

Assignment

A small health insurance company wants to standardize its underwriting process. In this process an applicant apply for a health insurance. But different workers decide differently (i.e. based on their own knowledge) and therefore the acceptance depends highly which worker gets the application on its desk. The insurance company now decides to collect the knowledge and define some guidelines that will be obligatory in future. In order to support them the insurance company also decided to support the workers by an knowledge-based system which give them some recommendation how to decide in a particular case.

There are two kinds of sources for such knowledge. The first sources are the experts themself. The second sources are the past applications stored in a sheet. The company wants to extract the knowledge out of that data as well.

Now your task is to build the knowledge-based system. Solve that task in two steps:

1. You need to transform the knowledge from experts into a format which can be used by an expert system, i.e. model that knowledge into an appropriate knowledge representation language.
2. You also need to extract the knowledge from the sheet and translate it also into the knowledge representation formalism.

Both tasks can be done in groups by two or three.

Your delivery will consists of the following things:

1. A brief documentation how you translated the knowledge from experts as well as from the sheet into the knowledge representation formalism.
2. A file which contains the represented knowledge in three parts
 - a. first part: knowledge which is extracted from the experts
 - b. second part: knowledge which is extracted from the sheet by some learning algorithm
 - c. and third part, which combines the results from both sources

Please send a zip file (zip ONLY) to your supervisors per e-mail not later than June 15th, 2016, 18:00 MEZ.

Knowledge from the Experts

The experts told in an interview that several applications could be rejected immediately because they are not satisfying some basic guidelines. A person younger than 21 year is not able to apply for a health insurance in your company. Also persons those are older than 70 are not able to apply. Also the place of residence is important. Applicants from Italy or Switzerland are eligible. But applicants that live in other that these both countries are not eligible and will be rejected.

Further some diseases are excluded. Customers with a hearing damage will be rejected immediately. Also customers older than 55 and some mental-health diseases are also rejected.

Further the experts agreed that the estimate some risk values for such an applicant. They agreed that the risk value should be *low*, *medium* or *high* (instead of numbers). Applicants with a high-risk value will be rejected. Customers with medium risks may pay an extra fee to compensate the risk.

Knowledge from the sheet of former applications

The health insurance company archive all former, successful application. The extract of these applications was transformed into an excel sheet. The extracted data from the applications are (i.e. the attributes):

- age in years (attribute "age")
- number of surgeries in the last 5 years (attribute "surgery")
- number of doctor visits in the last 5 years (attribute "docvisit")
- does the applicant have an allergy (attribute "allergy")
- does the applicant takes medications on a daily base (attribute "med")
- on which diseases does the applicant suffer (attribute "disease")
- what is her or his BMI (Body Mass Index) (attribute "bmi")?
-

Further the company checks how much money these customers did cost the company so far. If a customer did cost a lot of money (expensive surgeries, expensive medications etc) then its risk value is set to high. If the customer did not cost less money in the past (only some doctor visits over the years, some medications) the risk value is set to low. All cases in between are set to medium. The risk value is added to the sheet as well (attribute "class").

The aim of the knowledge extraction task here is to analyze which applications in the past results into which risk value. Generalized knowledge is highly appreciated!

