BUILDING A WEB EFFORT ESTIMATION MODEL THROUGH KNOWLEDGE ELICITATION

Emilia Mendes
Computer Science Department, The University of Auckland, Auckland, New Zealand

Keywords: Web engineering, Web effort estimation, Expert-based effort models, Knowledge elicitation, Case studies.

Abstract: OBJECTIVE – The objective of this paper is to describe a case study where Bayesian Networks (BNs) were used to construct an expert-based Web effort model. METHOD – We built a single-company BN model solely elicited from expert knowledge, where the domain experts were two experienced Web project managers from a medium-size Web company in Auckland, New Zealand. This model was validated using data from eleven past finished Web projects. RESULTS – The BN model has to date been successfully used to estimate effort for numerous Web projects. CONCLUSIONS – Our results suggest that, at least for the Web Company that participated in this case study, the use of a model that allows the representation of uncertainty, inherent in effort estimation, can outperform expert-based estimates. Another nine companies have also benefited from using Bayesian Networks, with very promising results.

1 INTRODUCTION

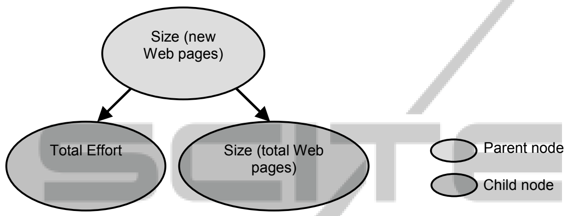
A cornerstone of Web project management is effort estimation, the process by which effort is forecasted and used as basis to predict costs and allocate resources effectively, so enabling projects to be delivered on time and within budget. Effort estimation is a very complex domain where the relationship between factors is non-deterministic and has an inherently uncertain nature. E.g. assuming there is a relationship between development effort and an application's size (e.g. number of Web pages, functionality), it is not necessarily true that increased effort will lead to larger size. However, as effort increases so does the probability of larger size. Effort estimation is a complex domain where corresponding decisions and predictions require reasoning with uncertainty.

Within the context of Web effort estimation, numerous studies investigated the use of effort prediction techniques. However, to date, only Mendes (2007a, 2007b, 2007c, 2008), Mendes and Mosley (2008), and Mendes et al. (2009) investigated the explicit inclusion, and use, of uncertainty, inherent to effort estimation, into models for Web effort estimation. Mendes (2007a, 2007b, 2007c) built a Hybrid Bayesian Network (BN) model (structure expert-driven and probabilities data-driven), which presented significantly superior predictions than the mean- and median-based effort (Mendes 2007b), multivariate regression (Mendes 2007a; 2007b; 2007c), case-based reasoning and classification and regression trees (Mendes 2007c). Mendes (2008), and Mendes and Mosley (2008) extended their previous work by building respectively four and eight BN models (combinations of Hybrid and data-driven). These models were not optimised, as previously done in Mendes (2007a, 2007b, 2007c), which might have been the reason why they presented significantly worse accuracy than regression-based models. Finally, Mendes et al. (2009) details a case study where a small expert-based Web effort estimation BN model was successfully used to estimate effort for projects developed by a small Web company in Auckland, New Zealand.

A BN is a model that supports reasoning with uncertainty due to the way in which it incorporates existing complex domain knowledge (Jensen, 1996). Herein, knowledge is represented using two parts. The first, the qualitative part, represents the structure of a BN as depicted by a directed acyclic graph (digraph) (see Fig. 1). The digraph's nodes represent the relevant variables (factors) in the domain being modelled, which can be of different types (e.g. observable or latent, categorical). The digraph's arcs represent the causal relationships between variables,

