Comparing Dissimilarity Representations

Adam Cardinal-Stakenas

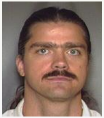
Department of Applied Mathematics & Statistics
Johns Hopkins University

Interface, Durham, NC

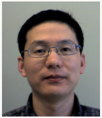
Comparing Dissimilarity Representations

Adam Cardinal-Stakenas

Department of Applied Mathematics & Statistics
Johns Hopkins University



Carey E. Priebe



Zhiliang Ma

Youngser Park

Interface, Durham, NC

The Problem: Classification

Setup:

$(X_i, Y_i) \overset{i.i.d}{\sim} F_{XY}$ with $X_i : \Omega \to \Xi$, and $Y_i : \Omega \to \{1, \cdots, J\}$.

The Problem: Classification

Setup:

$(X_i, Y_i) \overset{i.i.d}{\sim} F_{XY}$ with $X_i : \Omega \to \Xi$, and $Y_i : \Omega \to \{1, \cdots, J\}$.

Examples:

- Image analysis
- Text analysis
- Graph data
- Computational anatomy

The Problem: Classification

Setup:

$(X_i, Y_i) \overset{i.i.d}{\sim} F_{XY}$ with $X_i : \Omega \rightarrow \Xi$, and $Y_i : \Omega \rightarrow \{1, \cdots, J\}$.

Examples:

- Image analysis
- Text analysis
- Graph data
- Computational anatomy

Goal:

To obtain superior performance in the exploitation task using comparisons of the data.

Dissimilarities

Definition

A dissimilarity measure is a function $\delta : \Xi \times \Xi \rightarrow \mathbb{R}^+ \cup \{0\}$ with:

1. Positivity: $\delta(x_1, x_2) \geqslant 0$,
2. Identifiability: $\delta(x_1, x_2) = 0 \Rightarrow x_1 = x_2$.
3. Reflexivity: $\delta(x, x) = 0$,
4. Symmetry: $\delta(x_1, x_2) = \delta(x_2, x_1)$,

Dissimilarities

Definition

A dissimilarity measure is a function $\delta : \Xi \times \Xi \to \mathbb{R}^+ \cup \{0\}$ with:

1. Positivity: $\delta(x_1, x_2) \geqslant 0$,
2. Identifiability: $\delta(x_1, x_2) = 0 \Rightarrow x_1 = x_2$.
3. Reflexivity: $\delta(x, x) = 0$,
4. Symmetry: $\delta(x_1, x_2) = \delta(x_2, x_1)$,

A dissimilarity representation for a set of n objects is expressed as a symmetric and hollow matrix Δ with positive off-diagonal entries.

Dissimilarities

Definition

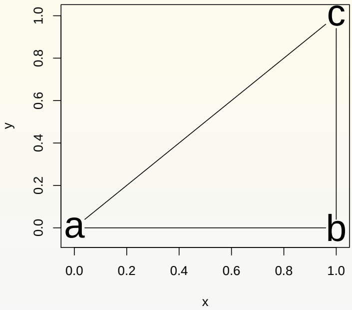
A dissimilarity measure is a function $\delta : \Xi \times \Xi \rightarrow \mathbb{R}^+ \cup \{0\}$ with:

1. Positivity: $\delta(x_1, x_2) \geqslant 0$,
2. Identifiability: $\delta(x_1, x_2) = 0 \Rightarrow x_1 = x_2$.
3. Reflexivity: $\delta(x, x) = 0$,
4. Symmetry: $\delta(x_1, x_2) = \delta(x_2, x_1)$,

A dissimilarity representation for a set of n objects is expressed as a symmetric and hollow matrix Δ with positive off-diagonal entries.

