International Conference
DIGITAL PRESERVATION AND PRESENTATION OF CULTURAL AND SCIENTIFIC HERITAGE
DiPP 2011
11-14 September 2011
Veliko Tarnovo, Bulgaria

Applying Associative Classifier PGN for Digitised Cultural Heritage Resource Discovery

Krassimira Ivanova, Iliya Mitov, Peter Stanchev
Milena Dobreva, Koen Vanhoof, Benoit Depaire

The results born from:

Crosspoint of the work of three projects:

- Project D002-308 "Automated Metadata Generating for e-Documents Specifications and Standards"
  (Bulgarian National Science Fund)

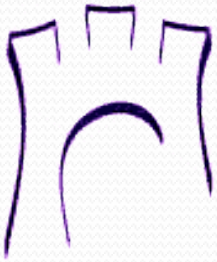
  - metadata

- Project R-1875 "Search in Art Image Collections Based on Color Semantics"
  (Hasselt University)

  - automatic extraction of visual descriptors from art images (APICAS)

- Project R-1876 "Intelligent Systems' Memory Structuring Using Multidimensional Numbered Information Spaces"
  (Hasselt University)

  - classifier PGN (DM system PaGaNe)

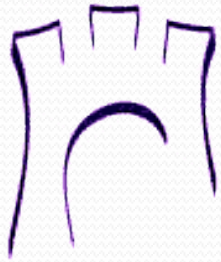


Associative Classifiers

- Arose from Associative-rule-miners (set with transactions – searching frequent combinations)
- Expansion: each transaction contains all features of an object in addition to the class label of the object.
  (such association rules are named "class association rules")

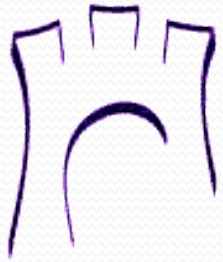
Generally the structure of associative classifiers consists of three major data mining steps:

1. Association rule mining
2. Pruning (optional)
3. Recognition



PGN Classifier

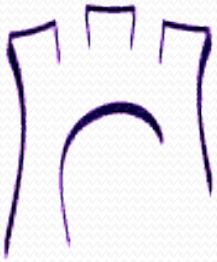
- Typically:
  first look support of association rule,
  after that confidence.
- We question this common approach which prioritizes support over confidence.
- We study a new associative classifier algorithm, which turns the priorities around and focuses on confidence first by retaining only 100% confidence rules.
- The main goal of was to verify the quality of the confidence-first concept.



PGN: Short Description

- In classical associative classifiers user must give the support and confidence level.
- One of the main specifics of PGN is that it is a parameter free classifier.

- The association rule mining goes from longest rules (instances) to the shorter ones until no intersections between patterns in the classes are possible.
- In the pruning phase the contradictions and inconsistencies of more general rules are cleared, after that the pattern set is compacted excluding all more concrete rules within the classes.

