

Assignment 3

Alfredo Clemente Ragone (alfredvc@stud.ntnu.no)
Per Magnus Veierland (permve@stud.ntnu.no)

April 19, 2016

1 INTRODUCTION

For this assignment we have chosen to create a decision support system for an every-day life decision; selecting which store to visit when shopping groceries. In reality, correctly modelling this choice requires several factors such as the available discounts on the desired goods, whether any of the desired items are sold out, whether there is parking space available at the store, how bad the traffic is on the way to the store, how crowded the store is likely to be etc. Given that many of the relevant factors are rapidly changing and require up-to-date information, and that several of the factors are stochastic, it is difficult for a user to make a highly rational choice. Decision support systems are beneficial because while it is possible to specify the available decision options and factors, tracking up-to-date information and combining the available evidence into evaluating the best decision is difficult.

Our model, see Figure 1, takes five decision inputs, with four of the inputs set by the user when evaluating the model: 1) *Item*; the desired item to purchase (*Hamburger*, *Eggs & Bacon*, *Coca-Cola*, or *Taco*), 2) *Hunger*; the degree of hunger (*None*, *Some*, or *Very*), 3) *Weekend*; a Boolean value indicating whether it is a weekday or weekend, 4) *Time of Day*; the point of time (*12-15*, *15-18*, *18-21*, or *21-24*).

The fifth model decision input is the *Store*; with the explicit decision alternatives *Bunnpris Elgeseter*, *Sesam Samfundet*, *Rema 1000 Elgeseter*, and *Shell Elgeseter*. When evaluating the model, a single utility value is output for a combination of inputs. The model is evaluated once for each possible store, given the other four inputs; and the store which outputs the greatest utility is the action recommended by the model, following the principle of *maximum expected utility*.

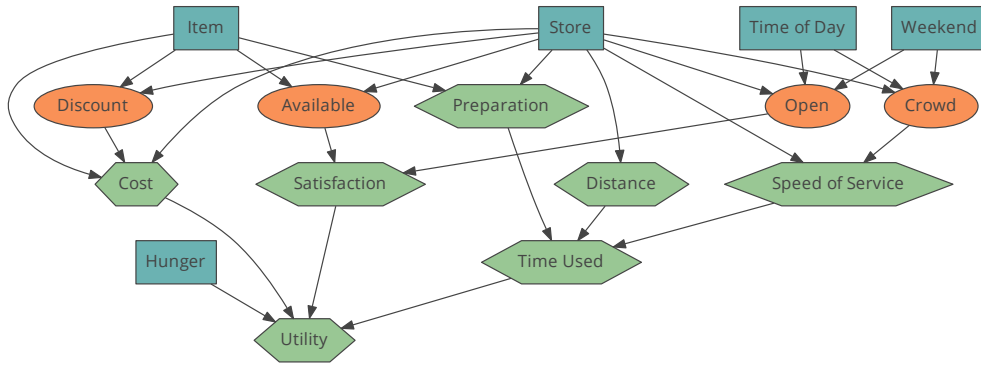


Figure 1: Model visualization showing *decision nodes* (blue), *chance nodes* (orange), and *utility nodes* (green).

Table 1 shows an overview over the internal chance and utility nodes, described further in Section 2.

2 MODEL

2.1 ASSUMPTIONS

The following assumptions have been made in the modeling representation to reduce complexity:

1. The user only purchases one item at a time when visiting a store. An item in this context can be a set of related goods, e.g. “Eggs & Bacon”.
2. Discounts are modelled as being binary, with an item being 50% cheaper if it is discounted. The probability of an item being discounted is a function of the store and the item, and an item can either be discounted or not.
3. To maintain some temporal resolution while keeping the model simple, only the hours 12:00 to 24:00 is modelled.
4. Item discount and item availability are both independent of whether it is a weekday or weekend, and the time of day.
5. Whether an item is available only depends on the item and on the store, it does not depend on the time or how crowded the store is.
6. The value of time and money are linear within the ranges used.