Mendes E..
BUILDING A WEB EFFORT ESTIMATION MODEL THROUGH KNOWLEDGE ELICITATION.
DOI: 10.5220/0003562701280135
In *Proceedings of the 13th International Conference on Enterprise Information Systems* (ICEIS-2011), pages 128-135
ISBN: 978-989-8425-55-3

where relationships are quantified probabilistically. The second, the quantitative part, associates a node conditional probability table (CPT) to each node, its probability distribution. A parent node's CPT describes the relative probability of each state (value); a child node's CPT describes the relative probability of each state conditional on every combination of states of its parents (e.g. in Fig. 1, the relative probability of Total effort (TE) being 'Low' conditional on Size (new Web pages) (SNWP) being 'Low' is 0.8). Each column in a CPT represents a conditional probability distribution and therefore its values sum up to 1 (Jensen, 1996).



CPT for node Size (new Web pages)	
Low	0.2
Medium	0.3
High	0.5

CPT for node Total Effort (TE)			
Size (new Web pages)	*Low*	*Medium*	*High*
Low	0.8	0.2	0.1
Medium	0.1	0.6	0.2
High	0.1	0.2	0.7

Figure 1: A small BN model and two CPTs.

Once a BN is specified, evidence (e.g. values) can be entered into any node, and probabilities for the remaining nodes automatically calculated using Bayes' rule (Pearl, 1988). Therefore BNs can be used for different types of reasoning, such as predictive and "what-if" analyses to investigate the impact that changes on some nodes have on others (Fenton et al. 2004).

Within the context of Web effort estimation there are issues with building data-driven or hybrid Bayesian models, as follows:

1. Any dataset used to build a BN model should be large enough to provide sufficient data capturing all (or most) relevant combinations of states amongst variables such that probabilities can be learnt from data, rather than elicited manually. Under such circumstance, it is very unlikely that the dataset would contain project data volunteered by only a single company (single-company dataset). As far as we know, the largest dataset of Web projects available is the Tukutuku dataset (195 projects) (Mendes et al., 2005). This dataset has been used to build data-driven and hybrid BN models; however results have not been encouraging overall, and we believe one of the reasons is due to the small size of this dataset.

2. Even when a large dataset is available, the next issue relates to the set of variables part of the dataset. It is unlikely that the variables identified, represent all the factors within a given domain (e.g. Web effort estimation) that are important for companies that are to use the data-driven or hybrid model created using this dataset. This was the case with the Tukutuku dataset, even though the selection of which variables to use had been informed by two surveys (Mendes et al., 2005). However, one could argue that if the model being created is hybrid, then new variables (factors) can be added to, and existing variables can be removed from the model. The problem is that every new variable added to the model represents a set of probabilities that need to be elicited from scratch, which may be a hugely time consuming task.

3. Different structure and probability learning algorithms can lead to different prediction accuracy (Mendes and Mosley, 2008); therefore one may need to use different models and compare their accuracy, which may also be a very time consuming task.

4. When using a hybrid model, the BN's structure should ideally be jointly elicited by more than one domain expert, preferably from more than one company, otherwise the model built may not be general enough to cater for a wide range of companies (Mendes and Mosley, 2008). There are situations, however, where it is not feasible to have several experts from different companies cooperatively working on a single BN structure. One such situation is when the companies involved are all consulting companies potentially sharing the same market. This was the case within the context of this research.

5. Ideally the probabilities used by the data-driven or hybrid models should be revisited by at least one domain expert, once they have been automatically learnt using the learning algorithms available in BN tools. However, depending on the complexity of the BN model, this may represent having to check thousands of probabilities, which may not be feasible. One way to alleviate this problem is to add additional factors to the BN model in order to reduce the number of causal relationships reaching child nodes; however, all probabilities for the additional factors would still need to be elicited from domain experts.

6. The choice of variable discretisation, structure learning algorithms, parameter estimation algorithms, and the number of categories used in the discretisation all affect the accuracy of the results and there are no clear-cut guidelines on what would be the best choice to employ. It may simply be dependent on the dataset being used, the amount of data available, and trial and error to find the best solution (Mendes and Mosley, 2008).

