Machine learning applied to quality management—A study in ship repair domain

Alira Srdoč a,*, Ivan Bratko b,c, Alojzij Sluga d

a Department of Research and Development, “3.MAJ” Shipyard, Ljubljanska 3, 51000 Rijeka, Croatia
b Faculty of Computer and Information Science, University of Ljubljana, Tržaska cesta 25, 1000 Ljubljana, Slovenia
c Department of Intelligent Systems, Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
d Faculty of Mechanical Engineering, University of Ljubljana, Askerceva 6, 1000 Ljubljana, Slovenia

Received 24 March 2005; accepted 28 September 2006
Available online 17 November 2006

Abstract

The awareness about the importance of knowledge within the quality management community is increasing. For example, the Malcolm Baldrige Criteria for Performance Excellence recently included knowledge management into one of its categories. However, the emphasis in research related to knowledge management is mostly on knowledge creation and dissemination, and not knowledge formalisation process. On the other hand, identifying the expert knowledge and experience as crucial for the output quality, especially in dynamic industries with high share of incomplete and unreliable information such as ship repair, this paper argues how important it is to have such knowledge formalised. The paper demonstrates by example of delivery time estimate how for that purpose the deep quality concept (DQC)—a novel knowledge-focused quality management framework, and machine learning methodology could be effectively used. In the concluding part of the paper, the accuracy of the obtained prediction models is analysed, and the chosen model is discussed. The research indicates that standardisation of problem domain notions and expertly designed databases with possible interface to machine learning algorithms need to be considered as an integral part of any quality management system in the future, in addition to conventional quality management concepts.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Quality management; Knowledge acquisition; Deep quality concept; Delivery time estimate; Dock works

1. Introduction

Ship repair is a complex, highly dynamic and stochastic process with high interdependencies. The process is also characterised with a high share of incomplete and unreliable information that is particularly expressed in some stages of the process. In such processes output quality is significantly influenced by the quality of assessments and decisions that cannot be ensured only by adherence to certain predefined procedures and instructions, on which, e.g. the standard ISO 9001 is based. In such systems expert knowledge and experience play a decisive role, and they are often of the nature that makes it practically impossible for them to be formalised with traditional methods. Also, because of so expressed technological complexity, and too many inter and intra dependent variables of influence, it is not easy (or even possible) to define efficient analytical models. Delivery time estimate in ship repair is one of typical examples of such processes. It includes the overall repair time estimate, as well as the estimate of duration of repair works in dock. The accuracy of these estimates significantly influences the quality of ship repair service. Also, it is critical for the business results of the shipyard. If the estimated times are too long, the shipyard will not be competitive. And if they are estimated too short, a production schedule may fail due to unrealistically estimated activity durations, which may result in final delivery time delay and penalties. Also, the quality of performed job might be influenced negatively given that delay often means doing things in hurry. This particularly goes for the overall repair time estimates.

On the other hand, developments in artificial intelligence provide powerful means for modelling expert knowledge. They also allow the automatic acquisition of such knowledge by means of machine learning or data mining techniques. Unfortunately,
the use of such techniques in quality management context is not of systematic but rather of an ad hoc manner. In industry this is caused by at least two main reasons. The first is Taylorian philosophy of manufacturing that still prevails in the current quality management models. Determinism of operations, predictable behaviour of the system, and a priori information that is reliable, complete and accurate, identified as the basic Taylorian presumptions of manufacturing systems by Peklenik [1], are still the main presumptions of the most well known quality management models (total quality management model (TQM), Malcolm Baldrige Criteria for Performance Excellence, EFQM Excellence Model, and standard ISO 9001). For example, fact-based management, i.e. the factual approach to decision making, are still listed among core quality concepts in the frame of all these models. Also, the use of information technology is not sufficiently systematic. One of the consequences of this is the lack of accurate and standardised bases of organisational as well as of technological data in some manufacturing organisations and domains. The second reason why the use of artificial intelligence techniques in quality management context is not of systematic but rather of an ad hoc manner is that knowledge of artificial intelligence techniques is typically modest. On the other hand, although the Malcolm Baldrige criteria included recently knowledge management into one of its categories, the emphasis in related research is mostly on learning, i.e. on knowledge creation and knowledge sharing, and not knowledge formalisation process (see, e.g. [2]). Also, distinction between the terms ‘knowledge’ and ‘information’ is not always clear in such research (see, e.g. [3]). A more detailed explanation of these limitations, as well as the DQC model—a new theoretical framework how to overcome these deficiencies are presented by Srdoc et al. [4]. In difference to other quality models that are typically concerned only with shallow knowledge, in this model particular attention is paid to standardisation of domain concepts, and domain deep knowledge. Integration of information systems, defined as systems whose purpose is to acquire and represent knowledge, and quality systems is also proposed in [5]. Dooley [6] also suggests that TQM paradigm based on predictability, control and linearity may be insufficient. How TQM approaches are inadequate because they do not address the uncertainties that impact significantly on results in some industries, is also described in [7]. On the other hand, a review of the use of intelligent systems in manufacturing can be found in, e.g. [8]. The review shows variety in the use of these techniques.

