Hotel booking prediction by means of fuzzy logic prediction

Michael Hoekstra

BMI Paper

Hotel booking prediction by means of fuzzy logic prediction

BMI Paper

VU University Amsterdam

Faculty of Sciences

Master Business Mathematics and Informatics

De Boelelaan 1081a

1081 HV Amsterdam

September 2010

Preface

This paper is part of the curriculum of the master program *Business Mathematics and Informatics* at the VU University in Amsterdam. The paper should cover at least two of the three components of Business, Mathematics and Informatics.

I would like to thank my supervisor Alwin Haensel for his support during the whole process of research and writing this paper.

Contents

Chapter 1

Introduction

This paper is divided in two parts. The first part describes the theoretical aspects of fuzzy logic and the theory behind fuzzy logic prediction systems. At the end of the first part an example of a fuzzy logic prediction sytem is given, including the modelling steps. The chapters 2 and 3 together are the first part, where chapter 2 describes the fuzzy logic theory in general and chapter 3 the theory of fuzzy logic prediction systems. The second part, which is chapter 4, is an application of the techniques described in the first part. The second part describes the development of a fuzzy logic prediction system for online hotel bookings. Finally, chapter 5 comes up with the conclusions and recommendations with respect to fuzzy logic prediction and the developed system for hotel bookings.

The main goal of this paper is:

Provide information about the fuzzy logic concept and fuzzy logic prediction in general. And develop a fuzzy logic prediction system for an online hotel booking company.

Companies always benefit from accurate predictions. These predictions will help to improve their customer service level and maximalize profit. To predict the hotel bookings fuzzy logic prediction is used, a so-called expert system. This expert system makes use of expert rules, which contain fuzzy statements. An example of an expert rule is IF "price is low" AND "number of visitors is high" THEN "number of bookings is high". A statement as "price is low" is a fuzzy statement because it is not clear when a price is low or not. Fuzzy logic models this fuzziness mathematically.

Besides predicting, fuzzy logic prediction systems are suitable to determine bottlenecks within a system. The hierarchical structure of a fuzzy logic prediction system makes is possible to determine in what manner, and what variables, influence the prediction. This is explained in chapter 4.

Chapter 2

Fuzzy logic[1]

Natural language consists of many statements which are not clear, vague, or fuzzy. "Today is a warm day" and "John is tall" are fuzzy statements because subjectivity plays a role in this. Someone might classify John as being tall in case John is over 175 cm, while someone else considers John tall only when he is over 190 cm. Similar things occur when classifying a day as a warm day. Different classifications of the same instance makes the classification process fuzzy; boundaries of fuzzy data cannot be sharply defined. Statements as "It is over 20 °C" and "John is over 190 cm" are easy to verify and does not lead to discussion of the result. Having a set of length of people and constructing a subset of people who are over 190 cm is also easy. Fuzziness comes around when one would create a subset of tall people out of the set of peoples length; deciding whether a person is tall or not might be a difficult issue. Defining people over 190 cm as being tall gives a clear subset of tall people. It might be interesting to take all peoples length into consideration and assign a membership degree to each length, with respect to the subset of tall people. Such a set is a fuzzy set and Lotfi A. Zadeh presented the modelling of fuzzy concepts in 1965.

2.1 Membership functions

According to Zadeh natural language is a matter of degree [8]. Therefore it is not possible to assert a true or false to statements such as "Today is a warm day". If the statement is "It is over 20 °C" and we denote the temperature in degree Celsius with x, a subset A can be formalized with $\{x \mid x \in \langle -273.15, \infty \rangle, x > 20\}$ (where -273.15 is the lowest temperature possible, e.g. 0 K). Warm cannot be defined as an ordinary subset of temperature. Clearly 33 °C and 35 °C are both warm, but 35 °C is warmer than 33 °C. Zadeh came up with the idea that membership of a fuzzy set should not be on a 0 or 1 basis, but rather on a 0 to

[1]This chapter is partly based on the first three chapters of [5]

1 scale, that is, the membership should be an element of the interval $[0,1]$. An ordinary, or crisp, subset A of a set U is determined by its indicator function $\mathbf{1}_A$ is defined by

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Whereas a fuzzy subset of a set U is a function

$$U \to [0,1]$$

Because an indicator function can take only value 0 or value 1 and the domain boundaries of a fuzzy subset are also 0 and 1, it is obvious that ordinary subsets are a special case of fuzzy subsets. It is common to refer to a fuzzy subset simply as a fuzzy set. Let A be the subset of a set U, than the membership function μ_A assigns a degree of membership to each member of U. For example: let U be the set of maximum temperatures for each day on a certain period and A the subset of "Warm days", than the membership function μ_A can be defined as

$$\mu_A(x) = \begin{cases} 0 & \text{if } x < 10 \\ \frac{x-10}{15} & \text{if } 10 \leq x < 25 \\ 1 & \text{if } 25 \leq x \end{cases}$$