Therefore, given the abovementioned constraints, as part of a NZ-government-funded project on using Bayesian Networks to Web effort estimation, we decided to develop several expert-based company-specific Web effort BN models, with the participation of numerous local Web companies in the Auckland region, New Zealand. The development and successful deployment of one of these models is the subject and contribution of this paper. The model detailed herein, as will be described later on, is a large model containing 37 factors and over 40 causal relationships. This model is much more complex than the one presented in (Mendes et al., 2009), where an expert-based Web effort estimation model is described, comprising 15 factors and 14 causal relationships. This is the first time that a study in either Web or Software Engineering describes the creation and use of a large expert-based BN model. In addition, we also believe that our contribution goes beyond the area of Web engineering given that the process presented herein can also be used to build BN models for non-Web companies.

Note that we are not suggesting that data-driven and hybrid BN models should not be used. On the contrary, they have been successfully employed in numerous domains (Woodberry et al., 2004); however the specific domain context of this paper – that of Web effort estimation, provides other challenges (described above) that lead to the development of solely expert-driven BN models.

We would also like to point out that in our view Web and software development differ in a number of areas, such as: Application Characteristics, Primary Technologies Used, Approach to Quality Delivered, Development Process Drivers, Availability of the Application, Customers (Stakeholders), Update Rate (Maintenance Cycles), People Involved in Development, Architecture and Network, Disciplines Involved, Legal, Social, and Ethical Issues, and Information Structuring and Design. A detailed discussion on this issue is provided in (Mendes et al. 2005).

The remainder of the paper is organised as follows: Section 2 provides a description of the overall process used to build and validate BNs; Section 3 details this process, focusing on the expert-based Web Effort BN focus of this paper. Finally, conclusions and comments on future work are given in Section 4.

2 GENERAL PROCESS USED TO BUILD BNS

The BN presented in this paper was built and validated using an adaptation of the Knowledge Engineering of Bayesian Networks (KEBN) process proposed in (Woodberry et al., 2004). Within the context of this paper the author was the KE, and two Web project managers from a well-established Web company in Auckland were the DEs.

The three main steps within the adapted KEBN process are the Structural Development, Parameter Estimation, and Model Validation. This process iterates over these steps until a complete BN is built and validated. Each of these three steps is detailed below:

<u>Structural Development</u>: This step represents the qualitative component of a BN, which results in a graphical structure comprised of, in our case, the factors (nodes, variables) and causal relationships identified as fundamental for effort estimation of Web projects. In addition to identifying variables, their types (e.g. query variable, evidence variable) and causal relationships, this step also comprises the identification of the states (values) that each variable should take, and if they are discrete or continuous. In practice, currently available BN tools require that continuous variables be discretised by converting them into multinomial variables, also the case with the BN software used in this study. The BN's structure is refined through an iterative process. This structure construction process has been validated in previous studies (Druzdzel and van der Gaag, 2000, Fenton et al., 2004, Mahoney and Laskey, 1996; Neil et al., 2000, Woodberry et al., 2004) and uses the principles of problem solving employed in data modelling and software development (Studer et al., 1998). As will be detailed later, existing literature in Web effort estimation, and knowledge from the domain expert were employed to elicit the Web effort BN's structure. Throughout this step the knowledge engineer(s) also evaluate(s) the structure of the BN, done in two stages. The first entails checking whether: variables and their values have a clear meaning; all relevant variables have been included; variables are named conveniently; all

states are appropriate (exhaustive and exclusive); a check for any states that can be combined. The second stage entails reviewing the BN's graph structure (causal structure) to ensure that any identified d-separation dependencies comply with the types of variables used and causality assumptions. D-separation dependencies are used to identify variables influenced by evidence coming from other variables in the BN (Jensen, 1996; Pearl, 1988). Once the BN structure is assumed to be close to final knowledge engineers may still need to optimise this structure to reduce the number of probabilities that need to be elicited or learnt for the network. If optimisation is needed, techniques that change the causal structure (e.g. divorcing (Jensen, 1996)) are employed.

