24th Bled eConference
eFuture:
Creating Solutions for the Individual, Organisations and Society
June 12 - 15, 2011; Bled, Slovenia

The Application of Bayesian Belief Networks

Barbara Krumay

WU, Vienna University of Economics and Business, Austria

barbara.krumay@wu.ac.at

Roman Brandtweiner

WU, Vienna University of Economics and Business, Austria

roman.brandtweiner@wu.ac.at

Abstract

The analysis of nominal data is often reduced to accumulation and description. Bayesian methods offer a possibility to analyse nominal data in a more sophisticated way. The possibility to indicate a structure via graphical representation, where variables are nodes and relationships are edges, enriches this method and makes it a powerful tool for data analysis. In this paper, an overview on Bayesian methods is given, the underlying rule is presented and some specialities will be discussed. Bayesian belief networks are described in brief and their potential to use them in case of uncertainty is presented. This includes not only the methods, but also possible applications in this context.

Keywords: Methods in Information Systems, Bayes Theorem

1 Introduction

Bayesian analyses are something special in the statistic world. They offer the possibility to include past experience, convenient assumptions or guesses which are represented by a prior distribution (Carlin and Louis, 2000). ”Bayesian probability theory is a set of rules for updating beliefs in an uncertain world” (Mayo and Mitrovic, 2001). Whenever the understanding of what is happening is unclear but causality is important it can be descried probabilistically (Charniak, 1991). The thinking of Bayes is based on the subjective probability which is the certainty in the personal valuation of an event (Mayo and Mitrovic, 2001). To understand Bayes's law it is necessary to know how conditional probability is defined:

$$P(b|a) = \frac{P(a|b) * P(b)}{P(a)}$$

- *P(a)* = Probability of ***a***

- $P(a|b)$ = Probability given *a* when *b* occurred

Based on this conditional probability it is possible to infer in both directions: deductive (a-priori) or inductive (a-posteriori). Deductive in this case means that it is possible to infer from a given event to the probability of the following event. Inductive means that the second event is given and it is possible to infer from this event on the prior. To describe the dependency of the variables from cause to effect, the product rule is used:

$$(a, b|c) = P(a|b, c)P(b|c) = P(b|a, c)P(a|c)$$

The Bayes rule describes the dependency from effect to cause:

$$P(a|b, c) = \frac{P(b|a, c)P(a|c)}{P(b|c)}$$

The Bayes's law can be used to test hypotheses:

$$P(H|E, c) = \frac{P(H|c) * P(E|H, c)}{P(E|c)}$$

P(H|E, c) = Posteriori probability which is the probability of hypothesis H after including the effect of the evidence E based on the observation c

- *P(H|c)* = A-priori probability which is *H* on given *c* alone
- *P(E|H, c*) = Likelihood or probability of the evidence when considering *H* and *c* is true
- *P(E|c)* = Is independent of *H* and can be seen as normalizing factor

Bayesian networks rely on the Bayes's law (Carlin and Louis, 2000). "Bayesian networks are directed acyclic graphs (DAGs) ... where the nodes are random variables, and certain independence assumptions hold" (Charniak, 1991). They combine graph theory and probabilistic theory in a graphical model using nodes and edges. The DAG models the probability distribution of dependent events. The events or random variables are states of affairs with the possible values true or false. The edges of the network represent the causal relationships between the nodes. Conditional dependencies are causal in both directions: from cause to (probable) effect and from effect to (probable) cause (Charniak, 1991). Three possibilities to connect variable (node) A with the next neighbours are (see also Figure 1) (Charniak, 1991):

1. linear or serial - edge from one node to another
2. convergent - edges from two or more different nodes to one
3. divergent - edges from one node to more than one others

Nodes are either d-separated or d-connected. D-separation of nodes occurs when A and C are nodes in a belief network and there are no direct paths between them or, in other words, there is always a node B in between. Hard evidence is stated when a node is 100 % in one state and 0 % in the other. Soft evidence is the opposite. When node B has hard evidence and is in between A and C, it blocks node C (Casella, 1985). Nodes which are not d-separated are d-connected (Charniak, 1991). "... if two things can cause the same state of affairs and have no other connection, the two things are independent" (Charniak, 1991). Probability distribution of Bayesian networks is based on the joint distribution and d-connection. The joint distribution of a Bayesian network is ”uniquely defined by the product of the individual distributions for each random variable” (Charniak, 1991). It is necessary to differentiate between beliefs and evidence. ”... belief is simply the conditional probability given the evidence” (Charniak, 1991). Beliefs are the subjective probability of a state based on the sum of evidence in a given situation. A-priori beliefs are based on the conditional probability tables of the network. Evidence, on the other hand, ”... is information about a current situation” (Charniak, 1991).

