UDC 004.413.5
Original Scientific Paper

Evaluating Expert Estimators Based on Elicited Competences

Hrvoje Karna *hkarna@fesb.hr*
Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
University of Split, Split, Croatia

Sven Gotovac *gotovac@fesb.hr*
Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture
University of Split, Split, Croatia

Abstract

Utilization of expert effort estimation approach shows promising results when it is applied to software development process. It is based on judgment and decision making process and due to comparative advantages extensively used especially in situations when classic models cannot be accounted for. This becomes even more accentuated in today's highly dynamical project environment. Confronted with these facts companies are placing ever greater focus on their employees, specifically on their competences. Competences are defined as knowledge, skills and abilities required to perform job assignments. During effort estimation process different underlying expert competences influence the outcome i.e. judgments they express. Special problem here is the elicitation, from an input collection, of those competences that are responsible for accurate estimates. Based on these findings different measures can be taken to enhance estimation process. The approach used in study presented in this paper was targeted at elicitation of expert estimator competences responsible for production of accurate estimates. Based on individual competences scores resulting from performed modeling experts were ranked using weighted scoring method and their performance evaluated. Results confirm that experts with higher scores in competences identified by applied models in general exhibit higher accuracy during estimation process. For the purpose of modeling data mining methods were used, specifically the multilayer perceptron neural network and the classification and regression decision tree algorithms. Among other, applied methods are suitable for the purpose of elicitation as in a sense they mimic the ways human brains operate. Data used in the study was collected from real projects in the company specialized for development of IT solutions in telecom domain. The proposed model, applied methodology for elicitation of expert competences and obtained results give evidence that in future such a model can be used in practice to reduce estimation error and enhance expert effort estimation.
Keywords: expert effort estimation, competences, elicitation, data mining, evaluation method.

1. Introduction

Expert effort estimation is extensively used in software project management. It is based on the premise that expert views can provide information and added knowledge in the situations where other methods fail for various reasons [2] and evidence shows that other methods don't guarantee better estimates [9]. In these situations projects highly dependent on the quality and reliability of inputs provided by the experts. The estimates depend upon particular experts providing it and environment in which estimates are produced [11], [41]. Estimates are solicited during sessions or meetings in different phases of project lifecycle. These sessions can take various forms, today the one most often used is some form of the Delphi method using the appropriate questionnaire with certain number of iterations. The method is based on elicitation of expert knowledge and judgments about the subject being analyzed [38].

Elicitation is the process of formulating the knowledge of an expert regarding one or more uncertain quantities and it is important because of its wide use [36].The technique has been studied within many disciplines. Examples of fields that have contributed to elicitation are decision analysis, psychology, risk analysis, Bayesian statistics and mathematics [13].

In situations where it is requested the expert constructs judgments in response to the requests being imposed. In doing this he retrieves relevant information from the memory and processes it to produce an answer. What information is retrieved and how it is processed depends on the issue in question and the context. The purpose of obtaining the experts judgment is to reduce the uncertainty that exists about this quantity. If the quantity is within the limits of experts knowledge there is greater chance that judgment will be right. This can be directly mapped to the situations which experts in field of software engineering are facing in everyday work. During effort estimation process in different phases of project they are typically requested to produce estimates about the efforts required to complete different project activities i.e. tasks. In such situations experts depend upon their knowledge, skills and abilities i.e. competences that they use during tacit thinking process to solve problems and/or make decisions [26].

Knowledge about the skills and competences of employees is of supreme importance for company success [22]. It enables efficient employee selection and staffing during project initialization as well as support during offer preparation and project planning [4]. Efficient competence management across organization ensures competitive position and a way to increase workforce productivity. Likewise structured competence framework provides transparency over available competences within organization and targeted development of those that are important through trainings and certifications [7].

Effort estimation is important part of software project management. The reliable effort estimates ensure planned project execution and compliance with the set time and budget constraints. Despite the long term efforts to produce accurate estimates based on formal and analogy based estimation methods expert estimation remains the most widely used technique of effort estimation [11]. Several reasons have contributed to this: studies consistently report that formal methods in comparison to expert estimation fail to produce more accurate estimates [9], expert estimation is easy to implement and finally expert estimation is more flexible regarding the type and format of the information used to produce estimates [19].

Artificial intelligence researchers have always been interested in developing intelligent decision aids with applications in various domains [3].The ways in which artificial intelligence has been applied to software engineering can be regarded as ways to optimize either the engineering process or its products. One of the areas in which artificial intelligence techniques have proved to be useful in software engineering research and practice is classification, learning and prediction for software engineering [12]. Here there has been great interest in modeling and predicting software costs as part of project planning. For example a wide variety of traditional machine learning techniques such as artificial neural networks, case-based reasoning and rule induction have been used for software project prediction [25]. An overview of machine learning techniques for software engineering can be found in the work of Menzies [27]. In this study we investigate the relationship between elicited experts competences and accuracy of their effort estimates. To figure out the relation between ones competences and success in effort estimation we have to apply methods of knowledge discovery. Data mining algorithms are such an example and as studies report software engineering can benefit from use of this approach [15], [23], [24]. Data mining in terms of software engineering consists of collecting software engineering data, extracting knowledge and when possible using this knowledge to improve the software engineering process. In this study we use two approaches: neural networks and decision trees. Based on the modeling results experts are ranked using weighted scoring model and they estimation performance evaluated.

The remaining part of this paper is organized as follows: Section 2 quotes the background in this area. Section 3 introduces the competence model used in the study. Section 4 describes the design of study. Section 5 explains the experiment setup and modeling performed in study. In Section 6 survey results and their implications are discussed. Section 7 presents evaluation of experts based on elicited competences. Finally in Section 8 we present discussion and in Section 9 conclusions and directions for the future research.

2. Background

Elicitation is performed in situations where it is not possible to make direct observations. It is a process of formulating the knowledge of an expert regarding one or more quantities [8]. In the context of expert effort estimation, "expert" is a person engaged on software project for which this analysis is performed. Experts statements regarding efforts required to perform given work tasks are produced during the decision making process that is hidden from outside observation [15]. During expert effort estimation these statements are essentially the only source of information. In this respect elicitation can be used to extract knowledge about experts inherent characteristics that play important role during this decision making process. How expert knowledge can be elicited accurately and reliably using the best current practices is the topic of dynamic research in various fields dating back for few decades [29].