Parameter Estimation: This step represents the quantitative component of a BN, where conditional probabilities corresponding to the quantification of the relationships between variables (Jensen, 1996; Pearl, 1988) are obtained. Such probabilities can be attained via Expert Elicitation, automatically from data, from existing literature, or using a combination of these. When probabilities are elicited from scratch, or even if they only need to be revisited, this step can be very time consuming. In order to minimise the number of probabilities to be elicited some techniques have been proposed in the literature (Das, 2004; Druzdzel and van der Gaag, 2000; Tang and McCabe, 2007); however, as far as we are aware, there is no empirical evidence to date comparing their effectiveness for prediction, compared to probabilities elicited from scratch, using large and realistic BNs. This is one of the topics of our future work.

Model Validation: This step validates the BN that results from the two previous steps, and determines whether it is necessary to re-visit any of those steps. Two different validation methods are generally used - Model Walkthrough and Predictive Accuracy. Model walkthrough represents the use of real case scenarios that are prepared and used by domain experts to assess if the predictions provided by a BN correspond to the predictions experts would have chosen based on their own expertise. Success is measured as the frequency with which the BN's predicted value for a target variable (e.g. quality, effort) that has the highest probability corresponds to the experts' own assessment.

Predictive Accuracy uses past data (e.g. past project data), rather than scenarios, to obtain predictions. Data (evidence) is entered on the BN model, and success is measured as the frequency with which the BN's predicted value for a target variable (e.g. quality, effort) that has the highest probability corresponds to the actual past data. However, previous literature also documents a different measure of success, proposed by Pendharkar et al. (2005), and later used by Mendes (2007a, 2007c), and Mendes and Mosley (2009). This was the measure employed herein.

3 PROCESS USED TO BUILD THE EXPERT-BASED BN

This Section revisits the adapted KEBN process, detailing the tasks carried out for each of the three main steps that form part of that process. Before starting the elicitation of the Web effort BN model, the Domain Experts (DEs) participating were presented with an overview of Bayesian Network models, and examples of "what-if" scenarios using a made-up BN. This, we believe, facilitated the entire process as the use of an example, and the brief explanation of each of the steps in the KEBN process, provided a concrete understanding of what to expect. We also made it clear that the knowledge Engineers were facilitators of the process, and that the Web company's commitment was paramount for the success of the process. The entire process took 54 person hours to be completed, corresponding to nine 3-hour slots.

The DEs who took part in this case study were project managers of a well-established Web company in Auckland (New Zealand). The company had ~20 employees, and branches overseas. The project managers had each worked in Web development for more than 10 years. In addition, this company developed a wide range of Web applications, from static & multimedia-like to very large e-commerce solutions. They also used a wide range of Web technologies, thus enabling the development of Web 2.0 applications. Previous to using the BN model created, the effort estimates provided to clients would deviate from actual effort within the range of 20% to 60%.

Detailed Structural Development and Parameter Estimation: In order to identify the fundamental factors that the DEs took into account when preparing a project quote we used the set of variables from the Tukutuku dataset (Mendes et al., 2005) as a starting point. We first sketched them out on a white board, each one inside an oval shape, and then explained what each one meant within the context of the Tukutuku project. Our previous experience eliciting BNs in other domains (e.g. ecology) suggested that it was best to start with a few factors (even if they were not to be reused by the DE), rather than to use a "blank canvas" as a starting point. Once the Tukutuku variables had been sketched out and explained, the next step was to remove all variables that were not relevant for the DEs, followed by adding to the white board any additional variables (factors) suggested by them. We also documented descriptions for each of the factors suggested. Next, we identified the states that each factor would take. All states were discrete. Whenever a factor represented a measure of effort (e.g. Total effort), we also documented the effort range corresponding to each state, to avoid any future ambiguity. For example, 'very low' Total effort corresponded to 4+ to 10 person hours, etc. Once all states were identified and documented, it was time to elicit the cause and effect relationships. As a starting point to this task we used a simple medical example from (Jensen, 1996) (see Figure 2). This example clearly introduces one of the most important points to consider when identifying cause and effect relationships – timeline of events. If smoking is to be a cause of lung cancer, it is important that the cause precedes the effect. This may sound obvious with regard to the example used; however, it is our view that the use of this simple example significantly helped the DEs understand the notion of cause and effect, and how this related to Web effort estimation and the BN being elicited.