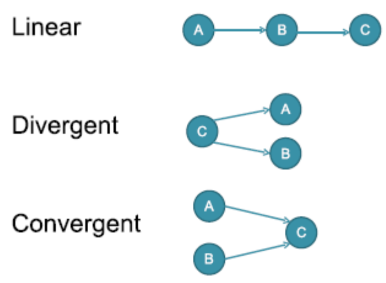


Figure 1: Causal Relationships / Connectors

The big advantage of belief networks is that it is possible to calculate the conditional probability of nodes in the network, but having only some of the nodes observed. Additionally, observed data can be combined from different data sets. The calculation is based on hidden Markov Chains or Monte Carlo model (John and Langley, 1995).

To design a Bayesian network is necessary to decide, which part of the potential variables should be included in the network in order to fulfil the required modelling goals. Structures should be defined but can also be gathered via learning. For each node in the network a conditional probability table (CPT) has to be calculated. This specifies a probability distribution of the network as product of local conditional probabilities:

$$P(A_1, ..., A_M) = \prod_{i=1}^{M} P(A_i|pa(A_i)) = \prod_{i=1}^{M} \Theta_i$$

where $pa(A_i)$ is the number of parent nodes.

Next, after constructing, the network is applied. To train the network, several learning situations can be simulated. Learning parameters can be based on maximum likelihood or counting the accumulations. This approach is legitimate when training data is complete:

$$\Theta_{ijk} = \frac{N_{ijk}}{N_{ik}}$$

N_{ijk} = cumulations of $X_i = j$ and $par(x_i) = k$ in empirical data D

N_{ik} = cumulations of $par(x_i) = k$ in empirical data D

Another possibility to train the networks is the Expectation-Maximization algorithm. It is possible to use it whenever the training data is incomplete but structures are available. It is processed in alternating steps of expectation (expected values for hidden nodes) and maximization (where the expected values for the hidden nodes are supposed to be observed).

Bayesian methods and especially Bayesian networks are used in different disciplines. Whenever a huge amount of numbers is necessary to calculate the probability of variables and not all data can be observed, they are applicable (Cesa-Bianchi et al., 1998). They are especially popular in decision scenarios, artificial intelligence projects, and medical studies (Rish et al., 2001). In the last years they encountered some popularity in spam protection (Jin et al., 2006), intrusion detection software (An et al., 2006) and the prediction of software productivity (Stamelos et al., 2003).

2 Theoretical Background and Data

As already stated in the beginning, collected nominal data reduces the possibilities for analysing, because it is not possible to compare ’yes’ and ’no’ (or 0 / 1 when binary data is used). Instead, a commonly used approach is to count occurrences to receive accumulations. This example is based on research of the impact of customer service offers on e-loyalty in B2C-e-commerce. In this example, we use some existing frameworks and models for our approach.

The e-loyalty framework (Gommans et al., 2001) covers important antecedents of e-loyalty. The influencing factors are website & technology, customer service, trust & security, brand building and value proposition (Figure 2). At first glance, this looks like a perfect framework to investigate the impact of online customer support on e-loyalty. The boundaries of this model are that there is no loop back to the influence of the loyalty on the further development of the customer support area.

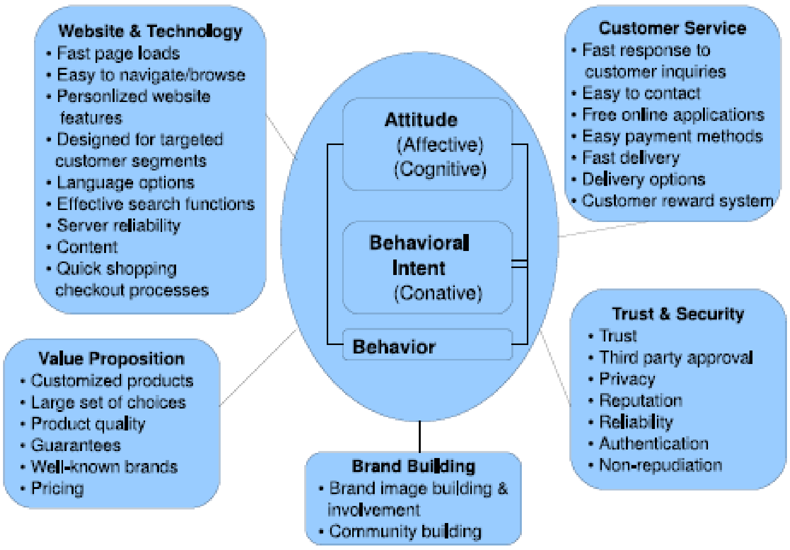


Figure 2: E-Loyalty Framework (Gommans et al., 2001)

