*Homework 3*: **Three-Way Decision through RST**

1. 3WD package

There are 4 functions in 3WD package.

+ Identify the concept X is subset of U. getConcept() function has 3 parameters namely dataset, decision attribute and the target.

+ Identify the equivalent relation and indiscernibility relation by getIND() function with 2 parameters dataset and the conditional attributes.

+ Calculate conditional probability by getProbability() function with three parameters dataset, indiscernibility relation and the concept X.

+ Identify POS, NEG, BND regions of 3WD with threshold alpha, beta and the result of conditional probability.

1. Diabetes Dataset

Dataset of diabetes, taken from the hospital Frankfurt, Germany. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

Link: <https://www.kaggle.com/johndasilva/diabetes>

Dataset has 9 columns and 768 rows.

Outcome: 1 means people having diabetes and 0 means people not having diabetes.

1. Design a data analysis method.

There are 9 attributes: Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, Age, Outcome

* Choose conditional attributes: Pregnancies, BloodPressure, Age
* Choose decision attribute: Outcome
* Choose goals: classify the people depending on their pregancies, blood pressure and age to identify whether they have diabetes or not or unidentified.

1. Write a Python program (Main Script), invoking 3WD package

**Step 1**: Run homework3.py file

**Step 2**: Input data filename: diabetes.csv

Output: head of dataset and set of attributes.

**Step 3**: Input decision attribute: Outcome

**Step 4**: Input decision value: 1

Output: set of objects X with the Outcome value equal to 1.

Note: You can Input 1 or 0 for this dataset.

**Step 5**: Input conditional attributes: Pregnancies BloodPressure Age

Output: a list of indiscernibility Relation IND and the conditional probability of Objects in concept

Note: it depends on our purposes we can input which condition attributes we want following the sample above. Name of attribute and space and Name of another attribute and so on. There is no comma between condition attributes.

**Step 6**: Input alpha threshold: 0.8

Note: parameters α can be computed from the loss function

**Step 7:** Input beta threshold: 0.2

Note: Note: parameters β can be computed from the loss function

Output: POS, NEG, BND

Thank to the concept of three-way decisions, decision makers can accept that 238 people in the positive regions have diabetes and reject that 471 the people in the negative regions don’t have diabetes and 59 noncommitment in the boundary regions.