CSIT
Yerevan 2009

Metric Categorization Relations Based on Support System Analysis

Krassimira Ivanova[1], Ilia Mitov[1],
Krassimir Markov[1], Peter Stanchev[1],
Koen Vanhoof[2],
Levon Aslanyan[3], Hasmik Sahakyan [3]

1 - Institute of Mathematics and Informatics, Sofia, Bulgaria
2 - Hasselt University, Hasselt, Belgium
3 - Institute for Informatics and Automation Problems, Yerevan, Armenia

Analysis of subject area learning sets and analysis of classification schemes raises a number of nonstandard questions, such as relations between categorization/metadata and logic-combinatorial structuring/clustering of the descriptive part of the input table.

1. INTRODUCTION

2. MATHEMATICAL DESCRIPTION

3. PROGRAM REALIZATION

4. EXAMPLE

5. CONCLUSION

*This work is partially financed by Bulgarian National Science Fund under the project **D 002-308 / 19.12.2008** "Automated Metadata Generating for e-Documents Specifications and Standards".*

INTRODUCTION

Set of objects consists of:

- subset of primary measurable features;
- subset of different classification values, which classify the primary subset from a number of viewpoints.

In standard case models follow the simplest case – considering a set of non intersecting classes with one type of classification.

INTRODUCTION

In case of complementary classifiers some problems could be mentioned:

- (I1) *understanding and evaluating* reasonable requirements for the primary object characterization tools, which are sufficient to validate the characterization given by metadata.

- (I2) *finding* logic-combinatorial meaning in feature area that takes informative burden of *describing classes and intersections* given in the metadata area.

PREPROCESSING STEPS

- *cluster analysis by columns* - identification of groups of features, which give similar descriptions of the objects given in descriptive part,

- *cluster analysis by rows* - the best suitable measure gives clusters highly correlated with classes defined in the metadata area by an individual classifier.

- *consistency structure of classifiers and classes* - coding the subsets of classifiers by vertices of dimensional unit cube we form the consistency Boolean function.

PROGRAM REALIZATION

We propose an extension of classical classification methods with the purpose of ***using of metadata for automatic concept identifying of the founded regularities by the system***.

We have used as a ground a part of already realized classification algorithm in the experimental classification system PaGaNe.

At the first stage the learning set is processed by the standard classification algorithm of PaGaNe.

The next step consists of traversing of all metadata positions and finding for each value of these positions one or several control nodes that correspond to this value.

PROGRAM REALIZATION

For every metadata value (that define some concept) can be found zero, one or more corresponded control nodes.

- *Zero* - The reason that corresponded control nodes not exist usually lays in the fact that chosen primary attributes are not enough to correctly define this concept.
- *One* - we can assume that this is the exact name of this control node. The content of this concept is represented as a conjunction of significant values of attributes, contained in corresponded control node.
- *Several* - It is represented as a disjunction of conjunctions of significant values of attributes, contained in connected control nodes.

EXAMPLE

We have taken a ZOO database from UCI Machine Learning Repository, which describes animals using 17 categorical attributes.

We have expanded ZOO manually with three columns of metadata, containing information for the animal respectively:

- in which "Phylum" it belongs to ("Chordata", "Mollusca", "Arthropoda", etc.);
- is "Predator" or "Pray";
- in which "Class" it belongs to ("Mammalia", "Fish", "Aves", "Insecta", etc.).

As a source for additional information we use Encyclopedia of Life.

EXAMPLE

Human definitions of:
-Phylum "Chordata" – backbone;
-Phylum "Mollusca" – a mantle and nervous system;
-Phylum "Atrhropoda" – exoskeleton, a segmented body, and appendages.

Phylum : Chordata
<def>
backbone: yes
</def>

Phylum : Mollusca
<def>
hair: no
feathers: no
eggs: yes
milk: no
airborne: no
backbone: no
venomous: no
fins: no
tail: yes
domestic: no
</def>

Phylum : Arthropoda
<def>
feathers: no
milk: no
toothed: yes
backbone: no
fins: no
catsize: no
</def>

Class : Mammalia
<def>
feathers: no
milk: yes
backbone: yes
breathes: yes
venomous: no
</def>

Class : Fish
<def>
hair: no
feathers: no
eggs: yes
milk: no
airborne: no
aquatic: yes
toothed: yes
backbone: yes
breathes: no
fins: yes
legs: 0
tail: yes
</def>

Class : Aves
<def>
hair: no
feathers: yes
eggs: yes
milk: no
backbone: yes
breathes: yes
venomous: no
fins: no
legs: 2
tail: yes
</def>

CONCLUSION

In this paper we try to apply some concepts, already known in pattern recognition area (such as support systems, parameterized distances and logic separation) to solve some novel specific problems of categorization such as discovering the ***relations*** *between descriptive part and metadata values* of the input table in order to use the metadata for ***automatic concept identifying*** of the founded regularities.

The concept analysis system PaGaNe is applied to treat the metric-categorization relations.

Categorization modeling raised questions that brought to the logic combinatorial recognition area.

Thank you for the attention!