We may observe numerous dissimilarities. That is, for a given Ξ, we may have available or be able to calculate, $\delta_1, \delta_2, \ldots, \delta_k$ where $\delta_i : \Xi \times \Xi \rightarrow \mathbb{R}$ is a dissimilarity function for $i \in \{1, \ldots, l\}$

Example: $\mathbb{R}^2$



Manhattan Distance:

$$\delta_{\text{Man.}} = |x_1 - x_2| + |y_1 - y_2|$$

Euclidean Distance:

$$\delta_{\text{Euc.}} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Minkowsi Distance:

$$\delta_{\text{Min. p}} = \sqrt[p]{|x_1 - x_2|^p + |y_1 - y_2|^p}$$

Supremum Distance:

$$\delta_{\text{Sup.}} = \max\{|x_1 - x_2|, |y_1 - y_2|\}$$

δ	$\delta(a,b)$	$\delta(b,c)$	$\delta(a,c)$
Man.	1	1	2
Euc.	1	1	1.41
Min. p=3	1	1	1.26
Sup.	1	1	1

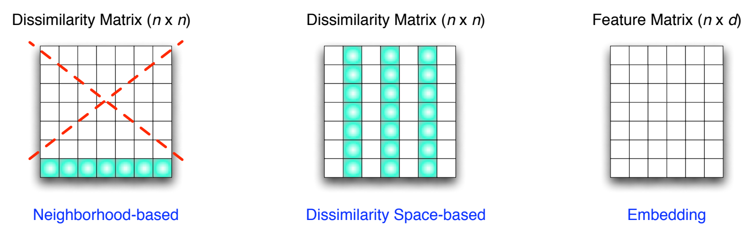
Introduction
- The Statistical Inference Problem
- Dissimilarities

Approaches
- Framework
- Multiple Dissimilarities
- Motivation

Examples
- Image/Caption Fusion

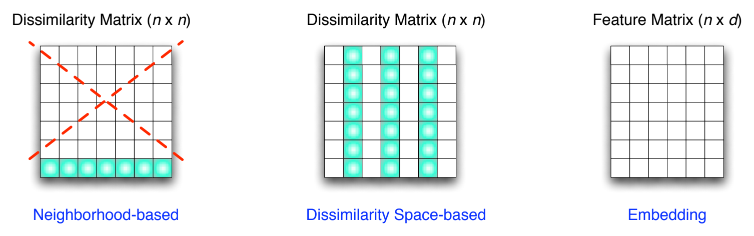
Framework

Three dissimilarity classification approaches



Framework

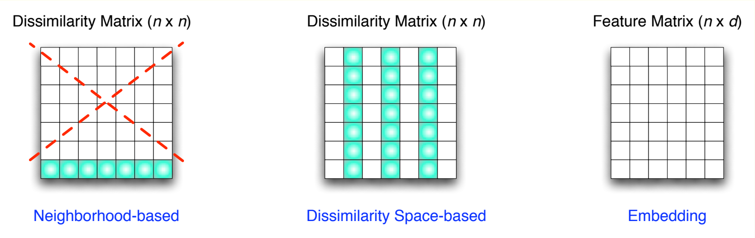
Three dissimilarity classification approaches



1. The *neighborhood-based approach* interprets dissimilarities as neighborhood relations.

Framework

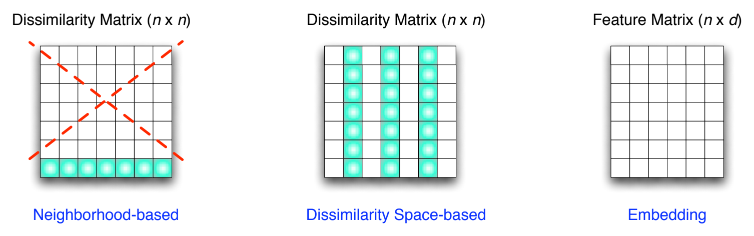
Three dissimilarity classification approaches