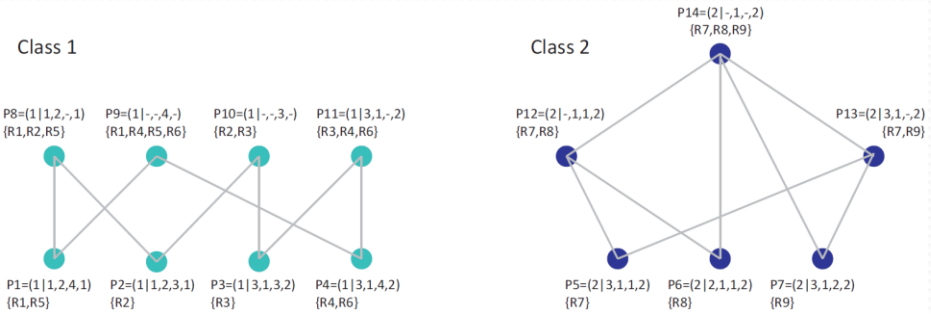


PGN: Training process

Association rule mining

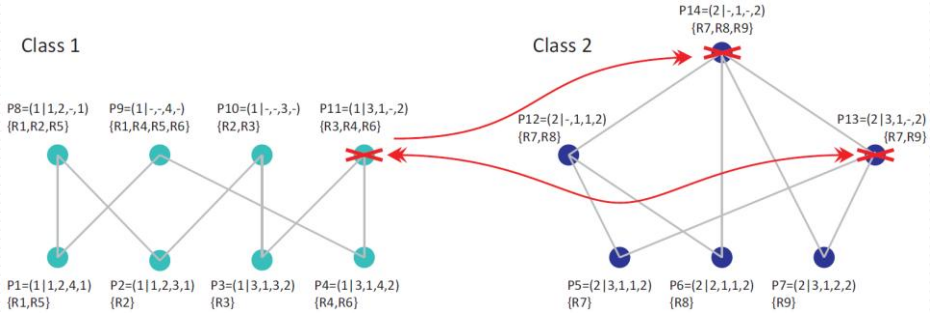
1.Adding instances to the pattern set.

2.Creating all possible intersection patterns between patterns within the class.

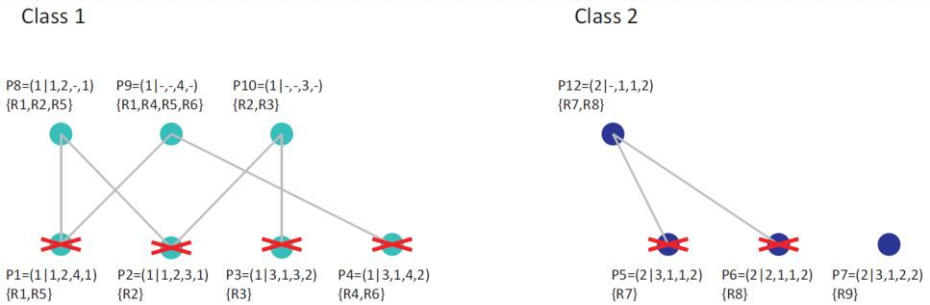


Pruning

1.Deleting contradictory patterns as well as general patterns that have exception patterns in some other class.

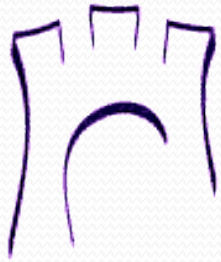


2.Removing more concrete patterns within the classes (ensures compactness of the pattern set)

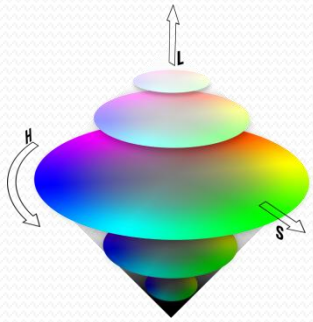


Experimental Art Collection

Movement	Artist
Icons (60)	Icons (60)
Renaissance (90)	Botticelli (30); Michelangelo (30); Raphael (30)
Baroque (90)	Caravaggio (30); Rembrandt (30); Rubens (30)
Romanticism (90)	Friedrich (30); Goya (30); Turner (30)
Impressionism (90)	Monet (30); Pissarro (30); Sisley (30)
Cubism (90)	Braque (30); Gris (30); Leger (30)
Modern Art (90)	Klimt (30); Miro (30); Mucha (30)

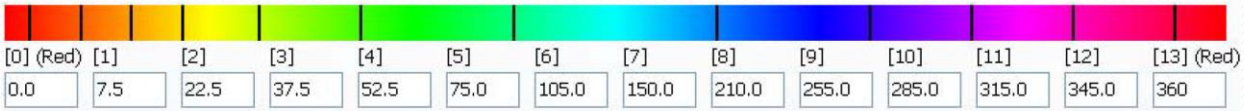


Low-level Visual Descriptors

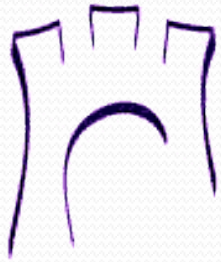


- The pixels in the images are converted into the HSL-artists color model.