The model of DeLone & McLean was developed to measure the success of information systems, fostering net benefits for users or usergroups (DeLone and McLean, 2004). The first model, which was created in 1992 (DeLone and McLean, 1992), was adapted to fit the e-commerce needs by adding a service quality factor (DeLone and McLean, 2004). In brief, the model consists of different components. On the one hand, there is the quality component, which is divided into system quality, information quality, and service quality. On the other hand, the usage-component is given, where the usage of the website and from usage resulting satisfaction is addressed. Last, the net benefits of the user are a component as well (Figure 3).

The components are described in brief as follows (based on DeLone and McLean, 2004):

- System quality: Desired characteristics of e-commerce system such as usability, availability, reliability, adaptability, response and download time
- Information quality: Content of the website, being personalized, complete, relevant, easy to understand, and secure
- Service quality: Overall support by the service provider, important for ecommerce, poor support leads to user defection
- Use: Measures every usage of the website (visits, navigation, information retrieval, execution of transactions)
- User satisfaction: Covers the full customer experience during the transaction process
- Net benefits: Balance between positive and negative impacts on users

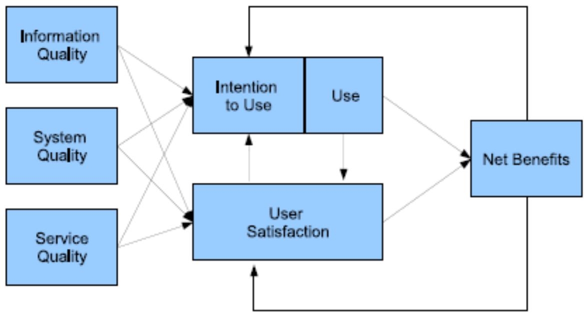


Figure 3: The DeLone & McLean IS Success Model (DeLone and McLean, 2004)

This model helps to sketch the different components of information systems. To apply the described research to this model, some modifications are necessary. As already mentioned above, customer support is a type of service and can be applied in the service-quality component. The online customer service area fits into both: information quality as well as system quality. This can be separated by the different tasks, the online customer support must fulfil: on the one hand it fulfils the information task, which is present in all phases of the transaction process. On the other hand the system quality of the online support is an influencing factor, because of the expectation of the customers that online services are available all the time and offer a high usability (Stern, 1996). The net benefits is the ultimate impact of the system, which can effect users as well as the company (DeLone and McLean, 2004). In this research it will be seen as net benefit for the company, which is the positive aspects of loyalty leading to positive word-of-mouth, less price sensitivity and lower switching probability as demonstrated in Figure 4. The more or less measurable success in this case are the net benefits of loyalty.

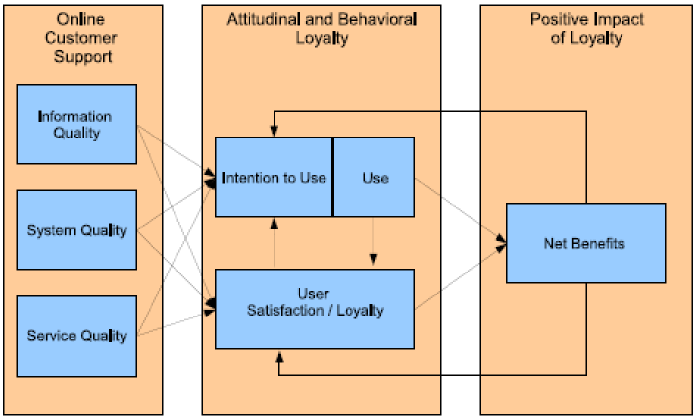


Figure 4: Research applied to DeLone & McLean IS Success Model (DeLone and McLean, 2004)

When looking at the DeLone & McLean model for IS success, it lacks some possibilities to use it for loyalty research. Nevertheless, it offers more possibilities to fulfil the goal of this research than the e-loyalty framework of Gommans et al. A combined approach, created from both models could support the goals in a more adequate way. Different items of the e-loyalty framework, which are in direct relation to customer service, were consulted and added to the DeLone & McLean model. As net benefits reduce low switching probability and positive word-of-mouth are added. The new model is shown in Figure 5. This model will be the basis for conducting the survey, processing the website analysis and building the prototype.