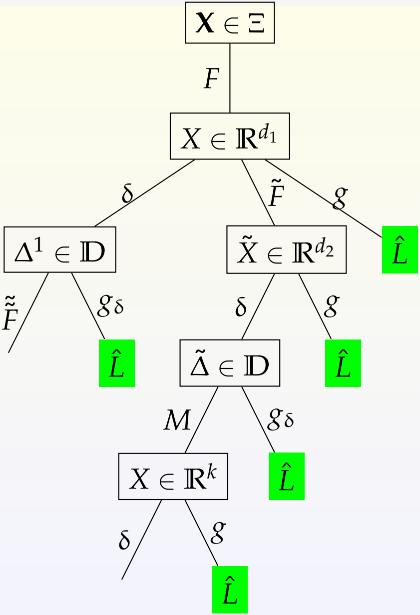
1. The *neighborhood-based approach* interprets dissimilarities as neighborhood relations.
2. The *dissimilarity space approach* defines a representation set $R = \{p_1, \cdots, p_r\}$, and interprets dissimilarities from a point to each element of the representation set as features of this point.

Framework

Three dissimilarity classification approaches



1. The *neighborhood-based approach* interprets dissimilarities as neighborhood relations.
2. The *dissimilarity space approach* defines a representation set $R = \{p_1, \cdots, p_r\}$, and interprets dissimilarities from a point to each element of the representation set as features of this point.
3. The *embedding approach* embeds dissimilarities into $\mathbb{R}^d$ in such a way that the configuration's interpoint distances approximate the original dissimilarities.



$F, \tilde{F}, \tilde{\tilde{F}}$: feature extraction procedures

$\delta \in \{\delta_1, \ldots, \delta_k\}$: dissimilarity function

g : classifer

g_δ : dissimilarity based classifier

Comparing Dissimilarities

- To choose a dissimilarity, a classifier-independent methodology for evaluating a dissimilarity representation is needed.

Comparing Dissimilarities

- To choose a dissimilarity, a classifier-independent methodology for evaluating a dissimilarity representation is needed.
- To achieve this, measure the congruence from an observed dissimilarity matrix to Δ_Y, the ideal dissimilarity matrix (from the training data).

Comparing Dissimilarities

- To choose a dissimilarity, a classifier-independent methodology for evaluating a dissimilarity representation is needed.
- To achieve this, measure the congruence from an observed dissimilarity matrix to Δ_Y, the ideal dissimilarity matrix (from the training data).
- We define congruence using the multidimensional scaling criterion function [Borg & Groenen]:

$$\mu : \mathbb{D} \times \mathbb{D} \to [0, 1], \qquad \mu(d(\boldsymbol{X}), \delta) = \frac{\sum_{i<j} d(X_i, X_j)\delta_{ij}}{\sqrt{\sum_{i<j} d(X_i, X_j)^2 \sum_{i<j} \delta_{ij}^2}}.$$

Comparing Dissimilarities

- To choose a dissimilarity, a classifier-independent methodology for evaluating a dissimilarity representation is needed.
- To achieve this, measure the congruence from an observed dissimilarity matrix to Δ_Y, the ideal dissimilarity matrix (from the training data).
- We define congruence using the multidimensional scaling criterion function [Borg & Groenen]:

$$\mu : \mathbb{D} \times \mathbb{D} \to [0,1], \qquad \mu(d(\boldsymbol{X}), \delta) = \frac{\sum_{i<j} d(X_i, X_j)\delta_{ij}}{\sqrt{\sum_{i<j} d(X_i, X_j)^2 \sum_{i<j} \delta_{ij}^2}}.$$

- We calculate $c(\Delta) = \mu(\Delta, \Delta_Y)$.