Organizations have always been concerned about the competences of their employees. Today in a knowledge-based economy the success of organization mostly depends on workforce competences and competent employees are their main resource [2]. Competences are the best predictors of job performance [18]. In the same way estimating effort and therefore time and costs in different phases of a project is particularly important as these form a base on which decisions are made. The problem is when these estimates are not prepared by competent estimators. The present knowledge of how experts competences affect estimation accuracy arouses research interests [6].

Competence is a combination of knowledge, skills and process abilities that are causally linked and provide a base for job performance [16]. In certain form they represent a company's resource that could be exploited to gain competitive advantage [20]. While human resource development literature is mostly concerned with development of highly transferable generic competences that are required for most jobs or roles, particular company management is often emphasizing competences that are unique and company specific.

There are different competence models, usually in a form of a hierarchical catalogue that describes those that are desirable for organization and particular role [14]. Models depend on approach used to classify competences and can be one-dimensional or multi-dimensional which today are de facto standard [16]. Organizations use specialized IT-based systems to support the strategic competence management process [8]. Our previous study confirmed employees experience and role on a project give a high level notion of one's ability to successfully perform effort estimation tasks [12]. When it comes to competences required to perform estimation tasks questions are still incompletely answered.

This study was conducted with the aim to identify competences of professional software engineers engaged on projects within the company and occupying different positions that are important in determining one's ability to produce accurate estimates of efforts required to perform certain project tasks. Methodology applied in the study is explained next. Initially the set of projects executed within the company were selected and experts identified. From projects that entered analysis work item data were extracted, together these items form data set used during modeling phase. Each item contains reference to a project, item owner and assigned efforts, this allowed linking of an item to competence profile and calculation of estimation error. The company competence model was used for the purpose of structuring of expert competence profiles. Next, from initial collection data sets were created based on applied evaluation criteria and they entered modeling phase in which neural network and decision tree were used to build predictive models. Predictive models report overall accuracy and group of predictors (competences) ranked by their relative importance. Inputs from modeling phase were then used to rank individual estimators based on their scores using weighted scoring model. Finally, the individual experts were evaluated for their estimation performance (it was expected that experts with higher ranking levels in competences identified as important by the models were producing more accurate estimates during estimation process). Developed methodology together with insights gained through application of various advanced knowledge discovery and evaluation techniques used in this and related studies is intended to help software engineers improve their everyday work practice specifically the process of expert effort estimation.

3. Competences

Competency models are used to align individual capabilities with the competence of organization. These models are viewed as descriptive tools to identify the skills, knowledge, personal characteristics and behaviors that are required to efficiently perform a job in the organization [17]. The relation between competences and performance is shown on Figure 1.

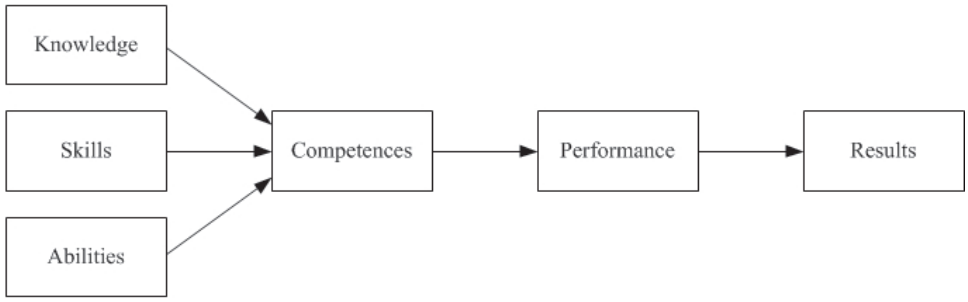

Figure 1. Relation between competences and professional results.

A company competence model establishes a common language which allows better communication between project managers and employees as it defines job expectations. It can also assist recruiting process where it can be used as some form of a guideline [13]. Knowing the skills, knowledge and abilities of employees allows better mapping of personnel to the company functions. For practical purposes of our study we are concerned with competences that a person working in a given occupational area should be able to do and demonstrate. Model of competences used by the company where study was performed covers tree segments: technical, professional and products and solutions competences. Each segment is further partitioned into sub-segments as it is shown in Table 1.

Segment	Competence	Description
Technical Competences	Operating Systems	Competence in use of operating systems
	Programming Languages	Competence in use of programming languages
	Development Environments	Competence in use of integrated development environments
	Database Systems	Competence in use of database management systems
	ALM Tools	Competence in use of application lifecycle management tools
	Project Process	Competence in application of different organization processes
Professional Competences	Development	Competence in different phases of software development process
	Operation and Maintenance	Competence in different operation and maintenance roles
	Project Types	Competence in various type of projects, current and past
	Role and Responsibility	Competence in relevant roles and responsibilities on projects, current and past
	Certifications	Level of certifications
Products and Solutions Competences	In-house Products and Solutions	Competence in development and use of in-house products and solutions
	Third Party Products and Solutions	Competence in development and use of third party products and solutions

Table 1. Model of competences used in the study.

It is important to note that in structured competence questionnaire used by the company to collect and store data each sub-segment represents an area that is further divided. For instance Programming Languages area specifically quotes languages in which skills are expected

(C, C++, C#, Java, etc.) or Project Types area quotes current and past types of projects that employee possibly participated in (Maintenance, R&D, Product development, etc.).

The process of creating of competencies collection is organized the following way: initially the structured competencies questionnaire is created and distributed to all employees in the department. All employees have to fill the questionnaire and return it to responsible person. The method of estimation is therefore a self-assessment and competency in each specific area can be marked with levels noted in Table 2. Once all questioners are collected they are imported to central department competence database.

