On the Informativeness of Asymmetric Dissimilarities

Yenisel Plasencia Calaña, Veronika Cheplygina, Robert P.W. Duin, Edel B. García Reyes, Mauricio Orozco Alzate, David M.J. Tax, Marco Loog

Simbad 2013, York, UK

Asymmetric dissimilarities

""Italy is similar to France""

1270 hits

"France is similar to Italy"

63 hits

Contents

- Causes of asymmetry
- What people do about it
- How we can do it better

Causes
Experts

	Italy	France
Italy		63
France	1270	

Causes
Matching

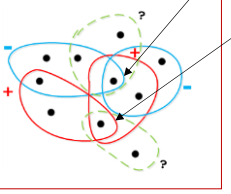
- Shapes = Strings of angles
- Incorporating invariance
- Inexact procedures
- Different search path

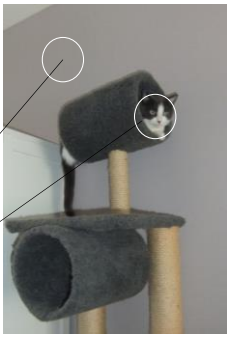
{90, 90, 0, 135, 135…}
starting at top left

{225, 225, 0, 270, 270…}
starting at arrow tip

Causes
Multiple Instance Learning

- Bags are **sets** of feature vectors
- Instance labels are not given
- **Concept instances** important for bag label





Causes

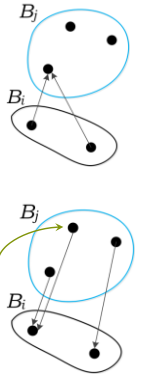
Multiple Instance Learning

- (Modified) Hausdorff distance

$$d(B_i, B_j) = \max_{\mathbf{x}_k \in B_i} \min_{\mathbf{x}_l \in B_j} d(\mathbf{x}_k, \mathbf{x}_l)$$

$$d_H(B_i, B_j) = \max(d(B_i, B_j), d(B_j, B_i))$$

Concept instance?



Asymmetry

What is usually done

- Nearest neighbor, clustering, SVM, embedding …
- Symmetrize dissimilarity
- *What if this throws away information?*

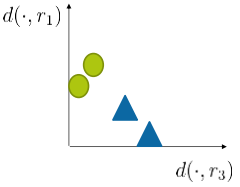
$$D_1 = d(A, B)$$
$$D_2 = d(B, A)$$
$$\max(D_1, D_2)$$
$$\min(D_1, D_2)$$
$$\frac{1}{2}(D_1 + D_2)$$

Asymmetry

Alternative

- **Dissimilarity space**
- Distances as features
- Any (even non-square) matrix
- Any supervised classifier

$$R = \{r_1, \ldots, r_k\}$$
$$D_i \in \mathbb{R}^k, i = 1, 2$$



Asymmetry

Our idea

- **E**xtended **A**symmetric **D**issimilarity **S**pace
- *All information available*

$$[D_1 D_2] \in \mathbb{R}^{k \times 2}$$

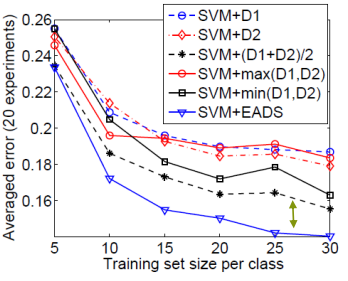
Datasets

- **Chicken Pieces**
  - 5 classes
  - Edge segments of 35 pixels
  - String of angles
  - Edit distance
- **Ajax Orange**
  - 25 objects, 10x6 backgrounds
  - 1 against all
  - Segments + simple features
  - Modified Hausdorff

Results

Chicken pieces

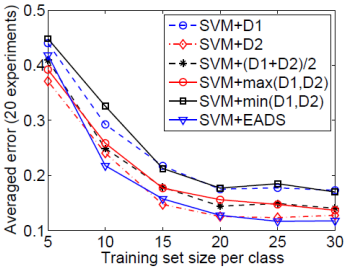
- D1 and D2 comparable
- EADS best, Averaging good
- Gap EADS & average smaller with more prototypes



Results

AjaxOrange

- D2 much better than D1 (Concept instances!)
- Averaging worse than best!
- EADS still good



Conclusions

- Asymmetric dissimilarities occur in many applications
- Asymmetry can be informative
- Dissimilarity space suitable for using asymmetry information
- Preserve most information with
  Extended **A**symmetric **D**issimilarity **S**pace

Do you also have asymmetric dissimilarities?