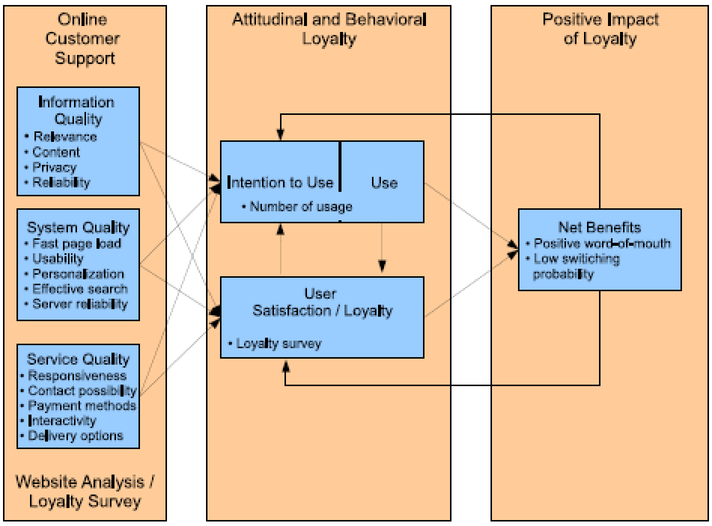


Figure 5: Transformation of the E-Loyalty Framework to the DeLone & McLean IS Success Model (DeLone and McLean, 2004)

3 Theoretical Background and Data

In this chapter the actual data analysis is described as an example. It is shown, how the given data was collected, how the network was built and how the process of applying was conducted. Problematic occurrences and difficulties are included as well as solutions to these problems. The conditional probabilities table and the interpretation are not given in this paper due to the huge amount of numbers. Only the network creation and redesign is described, because it is a possibility to reconfigure the basic thoughts.

3.1 Description of Given Data

The given data was collected using content analysis methods (Neuendorf, 2002) under the special conditions of Web site analysis (McMillan, 2008). The coding sheet was created on prior literature research. The model itself is created based on the DeLone & McLean success model for IS systems. The data was collected on Web sites of the companies of Fortune Global 500 2008 (see Fortune, 2008). The collection was done in autumn 2008 (October to December). Target to the analysis was the existence of specific parts of customer service on these Web sites. The existence was set to 'yes' and the absence of such variable to 'no'. 'Yes' or the existence was translated to 1 and 'no' to 0. Overall, 500 Web sites where accessed. Only 337 fulfilled the criterion of offering a product or service for end consumers (B2C). These 337 where then investigated in detail. Within these, 25 were not investigated due to language problems (Web sites in languages like Chinese or Taiwanese). The coders are not capable of this languages and a reliable result could not be guaranteed. Therefore the data basis consists of 312 data vectors holding 0 and 1. The net benefits (low switching probability, likeliness to spread word-of-mouth) were not part of the data analysis but taken from another data set.

3.2 Construction and Analysis

The construction process of the actual network was based on Helsper et al. (Helsper et al., 2005). First, the network was built on the theoretical model as described in figure 2.4. Every variable of the data analysis was set into relationship to a child node and conditional probability tables were built (see Figure 6). The R-package 'deal' is designed to create Bayesian networks. It uses Markov models and Monte Carlo models to recalculate the network and detect hidden nodes.