Level	Description
1 Initial	Performs routine tasks with supervision and guidance
2 Basic	Performs range of tasks, supervision is required for more complex tasks
3 Intermediate	Performs some complex and non-routine tasks, able to manage the subject without constant guidance, can oversee the work of others
4 Advanced	Performs a wide range of complex and non-routine tasks, can train others in this subject
5 Expert	Performs all tasks, applies a significant range of fundamental principles and techniques, has strategic view and can train others in this subject

Table 2. Competence levels.

4. Study Design

As it is mentioned the study was conducted in the Croatian branch of international company specialized for development of IT solutions used by a number of different telecom companies. This department has more than 50 employees occupying different positions of whom majority are software engineers responsible for software development and maintenance tasks on different projects. The solutions are developed using Microsoft technology stack (Team Foundation Server, Visual Studio, SQL Server, C#, etc.). In total 32 experts from 10 projects participated in study. Details of projects included in study are displayed in Table 3.

Project	Duration (months)	Development method	Team size	LOC[1]	Size[2]	Precedentedness[3]
1	20,40	Sequential	6	92.091	Small	True
2	26,66	Sequential	6	123.693	Small	True
3	34,15	Sequential	9	46.668	Small	False
4	31,90	Sequential	9	249.732	Medium	True
5	61,02	Sequential	12	457.745	Large	False
6	7,80	Sequential	12	167.644	Medium	True
7	27,11	Sequential	8	148.409	Small	True
8	17,01	Iterative	23	261.781	Large	False
9	34,94	Sequential	6	263.485	Large	True
10	66,37	Sequential	6	125.967	Small	True

[1] Size expressed in number of physical Lines of Code, calculated using LocMetrics tool (www.locmetrics.com)

[2] Company internal classification of project size (determined by financial indicators)

[3] Parameter that indicated presence of similar projects already executed in department

Table 3. Details of projects included in study.

The work is organized in teams consisting of a project manager, software developers and testers. Solution architects, quality and configuration managers are department functions and engage in projects at different phases. From selected projects profile competences of in total 32 employees were randomly selected for later analysis. Characteristics of this competence data set are the following: a) out of 32 profiles 29 were males and 3 were females, b) roles occupied by employees in data set are: 4 project managers, 3 solution architects, 18 developers, 3 testers, 3 quality managers and 1 configuration manager and c) regarding the position level there were 16 seniors, 14 advanced and 2 junior engineers. Initially collected data set used for modeling purposes contained the total of 2090 items.

4.1. Data Sources

From the above listed projects development task and employee competence data required for the research were collected using following sources:

- Application lifecycle management tool implemented on projects that support development process. In this case it primarily served as a central place for collection of work item data. For this purpose on all considered projects Microsoft Team Foundation Server was used. Advantage that this and similar tools offer is the capability of various forms of data presentation, manipulation and export.
- The estimators competence data were gathered during company internal assessment procedure performed by dedicated department functions. The data collection was organized in form of a structured questionnaire that each employee received, had to fill and return to department. The questionnaire covered different aspects of employee profile of which major part was concerned with professional competences that are required to perform every day engineering tasks.

For employees involved on projects, collected competence data were structured in appropriate form, this made the total of 32 estimator profiles that entered the analysis. Input variables that are used to represent estimators competence characteristics are logically organized into segments as defined in Table 1. Data exported from tracking system contain both reference to an item owner (employee) and assigned efforts. This allowed two things: first, linking of an item to estimators competence profile and second, calculation of estimation error.

4.2. Evaluation Criteria

The performance of models can be assessed in many ways, the typical current practice of software organizations in field is to apply the Magnitude of Relative Error (MRE), Mean Magnitude of Relative Error (MMRE) and Prediction level (Pred) or similar measures [28]. In this study we use all three noted measures and they are defined next. Used measures are computed from the Relative Error (RE) which is the relative size of a difference between the actual and estimated value of individual effort [39]. The Magnitude of Relative Error (MRE) used on an initially collected data set containing 2090 items is defined as [3]:

$$MRE = \frac{abs(actual\ effort - estimated\ effort)}{actual\ effort} \tag{1}$$

The MRE is the most widely used measure of effort estimation accuracy [1], [5], [21], it is basically a degree of estimation error in an individual estimate. The Mean Magnitude of Relative Error (MMRE) can be calculated as follows:

$$MMRE = \text{mean}\ MRE \tag{2}$$

The Pred(X) is a complementary criterion at level X that defines the predictions having a relative error of less than or equal to level X. It is defined as:

$$Pred(X) = \frac{100}{N} \sum_{i}^{N} \begin{cases} 1\ if\ MRE\ \leq X/100 \\ 0 \qquad otherwise. \end{cases} \tag{3}$$

In general, Pred(X) reports the average percentage of estimates that were within X percent of the actual values, for example Pred(30) = 50% means that half the estimates are within 30 percent of the actual [1]. Based on this we were able to create an restricted set Pred(30) = 67,5% counting 1412 items. This restricted set was aimed at determining the possible differences in terms of resulting model accuracy and predictor importance when only a subset of most accurate estimates by given criteria is analyzed.

4.3. Data Mining

Building of the data mining model considered in this research required the definition of business objectives. In this case it is the identification of the expert estimators competences and their relative importance in producing reliable effort estimates. This business objective was mapped to data mining objective with intention to create such a model that could later be implemented in practice. Methodological framework consists of following phases:

- Data collection: during which both work item and employee competence data were collected. This stage therefore included export of project tasks, identification of involved team members and structuring of their competence data.
- Data preparation: at this stage data was processed according to specific needs of model building process. The end product is data set that contains efforts data of each item and related employee (item was assigned to). This way single resulting data set from all ten analysed projects was generated. At this stage outliers, extremes and missing data are handled, more details are provided in Section 5.
- Data partitioning: input data is randomly divided into two segments, training and test sets. From the initial data set the ratio of 2/3 of the data is used for the training (building of a model) and 1/3 for the testing phase (assessing of model performance).
- Model building: during this phase the predictive models are built using a MLP neural network and C&R decision tree algorithms and evaluated for predictive performance.