Concerning the use of machine learning algorithms for quality management in manufacturing, there are also several approaches. For example, Shigaki and Narazaki [9] demonstrated an approximate summarisation method of process data for acquiring knowledge to improve product quality based on the induction of decision trees, one of machine learning techniques. They also demonstrated a machine learning approach for a sintering process using a neural network [10]. Concerning the ship repair domain there has been no work reported on the use of artificial intelligence for quality management. Thus the use of machine learning algorithms has also not been reported. Instead, approaches based mainly on statistical techniques and ISO 9000 standards can be found (e.g. [11,12]). On the other hand, some work concerning manufacturing databases in the ship repair domain has been reported (e.g. [13]).

In this study, the approach as suggested within the DQC model is applied. The mechanisms investigated are: (1) systematic recording of data into expertly designed database, (2) standardisation of the data, and (3) transformation of the data into a knowledge base by means of machine learning. The data studied in the research and collected from a real ship repair yard are: (1) parameters defining repair activities that were described within each repair project (attribute values), and (2) related times estimated by the human expert (the target attribute). The data are limited to dock works. The reasons for that are: (1) dock works are technologically self contained subset of repair works, present in almost every ship repair project, (2) dock works often contain activities that influence the overall delivery time the most, such as anti-corrosive and steel works, and (3) since docks appertain to the most valuable and bottleneck resources of any shipyard the duration of these works is always important, and estimated separately. The goal of machine learning from these data was to construct comprehensible delivery time predictors, such as regression or model trees for computer-supported estimate, eliciting the hidden implicit knowledge from the data. Attribute selection and data refinement are done manually, based on the deep understanding of the learning problem and what the attributes actually mean. Given that in the inquiries-answering stage detailed technical data typically are not known, they are not included into this study.

2. Delivery time estimate in ship repair

The correct estimate of delivery time largely influences the quality and cost of the ship repair service. The delivery time depends generally on factors concerning: (1) the particular works that have to be done within the ship repair project, (2) the features of the shipyard, such as, e.g. physical capacities and capacity loading, facilities, technologies, tools and manpower available, experience and skill of people, (3) delivery time of materials and components, and (4) the situation on the market, such as the corresponding delivery times of competitors. Although within the operations planning, shipyards define works dependencies for almost each ship repair project, and there have been some efforts to improve the situation (see, e.g. [14]), a satisfying generic and computerised model that would support time estimates in the inquiries-answering stage when a large amount of data are still not known, and has capabilities to learn with time, has not been developed. Instead, mainly software packages that allow the user to construct a hierarchical model of the shipyard’s facilities and their workload based on the user knowledge and information are available (see, e.g. Chryssolouris et al. [15]). Also, a ship-owner typically contacts a number of shipyards in order to submit the initial work list. Therefore, shipyards accept a large number of enquiries that have to be evaluated. Consequently, the situations in which shipyards model ship repair works separately for each project, and on a relatively high level, are not rare.
Ship repair data inherit the probabilistic nature of ship repair process. Besides, non-standardised terminology and descriptions of repair works and often incomplete and missing data, introduce noise into data. During this research, it was revealed that missing data are particularly present in parts of inquiries that influence delivery time the most, e.g. in already mentioned parts dealing with anticorrosive, i.e. antifouling, and steel works. Moreover, the possibility of a wrong estimate increases with lack of experience. Different experts also can give different estimates for the same project. On the other hand, within a shipyard there are usually only one or two specialists sufficiently experienced to give reliable estimates. Consequently, there is a strongly expressed need to formalise such knowledge. The aims of the formalisation are: (1) to prevent the loss of the valuable knowledge and experience, (2) to make transparent estimate structures, and (3) to increase the reliability of the estimates, and thus the quality of the ship repair service.