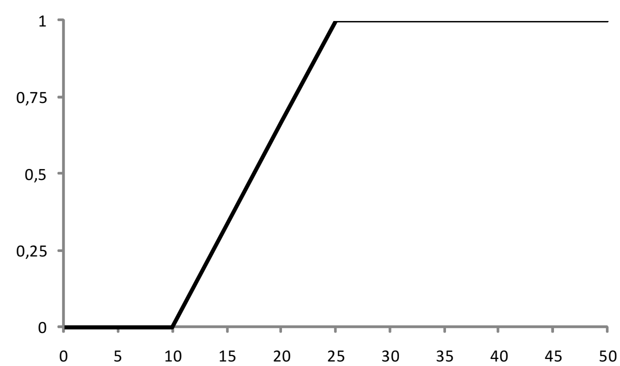

Figure 2.1: Membership function for "warm day in The Netherlands"

For people in The Netherlands this would be an acceptable membership function. People living in a warmer climate, say Mexico, would define a different membership function. This membership function could be

$$\mu_A(x) = \begin{cases} 0 & \text{if } x < 20 \\ \frac{x-20}{15} & \text{if } 20 \leq x < 35 \\ 1 & \text{if } 35 \leq x \end{cases}$$

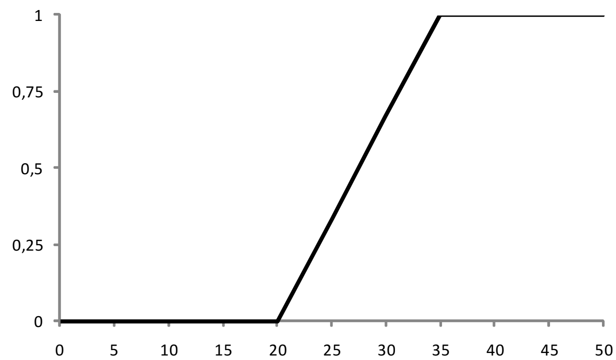

Figure 2.2: Membership function for "warm day in Mexico"

Figure 2.1 and figure 2.2 are examples of the most common membership functions: piecewise linear functions. Although piecewise linear functions are used most of the time, many different functions can be used. Another membership function for a warm day in The Netherlands could be defined as follows, see figure 2.3.

$$\mu_A(x) = \begin{cases} 0 & \text{if } x < 10 \\ \frac{\sqrt{x-10}}{\sqrt{15}} & \text{if } 10 \leq x < 25 \\ 1 & \text{if } 25 \leq x \end{cases}$$

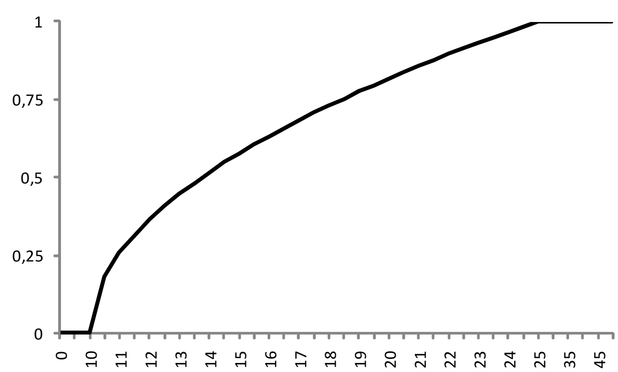

Figure 2.3: Another membership function for "warm day in The Netherlands"

To define a membership function, one must have knowledge about the topic the membership function concerns. In case of the example "warm day" a meteorologist would be the right person to define the membership function. Such a person is called an expert. Although experts are experienced at their field, they can have different opinions about how a membership function should be defined. Thus, the way membership functions are defined depends on the persons personal opinion.

The degree of membership of an element of a set should be viewed as a measure how well this element belongs to a certain subset. Although this degree of membership is within the interval $[0, 1]$, it is not the same as probability.

2.2 Alpha-cuts and fuzzy quantities

2.2.1 Alpha-cuts

Let A be a fuzzy subset of U, and let $\alpha \in [0, 1]$. The α-cut of A is simply the set of those $u \in U$ such that membership function $\mu_A \geq \alpha$. All that is needed to make this definition is that the image of the mapping A be in a partially ordered set. A partially ordered set A is a set with a binary relation " $\leq$ " which satisfies the following three properties:

1. Reflexivity: $a \leq a$ for all $a \in A$.
2. Antisymmetry: if $a \leq b$ and $b \leq a$ for any $a, b \in A$, then $a = b$.
3. Transitivity: if $a \leq b$ and $b \leq c$ for any $a, b, c \in A$, then $a \leq c$.

2.2.2 Fuzzy quantities

Denote $\mathcal{F}(U)$ as the set of all fuzzy subsets of U and $\mathbb{R}$ as the set of real numbers. The elements of $\mathcal{F}(\mathbb{R})$, that is, the fuzzy subsets of $\mathbb{R}$, are fuzzy quantities. It is possible to perform the arithmetic operations of addition, substraction, multiplication and division on fuzzy quantities. Performing arithmetic operations on $\mathbb{R}$ is the same as performing the corresponding operations on $\mathbb{R}$ viewed as a subset of $\mathcal{F}(\mathbb{R})$. For the operations addition, substraction and multiplication is the same in $\mathbb{R}$ as in $\mathcal{F}(\mathbb{R})$; for these operations fuzzy quantities can be treated as real numbers. Compared to division in $\mathbb{R}$, division in $\mathcal{F}(\mathbb{R})$ is different, especially when dividing by 0. In $\mathbb{R}$ a real number can not be divided by 0, in $\mathcal{F}(\mathbb{R})$ it is possible. Recall that $\mathbb{R}$ is viewed inside $\mathcal{F}(\mathbb{R})$ as the indicator function $\mathbf{1}_{\{r\}}$ for elements r of $\mathbb{R}$. The following proposition holds

