tr>
<tr>
<td>Tube 4</td>
<td>0.27</td>
<td>0.42</td>
<td>0.42</td>
<td>0.30</td>
</tr>
<tr>
<td>Averaged</td>
<td>0.24</td>
<td>0.27</td>
<td>0.20</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Bio-crystallization

image size: 2114 x 2114
Different food products / quality
2 classes, 54 examples/class

Busscher et al., Standardization of the incrystallization Method for Carrot Samples, Biological Agriculture and Horticulture, 2010, Vol. 27, pp. 1–23

Bio-crystallization: Dissimilarity Measures

<table>
<thead>
<tr>
<th>Dissimilarity Measure</th>
<th>NEF</th>
<th>1-NN</th>
<th>1-NND</th>
<th>SVM-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gauss</td>
<td>0</td>
<td>0.329</td>
<td>0.266</td>
<td>0.106</td>
</tr>
<tr>
<td>Laplace</td>
<td>0</td>
<td>0.229</td>
<td>0.313</td>
<td>0.125</td>
</tr>
<tr>
<td>Laplace Histogram</td>
<td>0.067</td>
<td>0.107</td>
<td>0.172</td>
<td>0.072</td>
</tr>
<tr>
<td>Averaged Dissimilarities</td>
<td>0.004</td>
<td>0.114</td>
<td>0.166</td>
<td>0.057</td>
</tr>
</tbody>
</table>

Bio-crystallization: classification errors

Gesture Recognition

Is this gesture in the database?
**Gesture Recognition**

- **20 signs (classes), 75 examples/sign**
- **Distance measure: DTW**

**Application: Graphs**

- **x1 x2 x3 x4 x5**
- **Graph with feature nodes**

**Interpolating structural and feature space dissimilarities**

- **Feature nodes**
- **Structure only (no features)**
- **Features only (no structure)**

**Conclusion**
Non-Euclidean Representations

- Why do we have them?
- Are they essential?
- Can we build classifiers for them? (to some extent)
- Can we transform them into Euclidean representations? (Yes, but at the cost of performance loss)

Computational Noise

Even for Euclidean distance matrices zero eigenvalues may show negative, e.g:
- $X = \text{N}(50, 20)$: 50 points in 20 dimensions
- $D = \text{Dist}(X)$: 50 x 50 distance matrix
- Expected: $49 - 20 = 29$ zero eigenvalues
- Found: 15 negative eigenvalues

Computational Problems

Large distances are overestimated due to computational problems

Weighted edit distance for strings

Intrinsically Non-Euclidean Dissimilarity Measures Invariants

Object space

D(A,C) > D(A,B) + D(B,C)
Non-metric object distances due to invariants
Boundary distances

A set of boundary distances may characterize sets of datapoints: Distances → features

Intrinsicly Non-Euclidean Dissimilarity Measures
Mahalanobis

Pairwise comparison between differently shaped data distributions

Different pairs → different comparison frameworks
→ non-Euclidean

Conclusions

- Real world objects are not points
- Objects have a size
- Relations are non-Euclidean
- Non-Euclidean generalization procedures are needed