cse537
Artifficial Intelligence

Professor Anita Wasilewska

PROJECT DESCRIPTION

BAKARY CLASSIFICATION DATA

BAKARY DATA - posted on the course web page

This is a **classification data** with TYPE DE ROCHE (Rock Type) as the CLASS attribute

There are 98 records with 48 attributes and 6 classes

Classes are:

C1: R. Carbonatees AND R. Carbonatees impures

C2: Pyrate

C3: Charcopyrite

C4: Galene

C5: Spahlerite

C6: Sediments terrigenes

Most important attributes (as determined by the expert) are:

S, Zn, Pb, Cu, CaO+MgO, CaO, MgO, Fe2O3

This is a real life experimental data and it contains a lot of missing data

PROJECT

Project GOAL is to use Internet based Classification Tools to build **2 classifiers:** one descriptive and one statistical, **discuss** the results and **compare** these two approaches on the basis of obtained results

1. Descriptive Classifier

Use a **Decision Tree** tool to generate and test sets of **discriminant rules** describing the content of the data, i.e. to **build your own classifier**

You can choose a tool you like, or use **WEKA**, the Waikato Environment for Knowledge Analysis

It can be obtained from

http://www.cs.waikato.ac.nz/˜ ml/weka/index.html)

PROJECT

2. Non-Decsriptive Classifier (statistical)

Use Neural Networks to build your **Classifier**

Choose your own tool

Here are some suggestions:

http://www.mathworks.com/products/neural-network/

http://www.simbrain.net/

PROJECT

The project has to follow all steps of **Learning Process**

1. Data Preparation

It includes attributes selection, cleaning the data, filling the missing values, etc...

Describe your motivation and results

2. Data preprocessing

For the Descriptive Classifier you must use 2 methods of data **discretization** and call the obtained data "My Data1", "My Data 2"

For the Non -descriptive Classifier use your chosen method of preprocessing

PROJECT

3. Learning Proper for the Descriptive Classifier

Use your "My Data1" and "My Data 2" and your **classification tool** for rules generation and testing to perform **Experiments 1- 3** described below

For the Non -descriptive Classifier perform the **Learning Proper** for each the **Experiments 1- 3**

Compare the results

EXPERIMENTS

Experiment 1

Use all records to find rules for the full classification; i.e. rules describing all classes **C1- C6** simultaneously

Experiment 2

Use all records to find rules **contrasting** class **C1** with all other classes

Experiment 3

Repeat **Experiments 1, 2** for all records with the most important attributes (as defined by the expert) only

Write a **Project Description** with methods, motivations, results and submit submit via e-mail to TA and Professor.