Proposition 1.1 *For any fuzzy set A, $A/\mathbf{1}_{\{\mathbf{0}\}}$ is the constant function whose value is $\mu_A(0)$*

Thus $\mathbf{1}_{\{r\}}/\mathbf{1}_{\{0\}}$ is the constant function 0 if $r \neq 0$ and 1 if $r = 0$.

2.2.2.1 Fuzzy numbers

Fuzzy numbers are special classes of fuzzy quantities. A fuzzy number is a fuzzy quantity A that represents a generalization of a real number r. Intuitively, $\mu_A(x)$ should be a measure of how well $\mu_A(x)$ approximates r, and certainly one reasonable requirement is that $\mu_A(r) = 1$ and that this holds only for r. Figure 2.4 shows an example of the fuzzy number 10

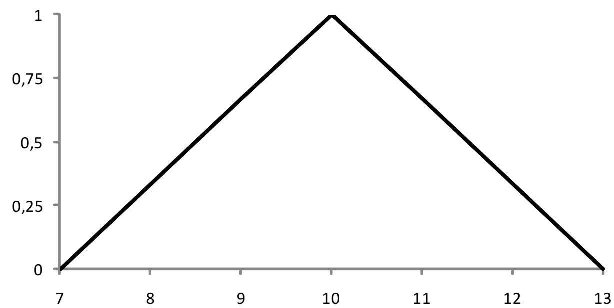

Figure 2.4: A fuzzy number 10

In this case the membership function of the fuzzy number 10 has a triangular shape. The shape does not have to be triangular, various kinds of shapes are allowed. The definition of a fuzzy number captures which shapes are allowed for a fuzzy number. A fuzzy number is a fuzzy quantity A that satisfies the following conditions:

1. $\mu_A(x) = 1$ for exactly one x.
2. The support $\{x \mid \mu_A(x) > 0\}$ of A is bounded.
3. The α-cuts of A are closed intervals.

2.2.2.2 Fuzzy intervals

Ordinary intervals are sharply defined. The boundaries of an ordinary inteval are clear and therefore it is easy to determine whether a real number is in an interval or not. In contrast to ordinary intervals, uncertainty plays a role in fuzzy intervals. Fuzzy quantities on a fuzzy interval $[a, b]$ have value 1 on $[a, b]$ and the other definining properties of fuzzy intervals should be like those of fuzzy numbers. In a formal definition a fuzzy interval is a fuzzy quantity A satisfying the following:

1. A has value 1.
2. The support $\{x \mid \mu_A(x) > 0\}$ of A is bounded.

3. The α-cuts of A are closed intervals

Figure 2.5 shows an example of the membership function for the fuzzy interval $[2, 4]$

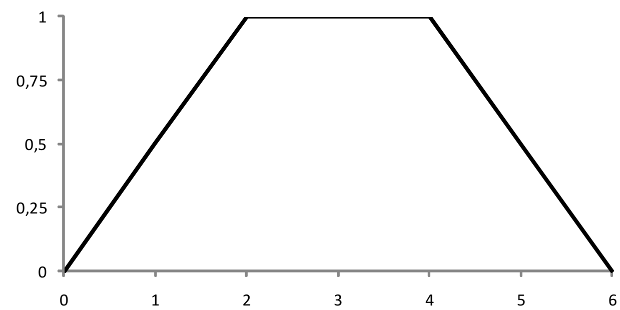

Figure 2.5: A membership function for the fuzzy interval $[2, 4]$

The definitions of fuzzy numbers and fuzzy intervals are very similar. Where the shape of a fuzzy number is not restricted to a triangular one, the shape of a fuzzy interval does not have to be trapezoidal. It can be any shape as long as it satisfies the definition.

Chapter 3

Fuzzy prediction

Several methods exist to predict an output variable with different input variables. A common used technique is regression analysis. Any regression analysis requires a set of assumptions such as linearity, normality and homoscadacity [3]. Furthermore, regression techniques are not capable of digesting linguistic fuzzy data. The fuzzy set theory allows the user to include unavoidable imprecision in the data records. Fuzzy inference is the actual process of mapping with a given set of input variables and output through a set of fuzzy rules. The essence of the modelling is to set up relevant fuzzy rules. General structure of a fuzzy system is shown in figure 3.1. Şen and Altunkaynak [4] set up the following steps which are necessary for succesful application of modelling through a general fuzzy system:

1. Fuzzification of the input and output variables by considering appropriate linguistic subsets such as high, medium, low, heavy, light, hot, warm, big, small.