3. Machine learning methods and algorithms used

Machine learning is the area of artificial intelligence concerned with the problem of building computer programs that automatically improve with experience [16]. Of the forms of learning, learning concepts from examples is the most common and best understood. Learning from examples is also called ‘inductive learning’. Inductive learning is the most researched kind of learning in artificial intelligence and this research has produced many solid results [17]. In attribute-based supervised inductive learning, a set of examples (instances) with known labels is given. An example is described by an outcome (label) and the values of a fixed collection of parameters (attributes). If the label is an element of a finite set of values, the problem is classification. If the label is a numeric quantity, the problem is regression. Monostori et al. [18] discuss several machine learning techniques based on learning from examples which are suitable for different kinds of knowledge synthesis in manufacturing planning domains.

What is learned by a machine learning scheme is a kind of ‘theory’ of the domain from which the examples are drawn, a theory that is predictive in that it is capable of generating new facts about the domain, such as the class of unseen instances [19]. Besides selecting a proper problem, and collecting and preprocessing the related data, successful machine learning also involves: (1) selecting a suitable machine learning scheme, and (2) suitable choice of parameter values. In both cases, the right choices depend on the data itself. In ship repair delivery time estimate problem the target attribute, i.e. the estimated duration of overall works in dock, is of numeric type. For this reason, two approaches to learning were possible (1) to discretise the target attribute values first, and then to use a classifying learning scheme or (2) to leave values of the target attribute as they are, and use learning schemes for numeric prediction.

Within this study the second approach was applied. It was assessed that discretisation of the works duration into classes is not of interest for practice—only too coarse duration estimates could be obtained compared with the range of estimated durations that appear in the practice (see Fig. 1). Generally, discretisation always invites the question whether it is justified. For this reason it always needs to be carefully studied. All these questions are avoided by using schemes for numeric prediction, such as, e.g. regression and model trees. For comparison reasons, instance based learning—another machine learning approach, and classical statistical approaches, have also been used in this study. An interesting example of numeric prediction problem dealing with ecological modelling can be found in [20].

3.1. Regression and model trees

Regression trees were introduced in the CART system of Breiman et al. [21]. Regression tree may be considered as a variant of decision trees, designed to approximate real-valued functions instead of being used for classification tasks. However, model trees did not appear until much more recently, being first described by Quinlan [22] and Karalik [23]. Quinlan...
generalised the regression trees in CART by using a linear regression model in the leaves to improve the accuracy of prediction. The advantage of using linear regression in the leaves of a regression tree is analysed in [23].

Regression trees are built through process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions, and then splitting it up further on each of the branches. Since the tree is grown from the training data set, when it has reached full structure it usually suffers from overfitting. It means that it tries to explain random elements of the training data that are not likely to be features of the larger population of data. This results in poor performance on real-life data. Therefore, it has to be pruned using the validation data set and the user-specified pruning factor. Larger values of the pruning factor result in smaller trees. If the pruning factor is specified as zero then pruning is simply finding the tree that performs best on validation data in terms of total terminal node variance. The same goes for model trees. According to Bratko [17] we can also view the pruning of trees as a method for learning from *noisy data*—when examples possibly contain errors. In this study for learning of model trees, as well as of regression trees, the system M5' was used. M5' is a reimplementation of Quinlan's model tree learner M5 [22], within the software package Weka [19]. M5 is one of the most well known programs for regression tree induction. Within Weka, it is available as m5.M5Prime class of algorithms. The test options of M5' were set to their default values, except for the pruning factor, which values we varied.

3.2. Instance-based learning

On the other hand, instance-based learning (IBL) algorithms Aha et al. [24], work directly from the examples themselves. They do not generalise. Instead, they memorise the set of training instances and on encountering a new instance they search the memory to find the most resembling training instance. IBL algorithms assume that similar instances have similar classifications: novel instances are classified according to the classification of their most similar neighbours. In fact, IBL algorithms are derived from the nearest neighbour (NN) pattern classifier [25]. Since IBL algorithms do not generalise, no model is generated. For instance-based learning we used algorithm IBk, also as implemented in Weka. The test options settings were left at their default values, except for k, which was set to 3—the target values of the three nearest neighbours of a new example were averaged to obtain a prediction of the target value for the new example. Linear regression, which is used for comparison purposes and whose only parameter controls how attributes to be included in the linear function are selected, is a standard statistical method, and was also used as implemented in Weka.