5. Experiment

In accordance with the data mining practice data was prepared to produce input sets comprising the total of 2090 records for initial and 1412 for restricted set corresponding to projects being analyzed. Variables considered in the input data sets are listed in Table 4:

	Variables	
Segment	*Name (Code)*	*Type*
Technical Competences	Operating Systems (OPS), Programming Languages (PRO), Development Environments (IDE), Database Systems (DBM), ALM Tools (ALM), Project Process (MET)	Predictor
Professional Competences	Development (DEV), Operation and Maintenance (OPR), Project Types (TYP), Role and Responsibility (ROL), Certifications (CER)	
Products and Solutions Competences	In-house Products and Solutions (IPS), Third Party Products and Solutions (TPS)	
	Magnitude of Relative Error (MRE)	Target

Table 4. Predictors and target in input data set.

From the input set of variables 13 are used as predictors and single variable (MRE) as a target. Experiment was conducted using IBM SPSS Modeler 14.2. For analyzed data sets a stream representing data flow was developed to perform experiment. The experiments followed the sequence in which data is initially fed into the stream after which it passed steps of preparation, transformation and partitioning before it entered the modeling element. The series of nodes represent operations that will be performed on the data, while links between the nodes indicate the direction of data flow. Typically a data stream is used to read data into SPSS Modeler after which it undergoes through a series of manipulations and then it is sent to a destination, such as a table or a viewer. SPSS Modeler supports the de facto industry standard, the CRoss-Industry Standard Process for Data Mining (CRISP-DM) [24].

Here we briefly describe the sequence of steps that data passes during modeling, it is also depicted on Figure 2. Data enters the stream through input node, here different types of sources are allowed (relational databases, Microsoft Excel, flat files etc.). Field labels and types are specified next together with measurement level used to describe characteristics of the data in a given field and its storage type. During this phase for each field the role in learning process is specified, being either input (predictor field) or target (predicted field).

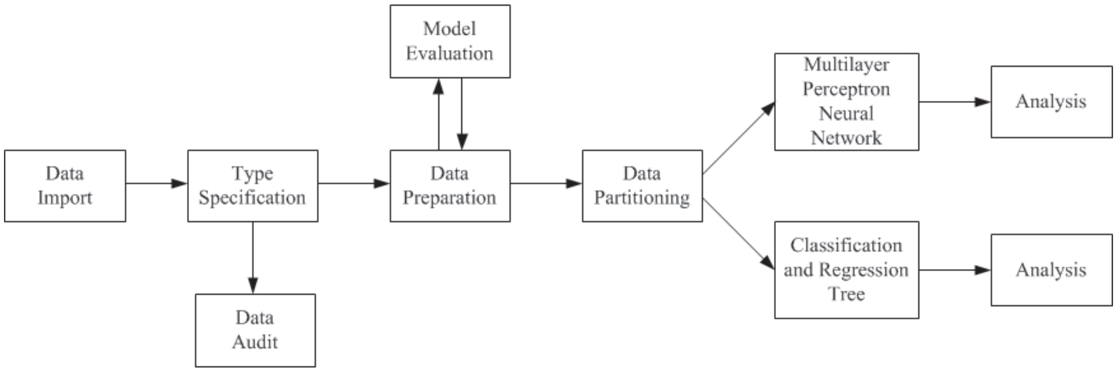

Figure 2. SPSS Modeler data stream.

Preparing data for analysis is one of the most important and time consuming steps. In this step data is analyzed and fixed by screening out problematic fields or those not likely to be used, and when appropriate transforming and constructing features. During initial data screening which was performed early in preparation phase missing data were handled. Typical cases are related to item owner and assigned efforts information. In both cases missing information prevents formation of valid entry that can be used in modeling. The reasons are the following, in case when item owner information is missing it is not possible to relate item and item owner while in case of missing actual and/or estimated effort it is not possible to calculate estimation error. This affects data set size as records with these missing data are useless and therefore removed from initially collected data set. Noted issue can be attributed to lack of discipline in the way owners handle items. During preparation of input and target fields for modeling to improve data quality missing continuous field values were replaced with mean while for categorical fields option to replace them with mode was used. Special handling is required in case of outliers and extremes. For the purpose of their identification and treatment the data audit node was used. Method used for outliers and extreme values detection was based on the number of standard deviations from the mean and was set to 3,0 for outliers and 5,0 for extremes. Approach used to handle outlier values was to replace them with cutoff value (set to 3.0 standard deviations). Initial set counted 75 while restricted data set counted 31 specified values. These actions can enhance the performance of algorithms used during modeling and improve predictive power of models. Specialized elements allow classification and comparison of different modeling methods. Partitioning is used to splits the data into separate subsets or samples for the training and testing stages of model building. By using one sample to generate the model and a separate sample to test it, we get an indication of how well the model will perform but also the indication of how well it will generalize to larger datasets that are similar to the current data if one is used in future. Modeler offers a variety of modeling methods taken from machine learning, artificial intelligence, and statistics. The methods allow us to derive new information from analyzed data and to develop predictive models. Each method has certain strengths and is best suited for particular types of problems. Based on performed evaluation following modeling elements implementing described data mining algorithms where used in the study:

- Neural network model uses a simplified model of the way the human brain processes information it implements the MPL (MultiLayer Perceptron) with the back propagation. Perceptron's architecture is organized into layers: input layer that receives information, hidden layer(s) and the output layer. During formation the model determines how the network connects the predictors to the target. This is done by hidden layer(s) that uses input values and modifies them using some weight. The activation function defines the output signal from the neuron. New value is then sent to the output layer where it is modified by some weight from connection between hidden and output layer. The back-propagation looks for the minimum of the error function. The combination of weights which minimize the error function is considered to be a solution of the learning problem.

- Decision tree model uses the Classification and Regression (C&R) algorithm, a tree-based classification and prediction method. Decision tree algorithm performs the procedure of examining the fields in dataset to find the ones that give the best classification or prediction by splitting data into subgroups. All splits are binary (only two subgroups). The process is applied recursively, splitting subgroups into smaller and smaller units until the tree is formed. The C&R algorithm minimizes the impurity at each step, where the node in the tree is considered "pure" if 100% cases in the node fall into a specific category of the target field. The output from a decision trees is a tree like structure that can be easily interpreted as a set of IF-THEN rules.