2. Construction of rules based on expert knowledge and/or the basis of available liteature. The rules relate the combined linguistic subsets of input variables to the convenient linguistic output subsets of input variables to the convenient linguistic output subset. Any fuzzy rule includes statements of **"IF...THEN..."** with two parts. The first part that starts with **IF** and ends before the **THEN** is referred to as the predicate (premise, antecedent) which combines in a harmonious manner the subsets of input variables. Consequent part comes after **"THEN"** which includes the convenient fuzzy subset of the output based on the premise part. This implies that there is a set of rules which is valid for a specific portion of the inputs variation domain. The input subsets within the premise part are combined most often with the logical "and" conjunction whereas the rules are combined with logical "or".

3. The implication part of a fuzzy system is defined as the sharping of the consequent part based on the premise (antecedent) part and the inputs are fuzzy subsets.

4. The result appears as a fuzzy subset and therefore, it is necessary to defuzzify the output for obtaining a crisp value that would be required by the administrators or engineeers.

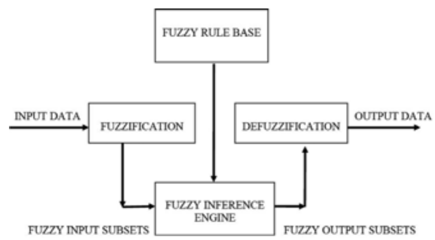

Figure 3.1: General structure of a fuzzy system

3.1 Simulating the behavior of experts

In many models the relation between variables is described by means of rules formulated by experts. For example the expert rule

IF length mother = "tall" AND length mother = "tall" THEN length child = "tall"

describes the relationship in length between father, mother and child. In order to make the model more precise the linguistic terms will be modelled as fuzzy sets. The membership functions belonging to these fuzzy sets must be specified as carefully as possible, because they will decisively inluence the valuation process.

The design of the membership functions can be done by an expert. The knowledge and experience of the expert is important in this process; good knowledge and much experience contributes to an accurate membership function. An pilot study at the Institute of Statistics and Mathematics of the University of Frankfurt am Main described procedures that are able to simulate the evaluation behavior of experts quite accurately, see [2]. These procedures are based on quantiles of the data of the fuzzy system. Where an expert determines the location of the 0-cut and 1-cut for a membership function, in the procedures of the pilot study these locations are obtained by the quantiles of the data. The pilot study developed three methods to simulate the behavior of experts.

Suppose we want to subdivide a variable into three categories (e.g. "low", "normal" and "high"). Each category has its own membership function. The shape of these membership functions is defined with help of quantiles of the data. The first method of the pilot study uses the 0.25-, 0.375-, 0.625- and 0.75-quantiles to locate the 0-cut and 1-cut for a membership function, see figure 3.2.

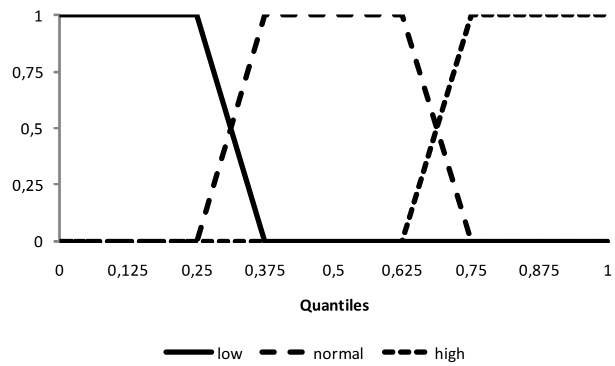

Figure 3.2: First method to simulate behavior of expert

The second method uses the 0.375-, 0.4375-, 0.625- and 0.75-quantiles to locate the 0-cut and 1-cut for a membership function, see figure 3.3.

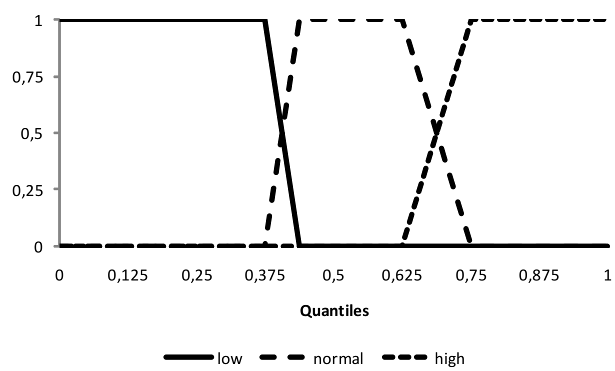

Figure 3.3: Second method to simulate behavior of expert

The third method uses the 0.375-, 0.4375-, 0.5625- and 0.625-quantiles to locate the 0-cut and 1-cut for a membership function, see figure 3.4.

3.2 Example: ice cream demand prediction

To describe how the fuzzy prediction process works we make use of a simple example. We want to predict the ice cream demand for an country on basis of

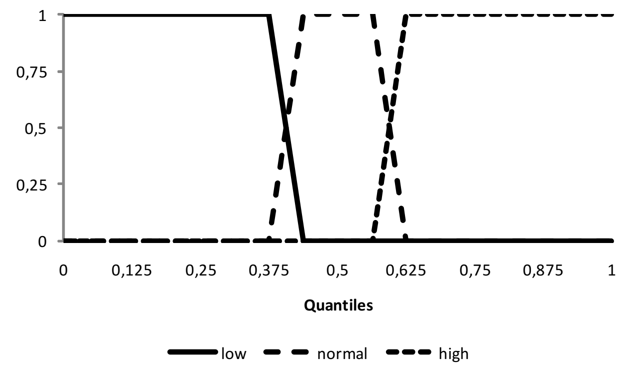

Figure 3.4: Third method to simulate behavior of expert

the percentage of people having vacation, with respect to the number of citizens in the country, and the temperature. For convenience both variables will be subdivided into three categories with piecewise linear membership functions, see figure 3.5.