4. The dock works data model

In order to explore the possibilities of synthesising the time estimating ship repair knowledge, and define the dock works data model, the sample containing 221 examples of ship repair projects was collected. The collected sample corresponds to approximately 70% of inquiries that during one calendar year have usually been received and analysed in that typical medium-sized repair yard. Collected repair projects were described in a shortened non-standardised form on paper, although the shipyard was already ISO certificated. For this reason comprehensive and lasting process of analysis was necessary to get the final data model.

4.1. The data analysis

In the first stage, the sample was analysed regarding the presence of the dock works in examples, and the following was found: (1) the works in dock were present in 207 examples (i.e. in 93.67% of the sample), (2) the works in dock were not present in nine examples (i.e. in 4.07% of the sample), (3) in four examples descriptions were not available (i.e. in 1.81% of the sample), and (4) the work was cancelled in one example (i.e. in 0.45% of the sample). Of 207 examples in which dock works were mentioned, further analysis showed that 197 of them satisfied conditions for further analysis containing relatively well-defined works in dock, planned to be performed in that shipyard. For this reason, the rest of the study was concentrated on these examples that were analysed in detail. Other examples were whether projects planned to be done in another shipyard, or non-typical repair projects (e.g. the large underwater damages and conversions). The aim was to define patterns for transforming non-standardised descriptions on paper into standardised computer database. The focus was put on four groups that the domain expert assessed as the most important for the dock works duration estimate, i.e. on: (1) renewal of steel on shell plating, (2) shell plating treatment, (3) works on propeller and propeller shaft, and (4) works on rudder. The analysis revealed that the most variations in the work descriptions were present in the rudder and propeller/propeller shaft frame of repair works, respectively. On the contrary, the shell plating treatment issues were always repeated under the same pattern, as well as the major part of issues dealing with renewal of steel on shell plating. Therefore, rudder and propeller/propeller shaft works were additionally analysed. The range and distribution of estimated durations for dock works are given in Fig. 1.

4.2. The data model

At the end of the data analysis, the work structures obtained for rudder and propeller/propeller shaft were merged together with the structures recognised previously within the shell plating treatment, and renewal of steel on shell plating. That way the data model consisting of 37 the most important dock works for duration estimate was obtained. By adding duration estimated by the expert in days as the target attribute, and the column for the example ID, the pattern for dock works data entry into database was finally defined (see Table 1). That way, the basis for interface to machine learning algorithms was also obtained. In the table, attribute A2 defines the total
4.3. Dock works data preprocessing

Within the data preprocessing process, the following was performed: (1) data were carefully checked once again and cleansed, (2) missing values in the sample were analysed, and (3) the data space reduction possibilities were explored. Due to their possible influence on the efficiency of learning, in this study particular attention was paid to missing values. Within the explored dataset, the values are mainly missing because in the stage of inquiries answering, a significant part of values are typically unknown. After that, the data dimensionality reduction was performed.

4.3.1. Missing values in the sample

Missing values are first analysed by instances. Then they are analysed by particular attributes. As discovered, about 50% of instances in the sample contained missing values. The average of missing values by instance was 3, and the greatest number of unknown values found within one instance was 17. It is to be noted that the expert estimate was given even for this instance. Concerning the attributes, the greatest number of missing, i.e. unknown values, on average about 18.24%, was found in the group of works dealing with the shell plating treatment (attributes from A1 to A15) (see Fig. 2). The next group by the number of unknown values was found in works dealing with renewal of steel on shell plating (attributes from A34 to A37). But within that group the percentage of unknown values was far less (on average about 6.98%).

On the other hand, attributes from A16 to A24 (works on rudder), and from A25 to A33 (works on propeller/propeller shaft), contained very few unknown values, on average about 0.39% for rudder and 0.68% for propeller. These results confirmed that shell plating treatment and renewal of steel on shell plating (attributes from A34 to A37). But within that group the percentage of unknown values was far less (on average about 6.98%).

4.3.2. Reduction of the data dimensionality

Because of the negative effect of irrelevant attributes on most machine learning schemes [19], in this study various possibilities for the data dimensionality reduction were explored. First, the possibility of discarding irrelevant, redundant, as well as useless attributes was examined, and then grouping of the attributes, as another method for reduction of the data space, was also performed. The attributes identified as irrelevant are those whose influence on the overall work duration in dock the domain expert assessed as not significant. Redundant attributes are those that contain information that is already included in information contained in some other attribute(s). On the other hand, attributes having the same values for all, or almost all instances in the sample, are marked as useless for the available dataset because it was obvious that such attributes are of no or very little influence on the learning