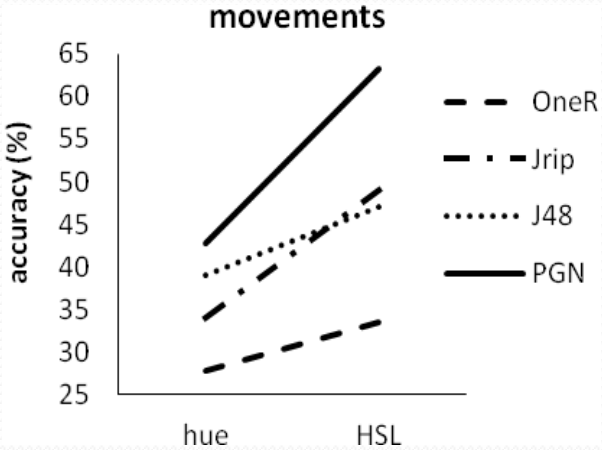
- The quantization of Hue is made to 13 bins
  ("-1": achromatic; "0".."11": "red"..."magenta")
  non-linear with respect to taking into account the misplacement of artists' color wheel and Hue definition in HSL color space
  –> 13 attributes for Hue

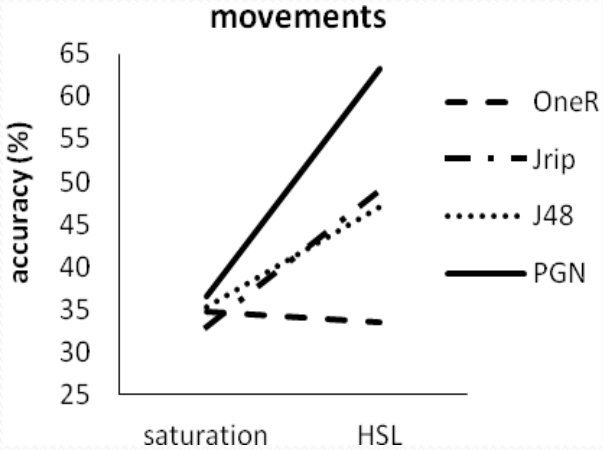


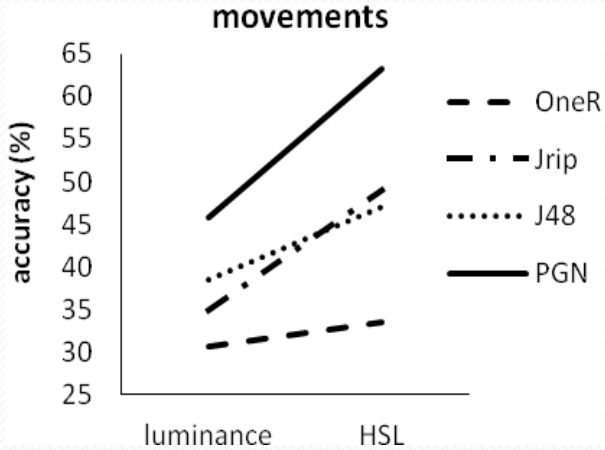
- The saturation and lightness are linearly quantized into 10 bins
  –> 10 attributes for Saturation
  –> 10 attributes for Luminance