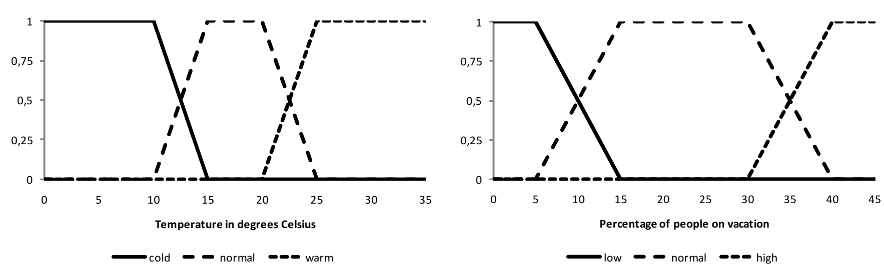

Figure 3.5: Membership functions of temperature and percentage of people on vacation

For the expert rules of this example, see table 3.1.

Rule No.	temperature	percentage of people on vacation	ice cream demand
1	cold	low	low
2	cold	normal	low
3	cold	high	normal
4	normal	low	low
5	normal	normal	normal
6	normal	high	normal
7	warm	low	normal
8	warm	normal	high
9	warm	high	high

Table 3.1: Expert rules ice cream demand

Suppose the temperature is 30 °C and 20 percent of the people is on vacation. These values result in a membership degree of 1 for "warm temperature" and "normal percentage of people on vacation", all the other membership functions have membership degree 0. Only the eighth rule of table 3.1 is active for these values. Consequently the prediction of ice cream demand is high with a support of 1.

In the range where the membership degrees are in $(0, 1)$ more than one rule is active at the same time. Such situations need another approach than the situation where only one rule is active. Now suppose the temperature is 12 °C and 38 percent of the people is on vacation. For temperature the category "cold" has a membership degree of 0.6 and the category "normal" one of 0.4. For percentage of people on vacation the category "normal" has a membership degree of 0.2 and the category "high" one of 0.8. Now the rules 2, 3, 5 and 6 are active at the same time. All rules have membership degrees for the input variables that are in $(0, 1)$ and therefore a method must be applied to assign a membership degree to the output variable. Rommelfanger examined various operators and came to the conclusion that the minimum operator should be used [6]. Applying the minimum operator to these rules assigns the following membership degrees to the output variable ice cream demand:

$$\begin{aligned}
&\text{rule}\,2: \min\left(\mu_{\text{temperature cold}},\, \mu_{\text{vacation normal}}\right) = \min\left(0.6,\, 0.2\right) = 0.2 \\
&\text{rule}\,3: \min\left(\mu_{\text{temperature cold}},\, \mu_{\text{vacation high}}\right) = \min\left(0.6,\, 0.8\right) = 0.6 \\
&\text{rule}\,5: \min\left(\mu_{\text{temperature normal}},\, \mu_{\text{vacation normal}}\right) = \min\left(0.4,\, 0.2\right) = 0.2 \\
&\text{rule}\,6: \min\left(\mu_{\text{temperature normal}},\, \mu_{\text{vacation high}}\right) = \min\left(0.4,\, 0.8\right) = 0.4
\end{aligned}$$

The output variable "ice cream demand" is also subdivided into three categories; "low", "normal" and "high". None of the rules 2, 3, 5 or 6 predicts a high ice cream demand, only rule 2 predicts a normal ice cream demand and the other three rules predict a normal ice cream demand. Since there is only one value which contributes to the ice cream demand category "low", there is no aggregation of the evaluations of the rules to be done. Rommelfanger examined different aggregation methods and suggests the use of the algebraic sum. The algebraic sum is defined as $\sum_{i=1}^{n} x_i - \prod_{i=1}^{n} x_i$ for $n > 1$.

The algebraic sum for the evaluation values of rules 3, 5 and 6 is $0.6 + 0.2 + 0.4 - 0.6 \times 0.2 \times 0.4 = 1.152$. To obtain the final evaluation values of the output variable, the aggregated values of the input variables have to be normalized. By doing so the final evaluation value for ice cream demand category "low" is $\frac{0.2}{0.2+1.152} = 0.15$, for category "normal" $\frac{1.152}{0.2+1.152} = 0.85$ and for category "high" $\frac{0}{0.2+1.152} = 0$.

Chapter 4

Hotel booking prediction

This chapter handles a case about prediction of hotel bookings by means of fuzzy prediction. The hotel bookings are done online at the web store of *weekendje weg* (http://www.weekendjeweg.nl). The number of reservations depends on several factors. Think about the price of a booking in combination with the month of the year; during Christmas and summer people are willing to pay a high price for a hotel room and in other cases they will not accept high prices. Because the reservations are done online, the web traffic is also an indicator for the number of bookings. Web traffic exists of several variables which will be described later on.