Application of data mining methods is well suited for our problem for several reasons. First of all they can operate on large data sets that are typical for research in field of software engineering. Next, they are used to extract knowledge from data and represent it in a form of rules for separation i.e. classification of input variable sets. This enables us to interpret and understand results of modeling. Finally, results from data mining process afterwards can be implemented in daily practice on projects, which can be a beneficial for business in multiple ways. In terms of our study these findings can enhance effort estimation process and thus result in more optimal utilization of project resources.

6. Results

The outputs resulting from the models report the relative importance of the top predictors. The importance of each predictor is relative to the model and it identifies the input variables that matter the most during prediction process. Results of modeling process for both neural network and decision tree performed on initial data set are displayed on Figure 3.

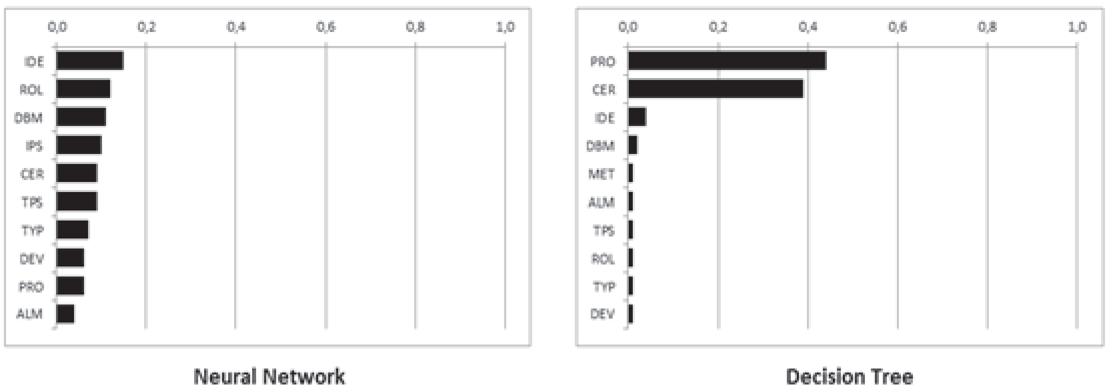

Figure 3. Relative importance of predictors in models for initial data set based on MRE.

The Multilayer Perceptron (MLP) neural network model returns the group of predictors with descending predictive power: IDE=0,15; ROL=0,12; DBM=0,11; IPS=0,10; CER=0,09; TPS=0,09; TYP=0,07; DEV=0,06; PRO=0,06; ALM=0,04. Resulting model has a single hidden layer with 10 neurons. Overall accuracy of resulting model is 57,9%. Although top predictors of estimation accuracy are competences i.e. know-how and skills in segments of development environment used on a project, current and previous roles and responsibilities, database management systems, in-house products and solutions know-how, certifications etc. from this model it is hard to designate typical predictors that could be used as classifiers.

On the other hand resulting model from the C&R decision tree clearly indicates predictors credible for the accurate effort estimates. This is obvious from distinctive values of their predictive importance: PRO=0,44; CER=0,39; IDE=0,04; DBM=0,02; MET=ALM=TPS=ROL=TYP=DEV=0,01. Model accuracy is similar to that of neural network. The resulting decision tree has depth=3 and can be expressed as:

```
CER in [ "Basic" ]
        DBM in [ "Advanced" "Basic" ]
        DBM in [ "Intermediate" ]
CER in [ "Advanced" "Expert" "Initial" "Intermediate" ]
        PRO in [ "Advanced" "Basic" "Expert" ]
        PRO in [ "Intermediate" ]
                IDE in [ "Advanced" ]
                IDE in [ "Intermediate" ]
```

The decision tree can be interpreted the following way: the most important predictor of one's effort estimation accuracy is the competence CER. This competence belongs to professional segment and indicates the level of employees certification in areas important for assigned job position. In resulting model this predictor is rated with second greatest predictive importance. CER divides the initial set into two subsets, those with basic level of certification and the rest that belong to group with levels initial, intermediate, advanced and expert. First subset is further divided by DBM criteria based on its corresponding levels. The important segment of second subset is further divided by PRO, competence that indicates experts level of competence in programming languages, this is predictor with greatest importance in resulting model. Those with PRO level intermediate are later divided into subsets by IDE. To conclude, the decision tree gives us simple and readable form of results.

The models built on restricted data set based on Pred(30) criteria have significantly greater overall accuracy of 82,1%. This indicates higher confidence in case models are used in prediction purposes. Results of modeling process for both neural network and decision tree based on Pred(30) criteria are displayed on Figure 4.

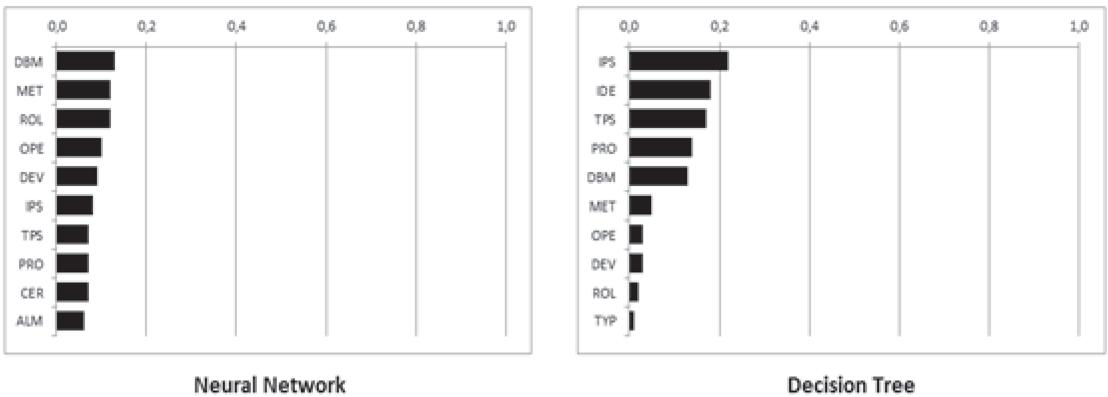

Figure 4. Relative importance of predictors in models for restricted data set based on Pred(30).