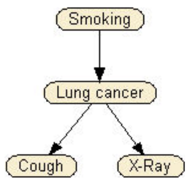


Figure 2: A small example of a cause & effect relationship.

Once the cause and effect relationships were identified, the original BN structure needed to be simplified in order to reduce the number of probabilities to be elicited. New nodes were suggested by the KE (names ending in '_O'), and validated by the DEs. The DEs also made a few more changes to some of the relationships. At this point the DEs seemed happy with the BN's causal structure and the work on eliciting the probabilities was initiated. All probabilities were created from scratch, and the probabilities elicitation took ~24 hours. While entering the probabilities, the DEs decided to re-visit the BN's causal structure after revisiting their effort estimation process; therefore a new iteration of the Structural Development step took place. The final BN causal structure is shown in Figure 3. Here we present the BN using belief bars rather than labelled factors, so readers can see the probabilities that were elicited. Note that this BN corresponds to the current model being used by the Web company (also validated, to be detailed next).

Detailed Model Validation: Both Model walkthrough and Predictive accuracy were used to validate the Web Effort BN model, where the former was the first type of validation to be employed. The DEs used four different scenarios to check whether the node Total_effort would provide the highest probability to the effort state that corresponded to the DEs' own suggestions. All scenarios were run successfully; however it was also necessary to use data from past projects, for which total effort was known, in order to check the model's calibration. A validation set containing data on 11 projects was used. The DEs selected a range of projects presenting different sizes and levels of complexity, where all 11 projects were representative of the types of projects developed by the Web company: five were small projects; two were medium, two large, and one very large.

For each project, evidence was entered in the BN model, and the effort range corresponding to the highest probability provided for 'Total Effort' was compared to that project's actual effort (see an example in Figure 4). The company had also defined the range of effort values associated with each of the categories used to measure 'Total Effort'. In the case of the company described herein, Medium effort corresponds to 25 to 40 person hours. Whenever actual effort did not fall within the effort range associated with the category with the highest probability, there was a mismatch; this meant that some probabilities needed to be adjusted. In order to know which nodes to target first we used a Sensitivity Analysis report, which provided the effect of each parent node upon a given query node. Within our context, the query node was 'Total Effort'.

Whenever probabilities were adjusted, we re-entered the evidence for each of the projects in the validation set that had already been used in the validation step to ensure that the calibration already carried out had not affected. This was done to ensure that each calibration would always be an improved upon the previous one. Within the scope of the model presented herein, of the 11 projects used for validation, only one required the model to be re-calibrated. This means that for all the 10 projects remaining, the BN model presented the highest probability to the effort range that contained the actual effort for the project being used for validation. Once all 11 projects were used to validate the model the DEs assumed that the Validation step was complete.

The BN model was completed in September 2009, and has been successfully used to estimate effort for new projects developed by the company. In addition, the two DEs changed their approach to estimating effort as follows: prior to using the BN model, these DEs had to elicit requirements using very short meetings with clients, given that these clients assumed that short meetings were enough in order to understand what the applications needed to provide once delivered. The DEs were also not fully aware of the factors that they subjectively took into account when preparing an effort estimate; therefore many times they ended up providing unrealistic estimates to clients.