The accuracy of a fuzzy prediction system depends on the variables which includes the system, the expert rules and the membership functions. These have to be chosen carefully in order to achieve a high accuracy. The fuzzy prediction system is made with the two variables price and web traffic, and also the variables that concern web traffic.

In contrast to modelling of a fuzzy system as described in chapter 3 and illustrated in figure 3.1, the output will not be defuzzified. The modelling process and the results are described in the next paragraphs.

4.1 The data set

In order to predict the hotel room reservations we used the price and the web traffic, which is again subdivided into other variables. The data is collected over a period of three years (years A, B and C) on a monthly basis. To build the fuzzy logic model and to tune the different parameters years A and B are used as training data. To test the final model year C is used as test data.

Each variable in the data set has 36 different values, for every month one over a period of three years. These values are presented as percentage of the yearly average or as the percentage of the average for three the same months in three different years. For example when the price for a hotel room is 60 euro in year A and the yearly average is 50 euro, then the percentage is 120%. When the price in January for year A is 60 euro, year B is 50 euro, year C is 70 euro, then the percentage is $\frac{60}{(60+50+70)/3} \times 100\% = 100\%$ for January of year A. The model to predict the hotel room reservations includes the following variables:

- percentage of month average price over years ABC
- percentage of month average year visits
- percentage of month average year bouncerate
- percentage of month average year new visits
- percentage of month average year time on site
- percentage of month average pages per visit

These six variables are the input variables which predict the output variable *percentage of month average reservations over years ABC*. Only the first variable is about price, while the others concern web traffic. Bouncerate is the ratio between the number of times a visitor of the site immediately leaves the site after entering it and the number of times the site is visited. A visit is "new" when someone visits the site for the first time during a day.

4.2 Model building

To build a model with a high accuracy, that is, predicts the output value very well, one has to think carefully about the choices that have to be made. The choices are about which variables to include, how to define the membership functions and how to set up the expert rules. These choices are fundamental steps to the fuzzy prediction model and its accuracy.

The data set contains the variables which are important for the model. Five out of six input variables have to do with web traffic. The input of these variables have to be aggregated in order to obtain fuzzy quantities for the categories of web traffic. Consequently the web traffic and price have to be aggregated to obtain fuzzy quantities for the output variable. Figure 4.1 illustrates the model structure for the hotel booking prediction.

Once the variables are included in the system, their membership functions can be defined. Each variable is subdivided into three categories: "low", "normal" and "high". Furthermore the membership functions are piecewise linear functions. Defining a membership function can be done on basis of experience, on basis of the data or a combination of both. The university of Frankfurt am Main described a method to describe the behavior of experts quite accurately on

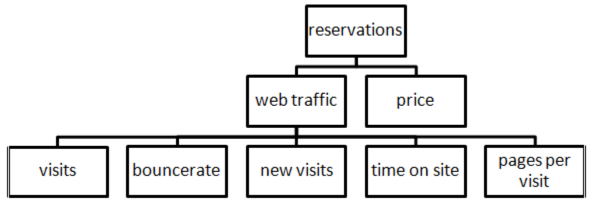

Figure 4.1: Model structure hotel booking prediction

basis of the data by means of quantiles, see [5]. We used this method, which is described as the first method in section 2.1, to define the membership functions. Recall that only year A and B are used for this, the data of year C will be used as test data. The 0.25-, 0.375-, 0.625- and 0.75-quantiles are used for the 0-cut and 1-cut. For the 0-cut and 1-cut of the membership functions, see table 4.1.

variable	normal / low	low / normal	high / normal	normal / high
% visits	94%	99%	103%	108%
% bouncerate	94%	97%	105%	108%
% new visits	96%	98%	100%	105%
% time on site	94%	96%	103%	105%
% pages per visit	94%	97%	102%	105%
% price over years ABC	98%	101%	103%	104%

Table 4.1: 0-cut / 1-cut membership functions

To clarify table 4.1 a graph of the membership functions of *percentage of month average year visits* is shown in figure 4.2.

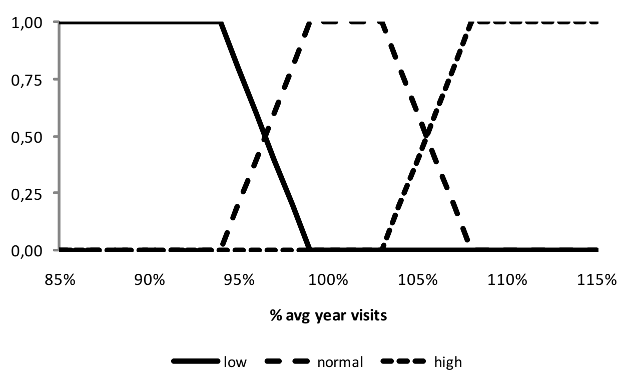

Figure 4.2: Membership functions of *percentage of month average year visits*

Finally the expert rules can be set up. According to figure 4.1 only web traffic and price are needed to predict the reservations. Although web traffic consists of several variables the expert rules for reservations can be done using web traffic and price only. These two variables have to be aggregated in order to obtain a fuzzy output for the reservations. During model building it turned out that when using three categories for reservations there was not enough subtlety in the output; minor changes in the input results in major changes in the output or the other way around. Therefore the output variable reservations is subdivided into five categories: "low", "normal minus", "normal", "normal plus" and "high".