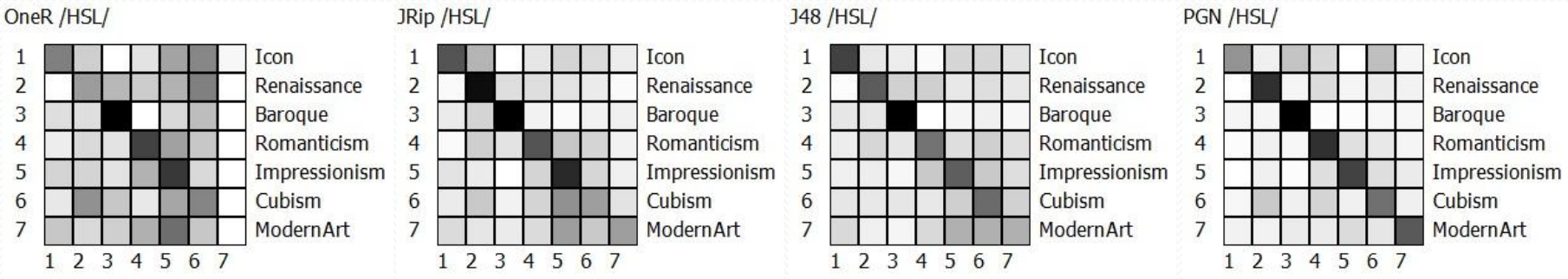
The accuracies of different classifiers by hue, saturation, luminance separately and all three together



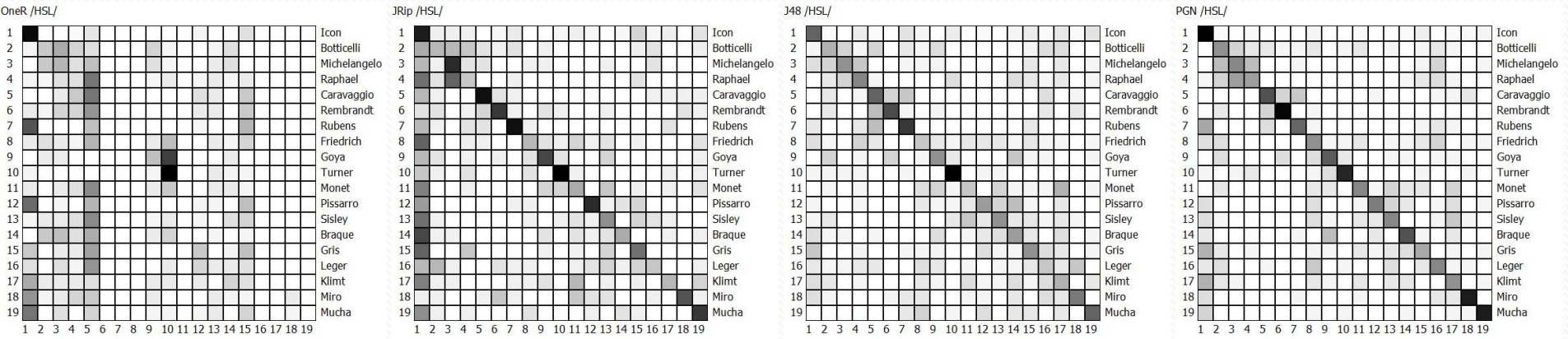




Confusion matrices



The Baroque movement is the easiest to predict; OneR fails to predict Modern Art;
PGN is the only classifier with a smooth consistent black/gray downwards diagonal.

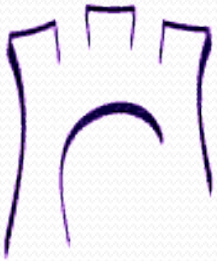


OneR is not able to classify the different artist paintings.
JRip predicts almost 25% of the paintings as Icon

(specific distribution of artists presentations – equal number, except Icons)

Conclusion

- The results of the experiments confirm our expectations that the proposed approach to prioritize confidence over the support has its reason and leads to outperforming PGN against other rule-based classifiers especially in the case of multi-class datasets.
- We believe that this approach can be successfully implemented in the resource discovery as a part of access functions in established digital libraries, repositories and aggregators and this way to increase the possibilities of such storages for ease access.



Thank you for the attention !

Krassimira Ivanova
kivanova@math.bas.bg