Once the BN model was validated, the DEs started to use the model not only for obtaining better estimates than the ones previously prepared by subjective means, but also as means to guide their requirements elicitation meetings with prospective clients. They focused their questions targeting at obtaining evidence to be entered in the model as the requirements meetings took place; by doing so they basically had effort estimates that were practically ready to use for costing the projects, even when meeting with clients had short durations. Such change in approach provided extremely beneficial to the company given that all estimates provided using the model turned out to be more accurate on average than the ones previously obtained by subjective means.

Clients were not presented the model due to its complexity; however by entering evidence while a requirements elicitation meeting took place enabled the DEs to optimize their elicitation process by being focused and factor-driven.

We believe that the successful development of this Web effort BN model was greatly influenced by the commitment of the company, and also by the DEs' experience estimating effort.

4 CONCLUSIONS

This paper has presented a case study where a Bayesian Model for Web effort estimation was built using solely knowledge of two Domain Experts from a well-established Web company in Auckland, New Zealand.

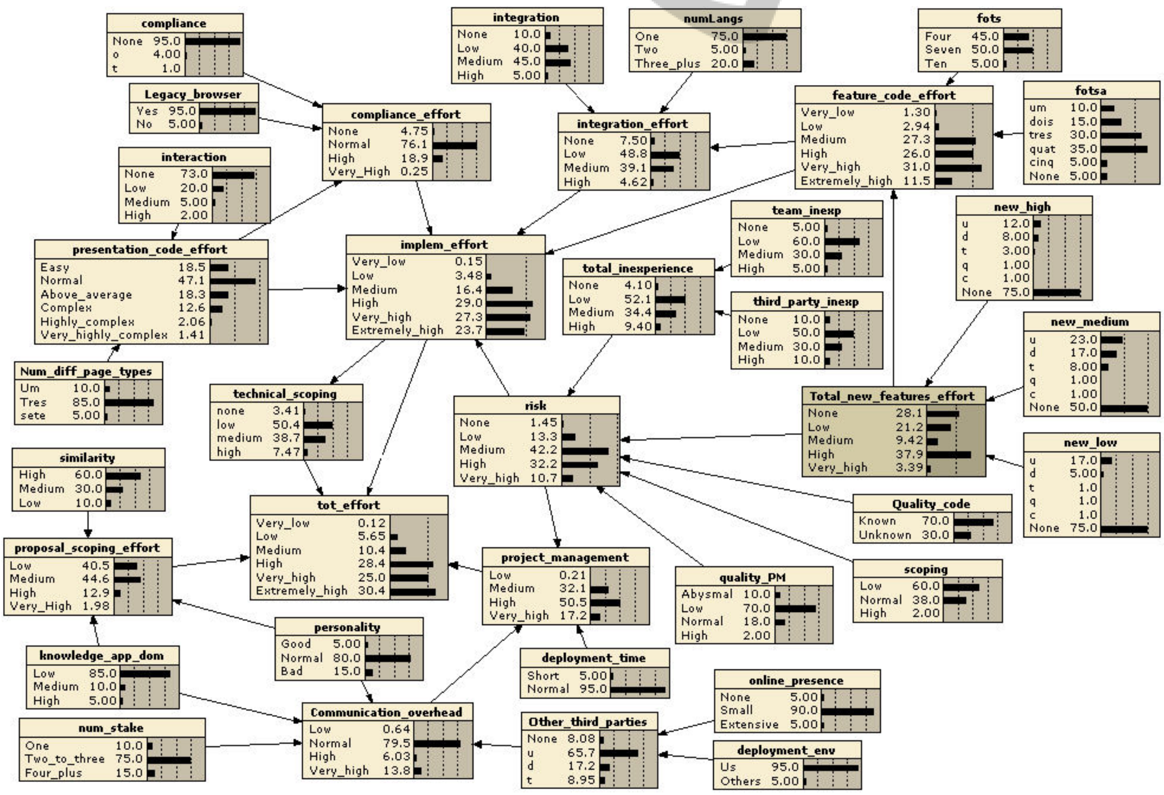


Figure 3: Final expert-based Web effort BN model.