Many different expert rules for the output variable can be made. The expert rules used for this fuzzy prediction system are made with common sense, see table 4.2.

Rule No.	price over years ABC	web traffic	reservations
1	low	low	normal
2	low	normal	normal plus
3	low	high	high
4	normal	low	normal minus
5	normal	normal	normal
6	normal	high	normal plus
7	high	low	low
8	high	normal	normal minus
9	high	high	normal

Table 4.2: Expert rules for reservations

The fuzzy quantities for the three categories of web traffic are the result of aggregating the five variables that concern web traffic. To predict the performance of web traffic the expert rules have to be set up. Again these rules are made with common sense. We choose to make one table which includes only two variables and one table with 25 rules which includes all the variables. After training the model on the training data (years A and B) one table will be chosen for the final model. For the expert rules for web traffic based on two variables, see table 4.3.

Rule No.	visits	bouncerate	web traffic
1	low	low	normal
2	low	normal	low
3	low	high	low
4	normal	low	high
5	normal	normal	normal
6	normal	high	low
7	high	low	high
8	high	normal	high
9	high	high	normal

Table 4.3: Expert rules for web traffic

For the expert rules based on all five variables concerning web traffic, see table 4.4.

Rule No.	visits	bouncerate	new visits	time on site	pages per visit	web traffic
1	low	low				normal
2	low	normal				low
3	low	high				low
4	normal	low				high
5	normal	normal				normal
6	normal	high				low
7	high	low				high
8	high	normal				high
9	high	high				normal
10			high	high	high	high
11			normal	normal	normal	normal
12			low	low	low	low
13		high	low			low
14		high	normal			low
15		low	high			high
16	low		low	low	low	low
17	normal		normal	normal	normal	normal
18	high		high	high	high	high
19	high		high			high
20	high		low			normal
21	low		high			normal
22	low		low			low
23	low				low	low
24	normal				normal	normal
25	high				high	high

Table 4.4: Expert rules for web traffic

Having tables 4.3 and 4.4 we can make two fuzzy prediction models. The predictions for the training data are most accurate for table 4.4, hence this table is included in the final model. For the results of the model analysis, see table 4.5.

# table 4 best	8
# table 5 best	14
# no difference	2
# minimal difference	4
# minimal difference AND table 4 best	3
# minimal difference AND table 5 best	1

Table 4.5: Results model analysis

Table 4.5 shows clearly that table 4.4 predicts the reservations the best. Although the predictions are not always accurate, they give a good indication. Furthermore it gives good insight what, and in what manner, affects the prediction; price or web traffic. Once the final model is known, the performance can be checked on the test data (year C).

4.3 Performance analysis

During model building a final model is chosen. The structure and variables of this model is shown in figure 4.1, the membership functions belonging to the input variables in table 4.1, and the expert rules in table 4.2 and table 4.4. To analyze the performance of this model, the model is tested on the test data. See table 4.6 for the results.

month of year C	percentage of month average years reservations over years ABC	predicted membership degrees reservations				
		low	normal minus	normal	normal plus	plus
1	99%	0	0	0	0.97	0.03
2	92%	0	0	0.67	0.33	0
3	83%	0	0	0	1	0
4	99%	0	0	0.14	0.29	0.57
5	99%	0	0.10	0.28	0.39	0.23
6	87%	0	0.01	0.34	0.60	0.05
7	94%	0	0	0.57	0.43	0
8	88%	0	0	0.45	0.55	0
9	94%	0	0.12	0.32	0.56	0
10	102%	0	0	0.29	0.71	0
11	97%	0	0	0.11	0.72	0.16
12	100%	0	0	0.20	0.31	0.49

Table 4.6: Predictions for reservations year C

Although the fuzzy logic prediction model predicts the reservations for years A and B fairly well, the predictions for year C are not close to the observed value. Possible explanations will be discussed in chapter *5 Conclusions and recommendations.*

Besides the predictions for the reservations the model is useable as a tool to determine the bottlenecks in a company. The model is built out of the components price and web traffic, where web traffic is a variable aggregated out of five others. For example, when *percentage of month average price over years ABC* is 100%, *percentage of month average year visits* is 96%, *percentage of month average year bouncerate* is 102%, *percentage of month average year new visits* is 100%, *percentage of month average year time on site* 106% and *percentage of month average pages per visit* is 102%, the system predicts an output which is illustrated in figure 4.3.

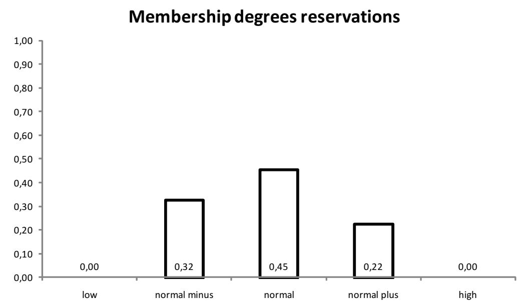

Figure 4.3: Reservations prediction

A closer look on the membership degrees of web traffic and price reveals which components have to be improved to obtain a higher reservation level. See figure 4.4 for the graphs of the membership degrees for price and web traffic.