Comparing Dissimilarities

- To choose a dissimilarity, a classifier-independent methodology for evaluating a dissimilarity representation is needed.
- To achieve this, measure the congruence from an observed dissimilarity matrix to Δ_Y, the ideal dissimilarity matrix (from the training data).
- We define congruence using the multidimensional scaling criterion function [Borg & Groenen]:

$$\mu : \mathbb{D} \times \mathbb{D} \to [0,1], \qquad \mu(d(\boldsymbol{X}), \delta) = \frac{\sum_{i<j} d(X_i, X_j)\delta_{ij}}{\sqrt{\sum_{i<j} d(X_i, X_j)^2 \sum_{i<j} \delta_{ij}^2}}.$$

- We calculate $c(\Delta) = \mu(\Delta, \Delta_Y)$.
- Larger values indicate stronger correlation to the ideal dissimilarity.

Motivation

- In the style of Devroye, Györfi, and Lugosi, the typical comparison measure for a classification problem involve calculating some measurement based on the posterior probabilities. For example

Motivation

- In the style of Devroye, Györfi, and Lugosi, the typical comparison measure for a classification problem involve calculating some measurement based on the posterior probabilities. For example
  - $L^* = E\{\min(\eta(X), 1 - \eta(X))\}$

Motivation

- In the style of Devroye, Györfi, and Lugosi, the typical comparison measure for a classification problem involve calculating some measurement based on the posterior probabilities. For example
  - $L^* = E\{\min(\eta(X), 1 - \eta(X))\}$
  - $L_{NN} = E\{2\eta(X)(1 - \eta(X))\}$

Motivation

- In the style of Devroye, Györfi, and Lugosi, the typical comparison measure for a classification problem involve calculating some measurement based on the posterior probabilities. For example
  - $L^* = E\{\min(\eta(X), 1-\eta(X))\}$
  - $L_{NN} = E\{2\eta(X)(1-\eta(X))\}$
  - $\rho = E\left\{\sqrt{\eta(X)(1-\eta(X)}\right\}$

Motivation

- In the style of Devroye, Györfi, and Lugosi, the typical comparison measure for a classification problem involve calculating some measurement based on the posterior probabilities. For example
  - $L^* = E\{\min(\eta(X), 1-\eta(X))\}$
  - $L_{NN} = E\{2\eta(X)(1-\eta(X))\}$
  - $\rho = E\left\{\sqrt{\eta(X)(1-\eta(X)}\right\}$

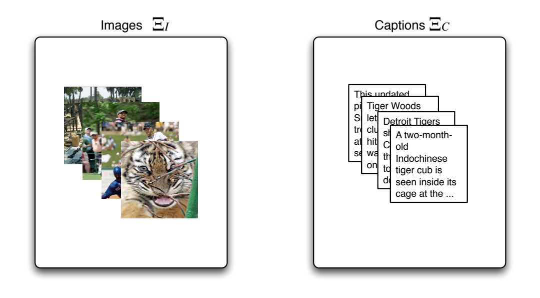
- The congruence coefficient is classifier independent in the sense that it only depends on the data, however it is not (obviously) asymptotically related to the posteriors.

Image/Caption Fusion

Motivation

Example: Image/Caption Fusion:

- Images and text are most naturally considered marginally.
- The geometries of image spaces and caption spaces are not well understood.
- Our fusion methodologies allow for subsequent "joint" analyses.



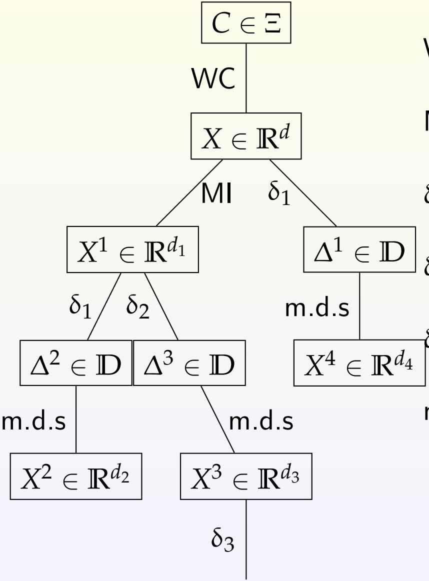
Image/Caption Fusion

The data

- 140,577 *images* & *captions* were collected from Yahoo! Photos website
- 1,600 pairs were selected using query word "tiger" on captions.
- They were labeled manually based only on captions:

label	**#**
animal tiger	148
Detroit Tigers baseball team	145
Tiger Woods the golfer	897
Tamil Tigers soldiers of Sri Lanka	330
Leicester Tigers rugby team	48
others	32

- Two class problem: "Tiger Woods" and "Tamil Tigers".