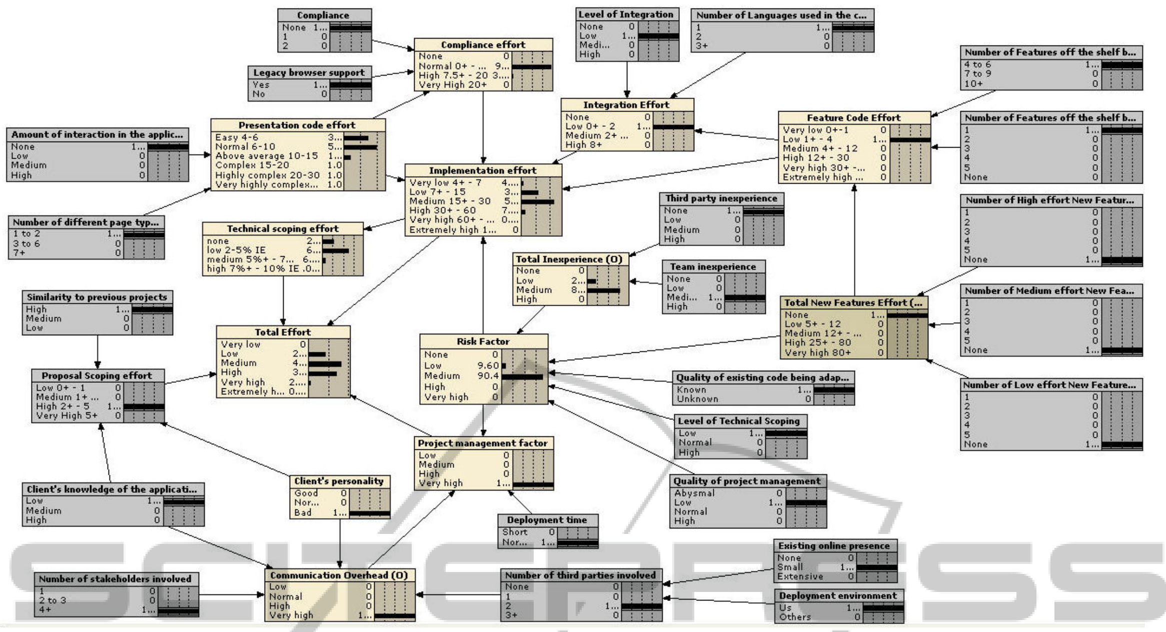


Figure 4: Example of evidence being entered in the Web effort BN model.

This model was developed using an adaptation of the knowledge engineering for Bayesian Networks process. Its causal structure went through three versions, because as the work progressed the experts' views on which factors were fundamental when they estimated effort also matured. Each session with the DEs lasted for no longer than 3 hours. The final BN model was calibrated using data on eleven past projects. These projects represented typical projects developed by the company, and believed by the experts to provide enough data for model calibration.

Since the model's adoption, it has been successfully used to provide effort quotes for the new Web projects managed by the company.

The entire process used to build and validate the BN model took 54 person hours, where the largest amount of time was spent eliciting the probabilities. This is an issue to those building BN models from domain expertise only, and is currently the focus of our future work.

The elicitation process enables experts to think deeply about their effort estimation process and the factors taken into account during that process, which in itself is already advantageous to a company. This has been pointed out to us not only by the domain experts whose model is presented herein, but also by other companies with which we worked on model elicitations.

To date we have completed the elicitation of six expert-driven Bayesian Models for Web effort estimation and have merged their causal structures in order to identify common Web effort predictors, and causal relationships (Baker and Mendes, 2010).