The following conditional independencies are assumed:

Type	Known	Name	Values
Chance	No	Discount	Probability
Chance	No	Available	Probability
Chance	No	Open	Probability
Chance	No	Crowd	<i>None, Some, Very</i>
Utility	Yes	Preparation	Duration in seconds
Utility	No	Cost	Cost in NOK
Utility	No	Satisfaction	Probability
Utility	Yes	Distance	Travel duration in seconds
Utility	No	Speed of service	Time in seconds
Utility	No	Time used	Time in seconds

Table 1: List of internal model variables. “Known” refers to whether the information is fully known at the time when the decision is made.

	<i>Bunnpris Elgeseter</i>	<i>Sesam Samfundet</i>	<i>Rema 1000 Elgeseter</i>	<i>Shell Elgeseter</i>
<i>Hamburger</i>	0.1	0.0	0.1	0.15
<i>Eggs & Bacon</i>	0.1	0.0	0.1	0.0
<i>Coca-Cola</i>	0.05	0.0	0.05	0.1
<i>Taco</i>	0.2	0.0	0.2	0.0

Table 2: Probability of an item being discounted

1. *Speed of Service* is assumed to be conditionally independent of *Weekend* and *Time of Day*, given the *Crowdedness* of the store.
2. *Open* and *Crowd* are conditionally independent of each other given *Weekend*, *Time of Day* and *Store*.
3. *Open* and *Available* are conditionally independent given *Store*.

2.2 DISCOUNT

As discussed in the assumptions section, store discounts are modelled as a single probability per store-item pair, indicating the chance of an item being discounted. If an item is discounted, the price of the item will be 50% of its nominal price.

The probability values listed in Table 2 are based on anecdotal evidence, both because our discount model is highly simplified, and due to the work required in tracking discounts at stores over long enough time to build correct averages.

2.3 AVAILABLE

Availability, see Table 3, is defined as the probability of all components of an item being available for purchase at a given store at any time.

	<i>Bunnpris Elgeseter</i>	<i>Sesam Samfundet</i>	<i>Rema 1000 Elgeseter</i>	<i>Shell Elgeseter</i>
<i>Hamburger</i>	0.85	0.95	0.90	0.9
<i>Eggs & Bacon</i>	0.95	0.0	0.95	0.0
<i>Coca-Cola</i>	0.95	0.8	0.97	0.95
<i>Taco</i>	0.80	0.0	0.85	0.0

Table 3: Probability of an item being available

2.4 OPEN

In order to calculate the probability of a store being open or closed we took into account the usual store opening times together with an assigned uncertainty for whether the store would actually be open when it was scheduled to be open. Using Equations 1-3, a probability for the store being opened, given whether it is a weekday or weekend and the time period of the day, is calculated for each store. The first two formulas are weekday specific, but are identical for the weekend.

$$P(\text{Open}|\text{Store}, \text{Weekday}, \text{TimePeriod}) = P(\text{Open}|\text{ScheduledOpen}) \cdot P(\text{ScheduledOpen}|\text{Store}, \text{Weekday}, \text{TimePeriod}) \quad (1)$$

$$P(\text{ScheduledOpen}|\text{Store}, \text{Weekday}, \text{TimePeriod}) = \frac{1}{n_{\text{weekdays}}} \cdot \sum_{d \in \text{Weekday}} P(\text{ScheduledOpen}|\text{Store}, d, \text{TimePeriod}) \quad (2)$$

$$P(\text{ScheduledOpen}|\text{Store}, \text{Day}, \text{TimePeriod}) = \frac{\text{HoursOpen}(\text{Store}, \text{TimePeriod})}{\text{Hours}(\text{TimePeriod})} \quad (3)$$

2.5 CROWD

When visiting a store, how crowded the store is will affect the time it takes to shop there. “Crowdedness” is modelled as a function of both the store and point of time, and can either be “none”, “some”, or “very”. Each of the three outcomes is assigned a probability between 0 and 1 for each point of time. The probabilities set in Table 4 are based on anecdotal evidence, and based on the observations that grocery stores are more busy during daytime; especially lunch and after work, and that all stores are more busy during weekends.