WC: Word count histogram

MI: Mutual Information

δ_1 =Euclidecan distance

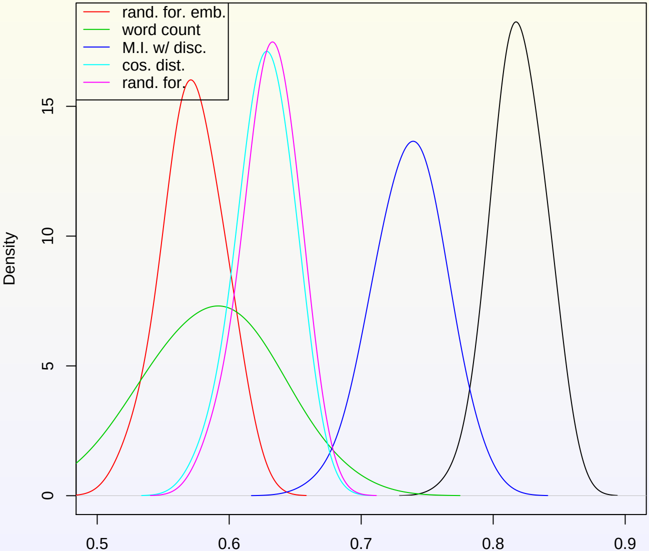
δ_2 =cosine distance

δ_3 =random forests prox.

m.d.s.=multidimensional scaling

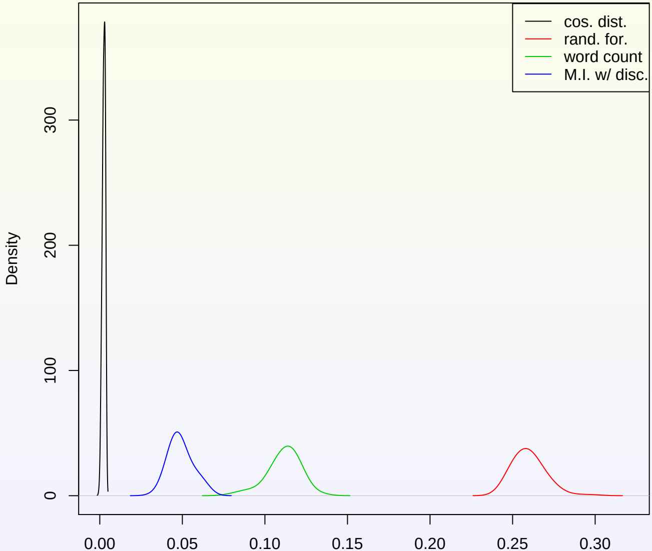
Comparing Dissimilarities

c-function: comparison of various caption dissimilarities



Comparing Dissimilarities

Estimated probability of misclassification



Conclusion and Future Work

- We can evaluate how closely a particular dissimilarity representation matches the dissimilarities between the class labels.
- This measure relates to classifier performance, but it does not depend on the use of a particular classifier.
- More work is needed to investigate how congruence coefficient changes with the dimensionality of the data.
- To improve the performace of the congruence coefficient for predicting classifier performace we wish to investigate tailoring the “ideal” dissimilarity to the observed data.

Questions?

Adam Cardinal-Stakenas
Department of Applied Mathematics and Statistics
Johns Hopkins University
adam_cs@ams.jhu.edu