<table>
<thead>
<tr>
<th>Att. ID</th>
<th>Work group</th>
<th>Work/attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Shell plating treatment</td>
<td>HP washing (m²)</td>
</tr>
<tr>
<td>A2</td>
<td>Shell plating treatment</td>
<td>Underwater part (m²)</td>
</tr>
<tr>
<td>A3</td>
<td>Scrapping, etc. (m²)</td>
<td></td>
</tr>
<tr>
<td>A4</td>
<td>Surface preparation Sa 1 (m²)</td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td>Surface preparation Sa 2/ Sa 1.5 (m²)</td>
<td></td>
</tr>
<tr>
<td>A6</td>
<td>Surface preparation Sa 2.5 (m²)</td>
<td></td>
</tr>
<tr>
<td>A7</td>
<td>Coating T/U (no. of coatings)</td>
<td></td>
</tr>
<tr>
<td>A8</td>
<td>Coating F/C (no. of coatings)</td>
<td></td>
</tr>
<tr>
<td>A9</td>
<td>Freeboard (m²)</td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>Scrapping, etc. (m²)</td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>Surface preparation Sa 1 (m²)</td>
<td></td>
</tr>
<tr>
<td>A12</td>
<td>Surface preparation Sa 2/ Sa 1.5 (m²)</td>
<td></td>
</tr>
<tr>
<td>A13</td>
<td>Surface preparation Sa 2.5 (m²)</td>
<td></td>
</tr>
<tr>
<td>A14</td>
<td>Coating T/U (no. of coatings)</td>
<td></td>
</tr>
<tr>
<td>A15</td>
<td>Coating F/C (no. of coatings)</td>
<td></td>
</tr>
<tr>
<td>A16</td>
<td>Rudder repair works</td>
<td>Clearances (false/true)</td>
</tr>
<tr>
<td>A17</td>
<td>Registry inspection (false/true)</td>
<td></td>
</tr>
<tr>
<td>A18</td>
<td>Dismounting (false/true)</td>
<td></td>
</tr>
<tr>
<td>A19</td>
<td>Renewal of pintle glands (false/true)</td>
<td></td>
</tr>
<tr>
<td>A20</td>
<td>Exchange of shaft bearing liner (false/true)</td>
<td></td>
</tr>
<tr>
<td>A21</td>
<td>Machining of rudder stock (false/true)</td>
<td></td>
</tr>
<tr>
<td>A22</td>
<td>Straightening of rudder (false/true)</td>
<td></td>
</tr>
<tr>
<td>A23</td>
<td>Rudder stock into workshop (false/true)</td>
<td></td>
</tr>
<tr>
<td>A24</td>
<td>Alignment on stock/blade seating (false/true)</td>
<td></td>
</tr>
<tr>
<td>A25</td>
<td>Propeller/propeller shaft repair works</td>
<td>Type of propeller (1 = classical type, 2 = controllable pitch prop.)</td>
</tr>
<tr>
<td>A26</td>
<td>No. of propellers where work is performed (no. of prop.)</td>
<td></td>
</tr>
<tr>
<td>A27</td>
<td>Clearances (false/true)</td>
<td></td>
</tr>
<tr>
<td>A28</td>
<td>Registry inspection (false/true)</td>
<td></td>
</tr>
<tr>
<td>A29</td>
<td>Dismounting (0 = there is no dism., 1 = dism. of prop. blades, 2 = dism. of prop., 3 = dism. of prop. and prop. shaft)</td>
<td></td>
</tr>
<tr>
<td>A30</td>
<td>Renewal of gaskets (0 = there is no work on gaskets, 1 = vulcanisation, 2 = renewal of gaskets with disassembly of stuffing box)</td>
<td></td>
</tr>
<tr>
<td>A31</td>
<td>Repair of cavities/static balancing (false/true)</td>
<td></td>
</tr>
<tr>
<td>A32</td>
<td>Inspection of shaft on machine (0 = there is no work on shaft, 1 = insp. only, 2 = insp. and adjusting of shaft)</td>
<td></td>
</tr>
<tr>
<td>A33</td>
<td>Polishing of propeller (false/true)</td>
<td></td>
</tr>
<tr>
<td>A34</td>
<td>Renewal of steel on shell plating</td>
<td>Ultrasound measuring of shell plating thickness (no. of points)</td>
</tr>
<tr>
<td>A35</td>
<td>Total quantity of steel (kg)</td>
<td></td>
</tr>
<tr>
<td>A36</td>
<td>Positions (0 = there are no positions, 1 = concentrated, 2 = scattered)</td>
<td></td>
</tr>
<tr>
<td>A37</td>
<td>Underwater seams (m)</td>
<td></td>
</tr>
<tr>
<td>A38</td>
<td>TIME</td>
<td>The estimated dock works duration (days)</td>
</tr>
<tr>
<td>A39</td>
<td>ID</td>
<td>Instance ID (system number)</td>
</tr>
</tbody>
</table>