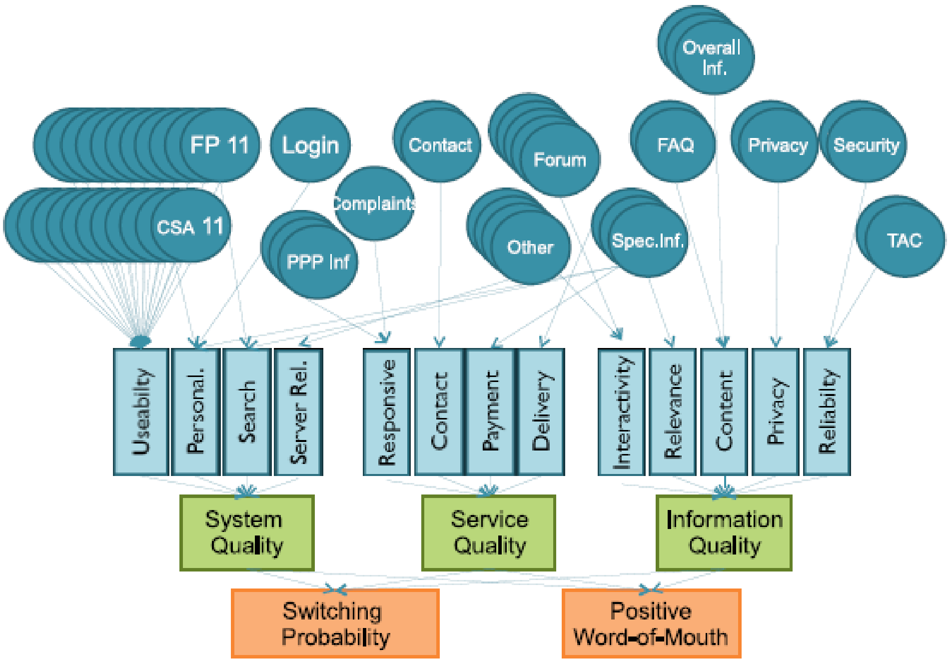


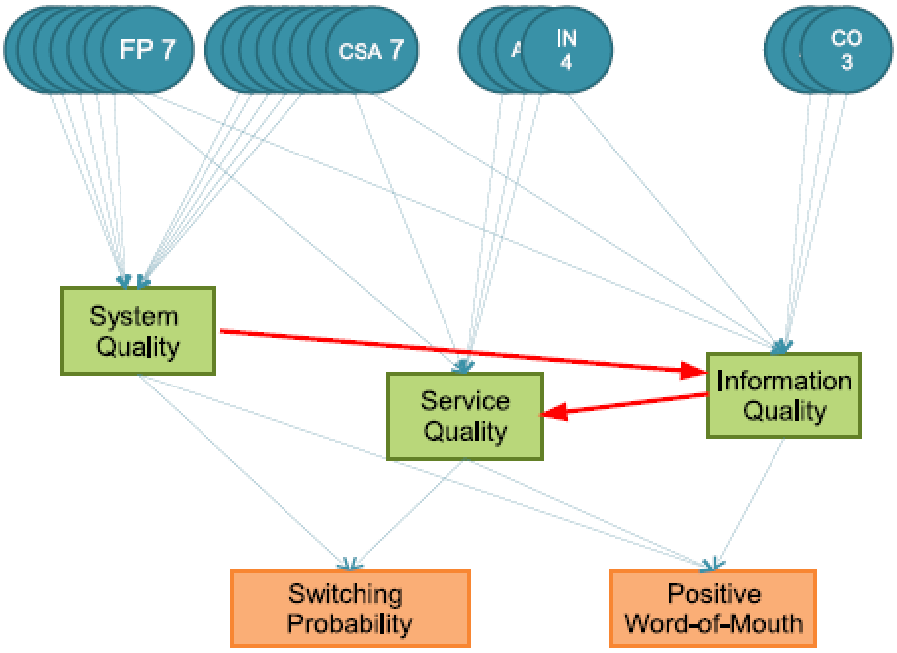
Figure 6: First Network

The huge amount of influencing items (73 at the beginning) caused problems according to the processing time of the software. Therefore - as a preprocessing step – a factor analysis was done to combine several variables to factors. The factors where estimated and complexity was reduced to 21 factors, representing the 73 variables. Factor scores are given in table 3.2.

Factors	dof	fit
FP1 – FP9	48	0.2671
CS1 – CS9	59	0.4387
IN1 – IN4	32	0.2562
CO1 – CO3	42	0.3353

Table 1: Factor analysis

Based on these factors, a new recalculation of the network was started. After five loops, a network with high network scores was detected, which represents the relationship and conditional probabilities in an appropriate way (see figure 3.2). In one of these loops, the in-between level of variables (usability, personalization, search ...) was deleted, leading to a better network score. This may be a sign that variable-level could be replaced totally by the operationalized items of the variables. The network structure changed in two ways: two causal relationships - one between system quality and information quality, another between information quality and service quality) were established (arrows in red, see figure 3.2). The causal relationship between information quality and switching probability disappeared. This seems to be an important evidence that the theoretical model was not representing the data - or the other way around. A further investigation is necessary to clarify these findings.