2.6 PREPARATION TIME

For this task we assume that the preparation time for each meal is only dependent on the meal, and on where it was bought from. We assume also a that preparation time varies very little from time to time, and this variation has little to no effect on the decision we are modeling, and therefore choose to model it as a constant. Table 5 shows common preparation times found on the internet. When the food is prepared at the store, then preparation time represents how much time the store takes to prepare the food.

		<i>Bunnpris Elgeseter</i>	<i>Sesam Samfundet</i>	<i>Rema 1000 Elgeseter</i>	<i>Shell Elgeseter</i>
Weekday	12-15	0.1, 0.6, 0.3	1.0, 0.0, 0.0	0.1, 0.6, 0.3	0.5, 0.3, 0.2
	15-18	0.1, 0.5, 0.4	0.8, 0.2, 0.0	0.1, 0.5, 0.4	0.5, 0.3, 0.2
	18-21	0.4, 0.4, 0.2	0.7, 0.2, 0.1	0.4, 0.4, 0.2	0.7, 0.2, 0.1
	21-24	0.8, 0.2, 0.0	0.1, 0.7, 0.2	0.8, 0.2, 0.0	0.8, 0.1, 0.1
Weekend	12-15	0.1, 0.5, 0.4	1.0, 0.0, 0.0	0.1, 0.5, 0.4	0.5, 0.3, 0.2
	15-18	0.1, 0.5, 0.4	0.7, 0.2, 0.1	0.1, 0.5, 0.4	0.5, 0.3, 0.2
	18-21	0.3, 0.5, 0.2	0.6, 0.3, 0.1	0.3, 0.5, 0.2	0.5, 0.3, 0.2
	21-24	0.5, 0.4, 0.1	0.1, 0.6, 0.3	0.5, 0.4, 0.1	0.6, 0.2, 0.2

Table 4: Probability of a store being crowded (*None, Some, Very*)

	<i>Bunnpris Elgeseter</i>	<i>Sesam Samfundet</i>	<i>Rema 1000 Elgeseter</i>	<i>Shell Elgeseter</i>
<i>Hamburger</i>	1800 [2]	300	1800 [2]	480
<i>Eggs & Bacon</i>	900 [3]	0	900 [3]	0
<i>Coca-Cola</i>	0	0	0	0
<i>Taco</i>	2400	0	2400	0

Table 5: Preparation time in seconds.

2.7 COST

The cost of an item is modeled as static, meaning that the cost of items do not change in time and the only factor that affects the final price of an item is discounts.

The costs stated in Table 2 are based on information obtained on the internet, and from the stores themselves.

2.8 SATISFACTION

We define satisfaction, see Equation 4, as the probability of obtaining the desired item. This is determined by the probability of the item being available, and the probability of the store being open. Given that they are conditionally independent of each other given the store, the function used is:

$$Satisfaction(Open, Available) = Open \cdot Available \quad (4)$$

	<i>Bunnpris Elgeseter</i>	<i>Sesam Samfundet</i>	<i>Rema 1000 Elgeseter</i>	<i>Shell Elgeseter</i>
<i>Hamburger</i>	60	95	50	115
<i>Eggs & Bacon</i>	45	0.0	40	0.0
<i>Coca-Cola</i>	20	25	20	28
<i>Taco</i>	50	0.0	45	0.0

Table 6: Non-discounted cost of items given the store.

	<i>Bunnpris Elgeseter</i>	<i>Sesam Samfundet</i>	<i>Rema 1000 Elgeseter</i>	<i>Shell Elgeseter</i>
<i>None</i>	300	120	360	120
<i>Some</i>	600	300	720	300
<i>Very</i>	900	1200	1080	900

Table 7: Speed of service in seconds given the store and level of crowdedness.

2.9 DISTANCE

The distance for each store was found through measurements using the *Google Earth* tool. Measurements were made with a starting point at the intersection between the two authors; at latitude 63.418095° and longitude 10.396682°. Each of the distances measured were then doubled to account for the travel in both directions.

The mean walking speed for young males is 1.5 $\frac{\text{meters}}{\text{second}}$ [1]. Using this as a conversion factor, each distance can be converted to a duration which can be unified with other metrics into a utility value.