ACKNOWLEDGEMENTS

We thank the Web company who participated in this case study, and also all the participating companies in this research. This work was sponsored by the Royal Society of New Zealand (Marsden research grant 06-UOA-201).

REFERENCES

Baker, S., and Mendes, E.: Aggregating Expert-driven causal maps for Web Effort Estimation, Proceedings of the International Conference on Advanced Software Engineering & Its Applications (accepted for publication). (2010).

Das, B.: Generating Conditional Probabilities for Bayesian Networks: Easing the Knowledge Acquisition Problem, arxiv.org/pdf/cs/0411034v1 (accessed in 2008). (2004)

Druzdzel, M. J., and van der Gaag, L. C.: Building Probabilistic Networks: Where Do the Numbers Come From? IEEE Trans. on Knowledge and Data Engineering, 12(4), 481-486. (2000)

Fenton, N., Marsh, W., Neil, M., Cates, P., Forey, S. and Tailor, M.: Making Resource Decisions for Software Projects, Proc. ICSE'04, pp. 397-406. (2004)

Jensen, F. V.: An introduction to Bayesian networks. UCL Press, London. (1996)

Mahoney, S. M., and Laskey, K. B.: Network Engineering for Complex Belief Networks, Proc. Twelfth Annual Conference on Uncertainty in Artificial Intelligence, pp. 389-396. (1996).

Mendes, E.: Predicting Web Development Effort Using a Bayesian Network, Proceedings of EASE'07, pp. 83-93. (2007).

Mendes, E.: The Use of a Bayesian Network for Web Effort Estimation, Proceedings of International Conference on Web Engineering, pp. 90-104, LNCS 4607. (2007).

Mendes, E.: A Comparison of Techniques for Web Effort Estimation, Proceedings of the ACM/IEEE International Symposium on Empirical Software Engineering, pp. 334-343. (2007).

Mendes, E.: The Use of Bayesian Networks for Web Effort Estimation: Further Investigation, Proceedings of ICWE'08, pp. 2-3-216. (2008).

Mendes, E., and Mosley, N.: Bayesian Network Models for Web Effort Prediction: a Comparative Study, Transactions on Software Engineering, Vol. 34, Issue: 6, Nov/Dec 2008, pp. 723-737. (2008).

Mendes, E., Mosley, N. and Counsell, S.: The Need for Web Engineering: An Introduction, *Web Engineering*, Springer-Verlag, Mendes, E. and Mosley, N. (Eds.) ISBN: 3-540-28196-7, pp. 1-28. (2005).

Mendes, E., Mosley, N., and Counsell, S.: Investigating Web Size Metrics for Early Web Cost Estimation, Journal of Systems and Software, *77(2), 157-172.* (2005).

Mendes, E., Pollino, C., and Mosley, N.: Building an Expert-based Web Effort Estimation Model using Bayesian Networks, Proceedings of the EASE Conference, pp. 1-10. (2009).

Neil, M., Fenton, N., and Nielsen, L.: Building Large-scale bayesian networks, The knowledge Engineering Review, KER, Vol. 15, No. 3, 257-284. (2000).

Pearl J.: Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Mateo, CA. (1988).

Pendharkar, P. C., Subramanian, G. H., and Rodger, J.A.: A Probabilistic Model for Predicting Software Development Effort, IEEE Trans. Software Eng. Vol. 31, No. 7, 615-624. (2005)

Studer, R., Benjamins, V. R. and Fensel, D.: Knowledge engineering: principles and methods. Data & Knowledge Engineering, vol. 25, 161-197. (1998).

Tang, Z., and B. McCabe: Developing Complete Conditional Probability Tables from Fractional Data for Bayesian Belief Networks, Journal of Computing in Civil Engineering, 21(4), pp. 265-276. (2007).

Woodberry, O., A. Nicholson, K. Korb, and C. Pollino: Parameterising Bayesian Networks, Proc. Australian Conference on Artificial Intelligence, pp. 1101-1107. (2004).