4 Conclusion and Discussion

Bayesian belief networks are very helpful in interpreting data under the conditions of incompleteness and uncertainty. They also offer the possibility to analyse nominal or binary data with causal relationships. After the design of the network it can be trained to find hidden nodes. In a loop the network can be trained to change the structure and find the best fitting network. At the end the result demonstrates the causal relationships, which must be discussed according to the theoretical model. Besides analysing nominal data, Bayesian belief networks are used in solving decision problems, control robots, or spam filtering. They are a simple but powerful way and therefor are worth being applied in the future more often. Limitations are lying in the method itself. A priori knowledge is necessary to be able to analyse data, which may be a drawback of this method.On the other hand, this necessity supports a well-defined design of the research and helps the researcher to gain a-priori knowledge about the topic of interest.

References

- **Books**

Carlin, Bradley P.; Louis, Thomas A. (2000). Bayes and Empirical Bayes Methods for Data Analysis. 2. Ed., Chapman and Hall CRC.

Neuendorf, Kimberly A. (2002). The Content Analysis Guidebook. Sage Publications, Inc. Thousand Oaks, California.

McMillan, Sally J. (2008). The Challenge of Applying Content Analysis to the World Wide Web. Sage Publications, Inc., Thousands Oaks, California.

Sterne, Jim. (1996). Customer Service on the Internet. Building Relationships. Increasing Loyalty and Staying Competitive. John Wiley & Sons, Inc.

- **Proceedings from conferences**

An, Xiangdong; Jutla, Dawn; Cercone, Nick. (2006). Privacy intrusion detection using dynamic Bayesian networks. In: ICEC ’06: Proceedings of the 8th international conference on Electronic commerce (pp. 208–215). ACM, New York, NY, USA.

Helsper, E. M.; von der Gaag, L. C.; Feelders, A. J. et al. (2005). Bringing Order into Bayesian-Network Construction. In: Proceedings of the 3rd international conference on Knowledge capture (pp. 121 – 128). International Conference On Knowledge Capture.

Jin, Xin; Xu, Anbang; Bie, R. et al. (2006). Spam email filtering with Bayesian belief network: using relevant words. In: International Conference on Granular Computing, IEEE (pp. 238 – 243).

John, George H.; Langley, Pat. (1995) Estimating Continuous Distribution in Bayesian Classifiers, Proceedings of the Eleventh Conference on Uncertainty in Artifical Intelligence (pp. 338-345). San Mateo, Morgan Kaufmann Publishers.

Rish, Irina; Hellerstein, Joseph; Thathachar Jayram. (2001). An analysis of data characteristics that affect naive Bayes performance. Proceedings of the Eighteenth Conference on Machine Learning-ICML2001. Morgan Kaufmann San Francisco. Morgan Kaufmann Publishers Inc.

- **Journal Articles**

[Case85] Casella, George. (1985). An Introduction to Empirical Bayes Data Analysis. In: The American Statistician 39 (2), pp. 83–87.

[CeHP98] Cesa-Bianchi, N.; Helmbold, D. P.; Panizza, S. (1998). On Bayes Methods for On-Line Boolean Prediction. In: Algorithmica 22, pp. 112–137.

[Char91] Charniak, Eugene: Bayesian Networks without Tears. (1991). AI Magazine 12 (4), pp. 50 – 63.

[DeMc92] DeLone, William H.; McLean, Ephraim R. (1992). Information Systems Success: The Quest for the Dependent Variable. In: Information Systems Research 3 (1), pp. 60 – 95.

[DeMc04] DeLone, William H.; McLean, Ephraim R. (2004). Measuring e-Commerce Success: Applying the DeLone & McLean Information Systems Success Model. In: International Journal of Electronic Commerce 9 (1), pp. 31–47.

[GoKS01] Gommans, Marcel; Krishman, Krish S.; Scheffold, Katrin B. (2001). From Brand Loyalty to E-Loyalty: A Conceptual Framework. Journal of Economic & Social Research 3 (1), pp. 43–58.

[MaMi01] Mayo, Michael; Mitrovic, Antonija. (2001). Optimising ITS Behaviour with Bayesian Networks and Decision Theory. In: International Journal of Artificial Intelligence in Education 12, pp. 124 – 153.

[StAD+03] Stamelos, I; Angelis, L.; Dimou, P. et al. (2003). On the use of Bayesian belief networks for the prediction of software productivity. In: Information and Software Technology 45 (1), pp. 51–60.

- **Web pages**

[Fort08] Fortune: Fortune Global 500. (download 22.10.2010), http://money.cnn.com/magazines/fortune/global500/2008/, 2008.