The distances (both travel and return), (and corresponding durations), were found to be 1. *Bunnpris Elgeseter* (400 meter; 267 seconds), 2. *Sesam Samfundet* (1000 meter; 667 seconds), 3. *Rema 1000 Elgeseter* (760 meter; 507 seconds), 4. *Shell Elgeseter* (500 meter; 334 seconds).

2.10 SPEED OF SERVICE

The speed of service, see Table 7, is defined as the amount of time used from entering a store to obtaining the item to be purchased. It is considered to be independent of the item being purchased, and depends only on the store and on how crowded it is.

2.11 UTILITY

The utility function we use for this model, see Equation 5 is a function of time used, satisfaction, cost and hunger. These are the main properties that a person takes into account when choosing a store to buy from. Satisfaction affects the utility independently of the other factors, and it's importance is dependant on the hunger level. The effect of time depends on hunger, if hunger is large time must be less to maintain the same utility. The affect of cost also depends on hunger, if hunger is low then cost must be low to maintain the same utility. And finally if hunger is kept constant and time increases, then the cost must decrease to maintain the same utility.

$$Satisfaction^{\frac{1}{0.8+Hunger}} \cdot \left(\frac{1}{((1-Hunger) \cdot Cost) + (Hunger \cdot \frac{TimeUsed \cdot 150}{3600})} \right) \quad (5)$$

To tune the relation between time and money we used the *Almanac Game*, where we each wrote down the 25th, 50th, and 75th percentile for how much money, in *Norwegian Kroner*,

Type of food	Day	Time of day	Hunger	Expected Action	Model Recommendation
<i>Hamburger</i>	<i>Weekend</i>	<i>21-24</i>	<i>Very</i>	<i>Sesam</i>	<i>Sesam</i>
<i>Hamburger</i>	<i>Weekday</i>	<i>15-18</i>	<i>Some</i>	<i>Rema1000</i>	<i>Sesam</i>
<i>Taco</i>	<i>Weekend</i>	<i>12-15</i>	<i>Very</i>	<i>Bunnpris</i>	<i>Bunnpris</i>
<i>Eggs & bacon</i>	<i>Weekend</i>	<i>12-15</i>	<i>Some</i>	<i>Bunnpris</i>	<i>Bunnpris</i>
<i>Coca-Cola</i>	<i>Weekend</i>	<i>21-24</i>	<i>Some</i>	<i>Shell</i>	<i>Shell</i>

Table 8: Gold standard of input-output pairs

we each believe one hour of time is worth. The results were as follows, Per Magnus had NOK 50, NOK 150 and NOK 400, and Alfredo had NOK 100, NOK 150, and NOK 200. We decided to use the two medians, NOK 150, for the final value. The *TimeUsed* value is then scaled to a monetary value through multiplying with the ratio $\frac{150}{3600} \frac{\text{NOK}}{\text{second}}$.

Finally the effects of hunger are as follows: when hunger is *None* an hour is worth NOK 37.5, when hunger is *Some* an hour is worth NOK 150, and when hunger is *Very* an hour is calculated at NOK 600.

3 EVALUATION

The model was verified and evaluated by creating the gold standard shown in Table 8. The table was created by the authors after the inputs and outputs had been the define, and before the model was developed further. It was used to verify that the model gave results similar to the ones expected. Once the model was finished, the examples from the gold standard were evaluated using the model, giving the outputs shown in Table 8.

With minimal tuning the model managed to output decisions similar to the ones expected on Table 8. The single observed difference between the expected and the obtained results was the model's suggestion of going to *Sesam Samfundet* given the parameters of Example #2. After discussing this scenario amongst the authors, we came to the conclusion that Per Magnus would happily follow this suggestion, however Alfredo would most likely not.

Given that the model gave good results for the examples it was tested on, it should also give reasonable suggestions for other input combinations, aiding in the arduous task of human decision making.

REFERENCES

- [1] Nick Carey. Establishing pedestrian walking speeds. *Portland State University*, 2005.
- [2] Matprat. Hamburger, 2016.
- [3] Meny. Klassisk norsk taco fersk og hjemmelaget, 2016.