

REDUCTING CLASSIFICATION RULES SYSTEMS APPLIED TO THYROID FUNCTIONAL DIAGNOSIS

Marc Almiñana¹, Laureano F. Escudero¹, Agustín Pérez¹, Alejandro Rabasa¹,
Cristina Sánchez² and Laureano Santamaría¹

¹*Centro de Investigación Operativa, Universidad Miguel Hernández. Spain*

²*Hospital General de Elda, Servicio Valenciano de Salud. Spain*

Thyroid gland abnormal operation diagnosis is usually obtained from radioimmunometric analysis where some hormone levels (TSH, T4, T4u, T3, etc.) are measured. The results of these analysis can be used to generate classification rules with discretized values of this variables as antecedent attributes and the clinical diagnosis as consequent (also called class attribute). So, each classification rule with the form: “if [antecedent/s] then [consequent]” is associated to a support and confidence values. The classification rules systems are really helpful for medical experts to find existing correlations between subsets of these hormone levels and the final diagnosis. But if the number of attributes and their cardinality increases, a correct interpretation of the rules system becomes harder (almost impossible) for the expert. For example, selecting 3 antecedent attributes each one having 4 possible discretized values, and the consequent attribute having 10 possible values, a final group of 640 rules could be generated. In this paper we use a classical Counting Frequencies Method to generate complete rules systems, and we propose a reduction system based on regions of relative significance that provides reduced rules systems (where the rules are labelled and ordered by significance). This method also provides an indicator of correlation level between different subsets of attributes (chosen as antecedent) and the clinical diagnosis. This information let the experts an easier interpretation of the most important rules, and brings them an objective measure to evaluate which subsets of variables are more correlated with the predicted clinical diagnosis. Computational experiment has been performed over standard and public datasets of UCI repository (University of California).