For the Multilayer Perceptron (MLP) model returns predictors with the predictive power: DBM=0,13; MET=ROL=0,12; OPE=0,10; DEV=0,09; IPS=0,08; TPS=PRO=CER=0,07; ALM=0,06. As in case with neural network model built on initial data set top predictors are closely grouped. Certain predictors as it is the case with DBM and ROL again appear at upper section, designating the most important predictors in the model. Certain predictors also have slightly greater average importance (example, DBM=0,11 in initial and DBM=0,13 in restricted model), while some other fall out of model (as in the case of IDE). In general, as it was the case with initial model the neural network model set of predictors is not favored for prediction purposed due to close grouping and consequently reduced importance distinction of returned predictors.

The resulting C&R decision tree model built on restricted data set is suitable for prediction. As it was the case with a tree model built on initial data set here we have group of prominent predictors. Their distinctive importance values are: IPS=0,22; IDE=0,18; TPS=0,17; PRO=0,14; DBM=13; MET=0,05; OPE=DEV=0,03; ROL=0,02 and TYP=0,01. C&R model again indicates PRO, IDE and DBM as competences important for success of effort estimation but now also IPS and TPS rank high in group. On the other hand certain predictors fall out of model, as it is the case with CER. The decision tree for restricted data set also has the depth=3 and is expressed as:

```
PRO in [ "Basic" "Intermediate" ]
        DBM in [ "Advanced" "Basic" ]
        DBM in [ "Intermediate" "Expert" ]
                IPS in [ "Basic" "Advanced" ]
                IPS in [ "Intermediate" "Expert" ]
PRO in [ "Advanced" "Initial" "Expert" ]
        IDE in [ "Advanced" "Basic" "Expert" ]
        IDE in [ "Initial" "Intermediate" ]
                TPS in [ "Advanced" "Expert" ]
                TPS in [ "Intermediate" ]
```

We can intuitively interpret these results in a way that technical competences have primary position in determination of one's estimation ability but also the competences in segment of products and solutions built either in-house or by third party vendors and extensively used in development process, where first are obviously more important.

Results of modeling performed based on both criteria indicate competences that can be used as predictors of experts effort estimation accuracy. In terms of neural network models they are relatively closely grouped by predictor importance what made it hard to derive conclusions. On the other hand decision tree models gave comprehensive models from which a set of rules can be derived. Those rules, in terms of prediction of expert estimators accuracy for both initial and restricted data sets can be expressed the following way: use level of competence primarily in segment of technical competences, particularly PRO, IDE and DBM, together with competences IPS and TPS from Products and Solutions segment as most relevant predictors. Only then consider group of predictors from segment of Professional competences that is foremost formed out of DEV, OPE and ROL competences. Other predictors can be ignored due to their low predictive power.

7. Evaluation of Expert Estimators

The evaluation of expert estimators was performed using weighted scoring matrix based on inputs from the modeling phase. The weighted matrix is a valuable decision-making tool that is used to evaluate alternatives based on specific evaluation criteria weighted by importance [37]. By evaluating alternatives based on their performance with respect to individual criteria, a value for the alternative can be identified. The values for each alternative can then be compared to create a rank order of their performance related to the criteria as a whole. The tool is important because it treats the criteria independently, helping avoid the over-influence or emphasis on specific individual criteria. Weighted scoring matrix used for evaluation purposes in this study is displayed in Table 5.

Criteria	Weight	Expert A		...		Expert N	
		rating	score	...	...	rating	score
C_1	w_1	$r_1(A)$	$w_1 \times r_1(A)$	...	...	$r_1(N)$	$w_1 \times r_1(N)$
C_2	w_2	$r_2(A)$	$w_2 \times r_2(A)$	...	...	$r_2(N)$	$w_2 \times r_2(N)$
C_3	w_3	$r_3(A)$	$w_3 \times r_3(A)$	...	...	$r_3(N)$	$w_3 \times r_3(N)$
...	...	...	...	...	...	...	...
...	...	...	...	$r_i(j)$	$w_i \times r_i(j)$	...	...
C_n	w_n	$r_n(A)$	$w_n \times r_n(A)$	...	...	$r_n(N)$	$w_n \times r_n(N)$
Total:	$\Sigma(w_1,...,w_n)$		$\Sigma(w_1 \times r_1(A),...$ $...,w_n \times r_n(A))$		...		$\Sigma(w_1 \times r_1(N),...$ $...,w_n \times r_n(N))$
	Rank:		R_A		...		R_N

Table 5. Expert weighted scoring matrix.

Here model predictors ($C_1,\dots, C_n$) are used as criteria and their respective importance values as weights ($w_1,\dots, w_n$). Typically, we want to focus our modeling efforts on the predictor fields that matter most but also consider those that matter less. The predictor importance helps us to do this by indicating the relative importance of each predictor in estimating the model. Since the values are relative, the sum of the weight values for all predictors is 1.0 i.e. $\Sigma(w_1,...,w_n) = 1.0$. Predictor importance relates to the importance of each predictor in making a prediction. For particular expert estimator rating $r_i(j)$ of an individual criteria C_i is based on levels as described in Table 2. Experts score in each criteria is calculated as product of corresponding weight and rating level, w_i x $r_i(j)$. Total score for each expert is obtained by summing over all criteria's $\Sigma(w_1 \text{ x } r_1(A),..., w_n \text{ x } r_n(A))$. Finally by sorting all totals we are able to determine experts rank R_j.

The ranking of the expert indicates his position within a group and weighted score in developed competences model. Here higher rank indicates greater development of competences indicated as important by the model. For the MRE model on Figure 5. we see resulting distribution with minimum scored values of 42,8%, grouping around scores of 75% and 85% where 17 estimators are positioned and maximum score of 99,6%. Pred(30) model built on restricted data set provided more balanced distribution of experts ranking. This allows us to perform fair separation of estimators into classes where higher class indicates greater chance of producing more accurate estimates. In this particular model minimum score is 48% and maximum 96,2%. There is clear distribution in range of scores between 65% and 95% in classes counting two to six estimators.