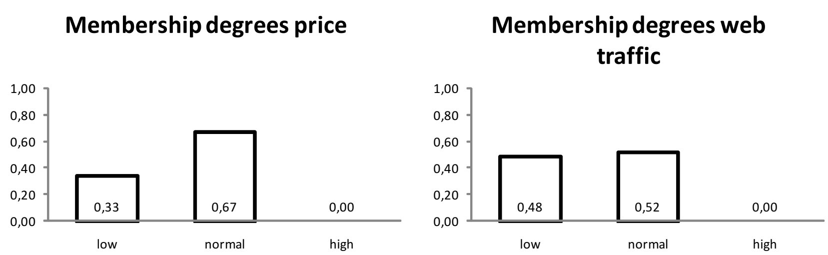

Figure 4.4: Membership degrees price and web traffic

Table 3 tells us that a low price and high web traffic results in a high reservation level. To improve the situation in figure 4.3, the price has to decrease and the web traffic has to increase. Web traffic is the result of aggregating five variables that concern web traffic. Figure 4.4 shows clearly how the input of these variables contribute to web traffic.

Chapter 5

Conclusions and recommendations

This paper describes the fuzzy logic concept and the fuzzy logic prediction system in general. The fuzzy logic prediction system is applied to hotel bookings which are done online. With this system we want to predict the reservations for hotel rooms on basis of the variables price and web traffic.

Although the developed model predicts the reservations fairly well for the training data, the predictions for the test data are clearly not accurate. An explanation could be that not the right decisions are made during model building or that fuzzy logic prediction is not an adequate method to predict hotel reservations. The decisions made during model building are: the variables to be included, the definition of the membership functions and the set up of the expert rules. All these decisions are not done by an expert, but made with common sense. The help of an expert would make the model more complete. It would make sense to extend the model with economic data and weather data such as economic growth, purchasing power and temperature. Most likely this will inrease the accuracy of the model. The membership functions are made by means of the quantiles of the data as described by [5]. Perhaps an other choice of the quantiles will lead to an increase of the accuracy, or a completely other approach for the definitions of the membership functions will do. Also an enhancement of the expert rules will lead to an increase of the accuracy.

A fuzzy logic prediction system makes use of a hierarchical structure of the variables. Because of this hierarchical order the bottlenecks within a system can be revealed. The system we made makes use of the variables price and web traffic, where web traffic is aggregated out of five other variables. Without knowing the input of these five variables the performance of web traffic can be analyzed. A comparison of web traffic and price shows in what manner these variales influence the reservations.

We think the idea to model fuzzy statements mathematically is a good starting point to build a prediction system. Especially when there are only linguistic variables available this method is a good choice. The membership functions that model the fuzzy statements can describe this fuzziness in a proper way. Problems can arise in the process of defining a membership function. How to catch the fuzziness of a variable in a mathematical function is not always clear and may need knowledge of an expert. Also the set up of the expert rules is a task which preferably must be done by an expert. Systems get easily too complex to set up the expert rules with common sense only. Having in mind earlier succesful studies of fuzzy logic prediction, see [3, 4, 6, 7], and provided that the fuzzy prediction model is built optimally, the predictions for the hotel reservations will be quite accurate.

Bibliography

[1] J.R. Benjamin and C.A. Cornell. *Probability Statistics and Decision Making in Civil Engineering.* New York: McGraw-Hill Book Inc, 1970

[2] J. Flach. *Ermittlung typisierter Fuzzy-Zugehörigkeitsfunktionen auf der Grundlage von Branchendaten als Komponente eines wissensbasierten Systems zur Unternehmensbeurteilung.* Diplomarbeit am Institut für Statistik und Mathematik der J.W. Goethe Universtität Frankfurt am Main, 1995.

[3] D. Gao, J. Chen and J. Lu. Fuzzy Prediction and Experimental Verification of Road Surface Cleaning Rate by Pure Waterjets. *Journal of Zhejiang University SCIENCE*, 6A(10):1115-1121, 2005

[4] S. Marsili-Libelli. Fuzzy Prediction of the Algal Blooms in the Orbetello Lagoon. *Environmental Modelling & Software*, 19:799-808, 2004

[5] H.T. Nguyen and E.A. Walker. *A First Course in Fuzzy Logic.* Chapman & Hall/CRC, Third edition, 2006

[6] H.J. Rommelfanger. *Fuzzy Logic Based Systems for Checking the Credit Solvency of Small Business Firms.* Institut für Statistik und Mathematik der J.W. Goethe Universtität Frankfurt am Main, 1999

[7] Z. Şen and A. Altunkaynak. Fuzzy System Modelling of Drinking Water Consumption Prediction. *Expert Systems with Applications*, 36:11745-11752, 2009

[8] L.A. Zadeh. Fuzzy Sets. *Information and Control*, 8:338-353, 1965