MSc thesis proposal

Main director: David Riaño

Research Group: Research Group on Artificial Intelligence

Co-directors: -

MSc thesis title: Background knowledge to improve induction of decision structures in medicine.

Master: Master in Artificial Intelligence (UPC-URV-UB)

Description:

The aim of this thesis is to investigate which heuristics are involved in the medical decision process and use them to improve the medical coherence of the decision structures used to assist the physicians.

Usually, decision structures are generated automatically without any kind of background knowledge of the domain using heuristics like the information gain whose objective is to obtain simple and efficient structures. It can be enough in some domains when we only need the final classification determined by the structure. However, in other cases, we are not interested only in the final classification but in the whole decision process. In these cases, obtaining structures as simple as possible is not useful because the decisions made along the decision process will probably be incomprehensible (or incorrect) by the experts. The solution is to acquire all the substantial background knowledge of the domain from the experts and use it when making decisions during the generation of the structure.

Our area of interest is medicine. We want decision structures representing the different steps followed by the physician during the process of diagnosis, for example. Our approach uses decision trees because they are a simple way for representing a decision process. In order to apply the medical background knowledge in their generation we use cost-sensitive decision trees. This kind of decision trees tries to minimize some kind of cost function. In our case, this cost function depends on different heuristics. We consider heuristics related to the length of the process (information gain, response time), the structure or the economy (economic cost, resource availability), medical reasons (adherence / coherence) and acceptability (health risk, comfortableness).

Each of the heuristics has a different way of representation. For example, the economic cost of a certain test could be measured numerically (in euros) but to represent its health risk it is probably better to do it categorically (no risk, low risk, etc.). One of the main problems of our thesis is to find the best way to represent each of the heuristics assuming that we want a trade off between information as exhaustive as possible and easiness for the physician to provide it. We investigate several approaches as numerical and categorical representations or partially ordered sets.

Another important point is the combination of heuristics. We want to join their values in a unique cost function so we study how to integrate such different representations.

It is expected that the algorithms developed in this thesis will be tested on real data coming from the Hospital Consortium SAGESSA (Reus) and the Clinical Hospital in Barcelona.