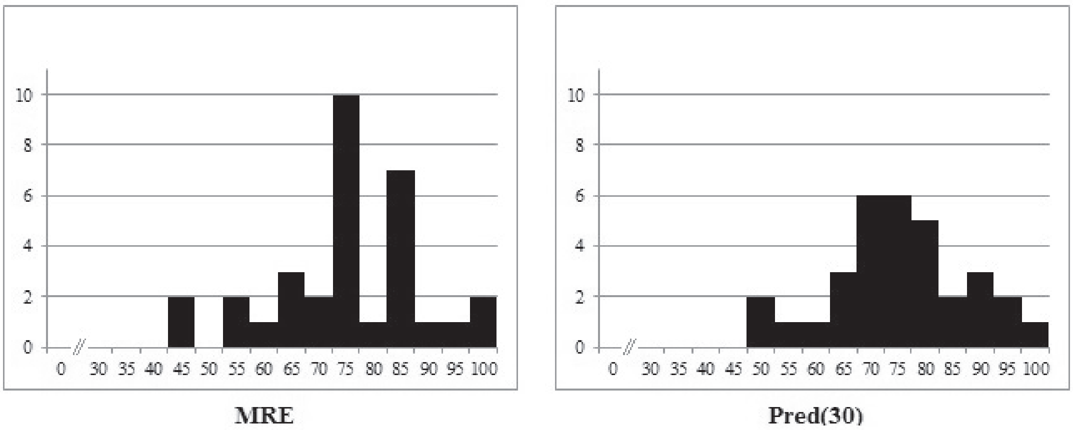

Figure 5. Distribution of experts based on evaluated competences for MRE and Pred(30) models.

Comparing results of estimation performance for particular estimator, based on mean values of estimation error i.e. MMRE we see significantly higher accuracy in case when second criteria is used. Here we can differentiate estimators groups in low end (45 to 65), middle (66 to 85) and high end (86-100) regions. Within these groups following results were found: estimators in low end and middle region showed smaller improvement (13 vs. 22%) in MMRE scores. On the other hand estimators in the high end region showed average improvement in MMRE scores of 28%. In certain cases for top ranked estimators in high end region there are significant individual MMRE score improvements of more than 30%. On the other hand the greatest overall improvement regarding the estimation accuracy is evidenced in those estimators grouped in middle class as it is the group counting most estimators. We can interpret these improved results as the consequence of application of more stricter criterion during elicitation process what can be useful in evaluating experts for the future estimates.

8. Discussion

This paper reports a detailed description of the study conducted with aim to develop predictive models in software engineering field of effort estimation that can be used in process of elicitation of expert competences responsible for reliable estimate production.

Motivation comes from the need of introducing modeled approach of assessing expert competences used in effort estimation. The study presents the methodology used to build predictive model, identifies adequate data mining techniques and provides a form for classification of expert competences used as predictors during elicitation process. Elicited competences can be used as a criterion to assess reliability of experts efforts estimates. This is particularly important in situations when group estimation is performed as higher experts rank can suggest more accurate estimation.

The methodology was applied on the real data. Segments belonging to effort quantities were extracted from the tracking system implemented on projects while the data concerning expert competences was collected via structured competence questionnaire. The proposed methodology for elicitation of expert competencies and obtained results of a performed study confirm its validity.

From a practical standpoint removing segment of less accurate estimates based on applied criteria produces a more efficient prediction set i.e. competences ranked by importance used to rank estimators. Results confirm our initial premise that applied evaluation of expert estimators based on elicited competences can be used to enhance estimation process.

9. Conclusions and Future Directions

In this study we have shown that the proposed methodology used to build predictive models based on which expert estimators are evaluated improved expert effort estimation accuracy. These results indicate that such models can be implemented in practice to improve management of software projects.

Results of this and future studies support the development of a model for enhanced expert effort estimation. Based on better understanding of effects that estimators competences have on reliability of effort estimates it would allow the application of corrective measures at early stage of estimation process. Such a model is intended to enhance reliability of effort estimates and could be applied to everyday practice of software engineers.

In future, based on this and our other studies [21] we are planning to do more experiments to better understand the implications of task and project parameters on effort estimation. Additionally, we plan to introduce additional set of delta variables into models that will help determine possible direction in which individual errors have to be corrected. Also, we hope to repeat experiments on other, if possible larger, data sets collected from different environments. This would facilitate building a more general and robust models.

References

[1] Basha, S; Ponnurangam, D. Analysis of Empirical Software Effort Estimation Models. International Journal of Computer Science and Information Security, 7(3), 2010.

[2] Booker M. J; Anderson, C. M; Meyer, A. M. The Role of Expert Knowledge in Uncertainty Quantification - Are We Adding More Uncertainty or More Understanding?. Army Conference on Applied Statistics, Santa Fe, NM, 2001.

[3] Braziunas, D. Computational Approaches to Preference Elicitation. Technical report, 2006, URL: http://www.cs.toronto.edu/~darius/papers/PreferenceElicitation.pdf

[4] Chouhan V. S; Srivastava, S. Understanding Competencies and Competency Modeling – A Literature Survey. International Organization of Scientific Research Journal of Business and Management, 16 (1):14-22, 2014.

[5] Conte, S. D; Dunsmore, H. E; Shen, V. Y. Software Engineering Metrics and Models. Benjamin-Cummings, Menlo Park, CA, 1986.

[6] Dorn, J; Pichlmair, M; Schimper, K; Tellioglu, H. Supporting Competence Management in Software Projects. Proceedings of International Conference on Concurrent Enterprising, 451-458, 2008.

[7] Ferrucci, F. et al. Genetic Programming for Effort Estimation an Analysis of the Impact of Different Fitness Functions. 2nd International Symposium on Search Based Software Engineering, Benevento, Italy, 89-98, 2010.

[8] Garthwaite, H. P; Kadane J. B; O'Hagan, A. Elicitation. Carnegie Mellon University, Pittsburgh, 2004, URL: http://www.stat.cmu.edu/tr/tr808/old808.pdf

[9] Giammalvo, P. Creating a Competency Assessment for Cost Estimators. Project Management World Journal, 1(2), 2012.

