Assignment

A small health insurance company wants to standardize its underwriting process. In this process an applicant can apply for a health insurance and the company decides, whether an application can be accepted. However, different workers decide differently (i.e. based on their own knowledge) and therefore the acceptance depends highly on which worker gets the application on its desk. The insurance company decided to support their workers with a knowledge-based system, which gives them recommendations how to decide in a particular case.

There are two kinds of sources for such knowledge. The first sources are the experts themselves, who have rules about which candidates are eligible. The second sources are the past applications stored in a database. It represents knowledge about the risks of candidates. The company wants to extract the knowledge out of that data as well.

Now your task is to build the knowledge-based system, which only selects candidates, who are eligible according to the expert's knowledge and have low or medium risk. 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 data and translate it also into the knowledge representation formalism.

Knowledge from the data of former applications

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

- 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 much 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!

Knowledge from the Experts

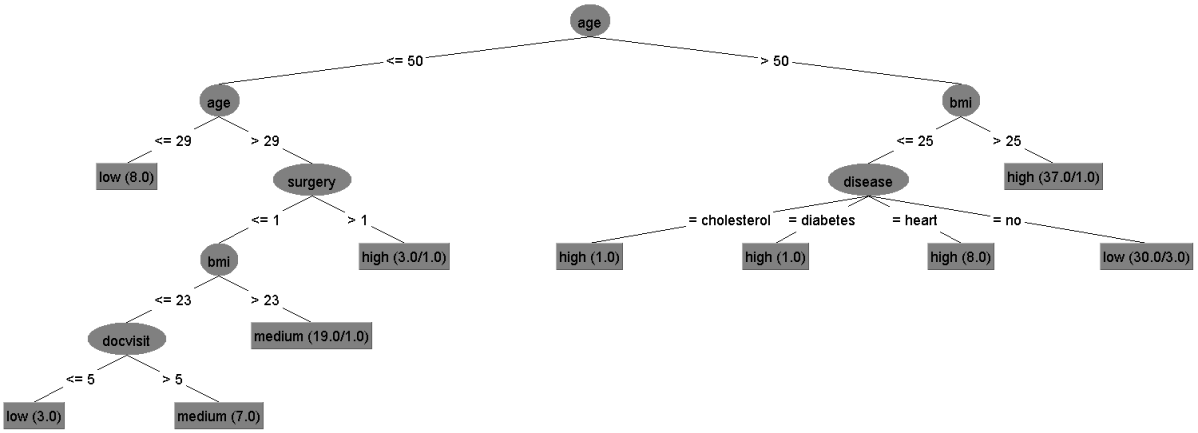
The expert knowledge represents information about the eligibility of candidates. The experts told in an interview that several applications could be rejected immediately because they are not satisfying some basic guidelines. A person younger than 21 year is not able to apply for a health insurance in 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.

Possible Solution:

Decision Tree learnt from Data with ID3 (using J48 in WEKA).



Translation of rules into Prolog:

risk(P,low) :- age(P,A), A=<29.

risk(P,medium) :- age(P,A), A>29, A=<50, surgery(P,S), S=<1, bmi(P,B), B=<23, docvisit(P,V), V>5.

risk(P,low) :- age(P,A), A>29, A=<50, surgery(P,S), S=<1, bmi(P,B), B=<23, docvisit(P,V), V=<5.

risk(P,medium) :- age(P,A), A>29, A=<50, surgery(P,S), S=<1, bmi(P,B), B>23.

risk(P,high) :- age(P,A), A >29, A=<50, surgery(P,S), S>1.

risk(P,high) :- age(P,A), A>50, bmi(P,B), B=<25, disease(P,cholesterol).

risk(P,high) :- age(P,A), A>50, bmi(P,B), B=<25, disease(P,diabetes).

risk(P,high) :- age(P,A), A>50, bmi(P,B), B=<25, disease(P,heart).

risk(P,low) :- age(P,A), A>50, bmi(P,B), B=<25, disease(P,no).

risk(P,high) :- age(P,A), A>50, bmi(P,B), B>25.

Expert rules:

reject(P) :- age(P,A), A <21.

reject(P) :- age(P,A), A >70.

reject(P) :- not(residence(P,switzerland)), not(residence(P,italy)).

reject(P) :- disease(P,hearingdamage).

reject(P) :- age(P,A), A >55, disease(P,mental).

Combining expert rules and rules from expert to make final decision:

accept(P,no) :- reject(P).

accept(P,no) :- risk(P,high).

accept(P,yes) :- risk(P,low), not(reject(P)).

accept(P,extrafee) :- risk(P,medium), not(reject(P)).

Information about applicants

We assume that information about applicants are represented as facts like

age(peter,25).

surgery(peter,0).

docvisit(peter,2).

allergy(peter,no).

med(peter,no).

disease(peter,cholesterol).

bmi(peter,24)