Capita Selecta AI - Entity Relationship Model

Daan Fierens

daan.fierens@cs.kuleuven.be

1 Problem Description

1.1 Entities, attributes and relations

Consider the following entity relationship model about a real-estate agency ("immobiliënmaatschappij") for luxury houses. There are three types of entities: houses, facilities (garden, swimming pool, tennis court, ...) and customers.
Associated to these entity types are some attributes.

- Attributes of houses:
  - neighbourhood (possible values {bad,good})
  - cheap (possible values {no,yes})
- Attributes of facilities:
  - exclusivity (possible values {high,low}), for instance, exclusivity might be low for a garden but high for a tennis court
- Attributes of customers:
  - age (possible values {old,young})
  - rich (possible values {no,yes})

There also exists some relationships between entities:

- wants: indicates that a customer wants a facility,
- has: indicates that a house has a facility,
- interested: indicates that a customer is interested in a house,
- buys: indicates that a customer buys a house

1.2 Probabilistic dependencies between attributes and relationships

Assume that the set of entities (the houses, facilities and customers) is given, but that there is uncertainty about the attributes of these entities and about the relationships between these entities. Furthermore, we have the following information.

1. The age of a customer is old with probability p_1 (hence, the age is young with probability $1 - p_1$).
2. The exclusivity of a facility is high with probability p_2.
3. The neighbourhood of a house is bad with probability p_3.
4. For a given customer and a given facility, the probability that the customer wants the facility is p_4.
5. Whether a customer is rich depends on his age.
   - If age is young, then he is rich with probability p_5.

- If age is old, then he is rich with probability p_6.

6. Whether a given house has a given facility depends on the exclusivity of the facility and on the neighbourhood of the house.
   - If the exclusivity is low and the neighbourhood is good, then the probability that the house has the facility is p_7.
   - If the exclusivity is high and the neighbourhood is bad, then the probability that the house has the facility is p_8.
   - Otherwise, the probability that the house has the facility is p_9.
7. Whether a house is cheap depends on the facilities that the house has.
   - The probability that a house is cheap is p_{10}^n, with n the number of facilities that the house has. (In other words, the probability that a house is cheap decreases exponentially with the number of facilities that it has.)
8. Whether a given customer is interested in a given house depends on whether there exists at least one facility that the customer wants and that the house has.
   - If there is at least one such facility, then the probability that the customer is interested in the house is p_{11}.
   - Otherwise the probability that the customer is interested in the house is p_{12}.
9. Whether a given customer buys a given house depends on whether the customer is interested in the house, whether the customer is rich and whether the house is cheap.
   - If the customer is not interested in the house, the probability that he buys the house is p_{13}.
   - If the customer is interested in the house and he is rich, the probability that he buys the house is p_{14}.
   - If the customer is interested in the house and he is not rich and the house is cheap, the probability that he buys the house is p_{15}.
   - If the customer is interested in the house and he is not rich and the house is not cheap, the probability that he buys the house is p_{16}.
10. Each house is bought by at most one customer. (In other words, two different customers cannot buy the same house.)

2 Assignment 1: Modelling

Try to model the above problem using your PLL language.

- To create a model in your PLL system you will probably have to use concrete values for the probabilities p_1 to p_{16}. You can choose these values yourself.
- Try to model as many of the above dependencies as possible. If you decide not to model a certain dependency:
  - Assume instead that the relevant property is independent with some prior distribution (for instance, if you do not model that whether a house is cheap depends on the facilities that the house has, then assume that a house is cheap with a certain probability p).
  - Motivate why you do not model that dependency. It might be impossible to model the dependency in your PLL language. If you think that this is indeed the case, then explain why you think this (for instance, which restriction of the language makes it impossible?).