[10] Hale, J. Performance-Based Certification: How to Design a Valid, Defensible, Cost-Effective Program. Pfeiffer, 2012.

[11] Humphrey,S; W., Konrad, D. M; Over, W., J; Peterson, C. W. Future Directions in Process Improvement. CrossTalk The Journal of Defense Software Engineering, 20(2):17-22, 2007.

[12] Harman, M. The role of Artificial Intelligence in Software Engineering. First International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE), 2012.

[13] Hora, C. S. Expert Judgment Elicitation. Create Homeland Security Center - Research Project Summaries, Paper 65, 2008.

[14] Hustad, E; Munkvold, B; Moll, B. Using IT for Strategic Competence Management - Potential Benefits and Challenges. Proceedings of European Conference on Information Systems, 2004.

[15] Jørgensen, M. et al. Human judgment in effort estimation of software projects. Presented at Beg, Borrow, or Steal Workshop, International Conference on Software Engineering, June, Limerick, Ireland, 2000.

[16] Jørgensen, M. A review of studies on expert estimation of software development effort. The Journal of Systems and Software, 70(1-2):37-60, 2004.

[17] Jørgensen, M; Sjoberg, D. I. K. Impact of experience on maintenance skills. The Journal of Software Maintenance and Evolution, 14:1-24, 2002.

[18] Jørgensen, M; Boehm, B; Rifkin, S. Software Development Effort Estimation: Formal Models or Expert Judgment?. IEEE Software, 26(2):14-19, 2009.

[19] Jørgensen, M. Estimation of Software Development Work Effort: Evidence on Expert Judgment and Formal Models. International Journal of Forecasting, 23(3):449-462, 2007.

[20] Karna, H; Gotovac, S. Mining Competences of Expert Estimators, Proceedings of the 23rd International Conference on Information Systems Development (ISD2014), Varaždin, Croatia, 14-22, 2014.

[21] Karna, H; Gotovac, S. Modeling Effort Estimation of Software Projects. Proceedings of the 22nd Conference on Software, Telecommunications and Computer Networks (SoftCOM 2014), Split, Croatia, 2014.

[22] Kenexa Corporation an IBM Company. How Competencies Enhance Performance Management, 2012, URL: http://www.kenexa.com.

[23] Klendauer, R. et al. Towards a competency model for requirements analysts. Information Systems Journal, 22(6):475-503, 2012.

[24] Maimon, O; Rokach, L. Data Mining and Knowledge Discovery in Real Life Applications. Second Edition, Springer, New York, 2010.

[25] Mair, C; Kadoda, G; Phalp, K; Webster, S. An investigation of machine learning based prediction systems, The Journal of Systems and Software, 53(1):23–29, 2000.

[26] Mendes, E. Cost Estimation of Web Applications through Knowledge Elicitation. Enterprise Information Systems - Lecture Notes in Business Information Processing, 102:315-329, 2012.

[27] Menzies, T. Practical machine learning for software engineering and knowledge engineering. Handbook of Software Engineering and Knowledge Engineering, World-Scientific, 2001, URL: http://menzies.us/pdf/00ml.pdf.

[28] Menzies, T; Chen, Z; Hihn, J; Lum, K. Selecting Best Practices for Effort Estimation, IEEE Transactions on Software Engineering, 32(11):883-895 2006.

[29] Meyer, A. M; Booker, M. J. Eliciting and Analyzing Expert Judgment: A Practical Guide. ASA-SIAM Series on Statistics and Applied Probability, 2001.

[30] Layman, L. et al. Mining Software Effort Data: Preliminary Analysis of Visual Studio Team System Data. Proceedings of the 2008 International Working Conference on Mining Software Repositories, 43-46, 2008.

[31] Le Deist, D. F; Winterton J. What is competence?. Human Resource Development International, 8(1):27-46, 2005.

[32] Lucia, A.D; Lepsinger, R. The Art and Science of Competency Models: Pinpointing Critical Success Factors in an Organization. Pfeiffer, San Francisco, CA, 1999.

[33] McClelland D.C. Testing for Competence Rather Than for Intelligence. American Psychologist, 28(1):1-14, 1973.

[34] Jørgensen, M; Moløkken-Østvold, K. J. Expert Estimation of Web-Development Projects - Are Software Professionals in Technical Roles More Optimistic Than Those in Non-Technical Roles, Empirical Software Engineering, 10:7-29, 2005.

[35] Nader, D. A; Tushman, M. The Organization of the future: strategic imperatives and core competencies for the 21st century. Organizational Dynamics, 27(1):45-58, 1999.

[36] O'Hagan, A. Research in elicitation. In Bayesian Statistics and its Applications, 375-382. Anamaya, New Delhi, 2006.

[37] Pradeep K. S. V. Data Mining- Scoring (Weighted Scoring Tables). Applexus Technologies (P) Ltd, SAP Community Network, 2011.

[38] Scapolo, F; Miles, I. Eliciting Experts' Knowledge: A Comparison of Two Methods, EU-US Seminar: New Technology Foresight, Forecasting and Assessment Methods, Paper 1, Seville, 2004.

[39] Singh, B. K; Misra, A.K. An Alternate Soft Computing Approach for Efforts Estimation International Journal of Innovation, Management and Technology, 3(3), 2012.

[40] Stensrud, E. et al. A Further Empirical Investigation of the Relationship Between MRE and Project Size. Empirical Software Engineering, 8(2):139-161, 2003.

[41] Wang, Y. On Laws of Work Organization in Human Cooperation. International Journal of Cognitive Informatics and Natural Intelligence, 1(2):1-15, 2007.

[42] Winterton, J; Le Deist, F; Stringfellow, E. Typology of knowledge, skills and competences: Clarification of the concept and prototype. Luxembourg, 2006.

[43] Xie, T; Thummalapenta, S; Lo, D; Chao Liu, C. Data Mining for Software Engineering. IEEE Computer, 42(8):35-42, 2009.