HP: high pressure; T/U, touch up; F/C, full coating; Att., attribute; Prop., propeller; Dism., dismounting; Insp., inspection.

surface of the underwater part of the ship on which works according to inquiries are to be carried out, and A9 the total surface of the freeboard on which works are also to be carried out.
model that is to be synthesised. Of course, if the situation with
the sample changes, inclusion of such attributes into the
learning process has to be reconsidered. A summary of this
analysis is given in Table 2.

Grouping of the attributes is a variant of constructing new
attributes, and depends strongly on the semantics of the data-
mining problem [19]. In the explored learning problem,
attributes that are assessed to be sensibly grouped are those of
the first group of repair works (shell plating treatment). The
approach used was to split the group (excluding the HP
washing, i.e. work marked as A1) into two symmetric parts
(underwater part and freeboard), and then sum up the
corresponding values (e.g. values for A2 and A9, A3 and
A10, etc.). In this research, grouping of attributes was used to
improve not only the accuracy, but also readability and
comprehensibility of the generated models. For the reason of
comprehensibility, the whole group of shell plating treatment
works was retained, although some of these works were, during
the previous analysis, assessed as discorable.

5. Learning and analysis of the generated models

Since there is no definite way to choose the most appropriate
learning methods in the specific domain [26], in this research
several learning tests were carried out. Because of so many
possibilities discovered for the reduction of the data
dimensionality, tests were performed on different datasets,
using different prediction techniques outlined in Section 3. In
initial experiments, all attributes were included. After that,
attributes estimated as most likely to be irrelevant, redundant or
useless for the learning problem were discarded, in various
combinations, exploring the accuracy of the obtained pre-
dictors. In the second stage, as described in the previous
subsection, the grouping of attributes was also performed.

After the first experiments, it became obvious that model
trees and simple linear regression achieved the best results.
Five measures of performance are examined: (1) the multiple
correlation coefficient (CC), (2) the mean absolute error
(MAE), (3) the root mean squared error (RMSE), (4) the
relative absolute error (RAE), and (5) the root relative squared
error (RRSE). Results are compared varying (1) the learning
method, (2) the dataset, i.e. the attributes used, and (3) the
pruning factor. Good performance, when measured with the
correlation coefficient, is indicated with the large values. For
all other methods good performance is indicated by small
values. Although the 10-fold cross-validation is generally
applied, where appropriate, i.e. for regression and model trees
and linear regression, to explore the performance bounds,
models generated on the entire dataset of 184 instances are
also analysed. Finally, for predicting the durations of repair
works in dock the model tree induced by M5’ scheme from
dataset containing 21 attributes, and using the default pruning
factor 2.0 was chosen. On the other hand, although
corresponding linear regression model achieved similar
accuracy, it was not chosen because it was not as readable
as the model tree. To predict the value of the target attribute, it
used all variables, while the model trees (as well as regression
trees) focus only on the variables that dock works duration
influence the most.

5.1. The model tree chosen

The attributes used for learning were: A1, A2+9, A3+10,
A4+11, A5+12, A6+13, A7+14, A8+15, A16, A18, A19, A20,
A21, A25, A26, A27, A28, A29, A30, A31, and A35. The
performance accuracy of this predicting model on unseen cases
as estimated by 10-fold cross validation is among the highest
achieved (CC = 0.8439, MAE = 1.9898, RMSE = 3.0417,
RAE = 57.7887%, and RRSE = 54.3618%). Also, it covers well almost the whole range in which dock works durations occurred. The chosen model tree induced by M5’ on all the data of 184 instances is given in Table 3. The tree can be used to predict the duration of repair works in dock from the works anticipated in the ship-owner’s enquiry. Depending on the concrete values of attributes that appear in the model, five different linear models are used. These models represent linear combination of variables. The duration of the works in dock increases as the values of each of these variables increase, in the case when coefficients in the model are positive. If they are negative, the duration of the dock works decreases as the values of the variables increase.

Of these five linear models, only LM3 is not quite consistent with the domain knowledge, given that it predicts that duration decreases as the value of A1 increases, which is not quite logical. Also not logical in this model is positive influence of A30 = 0 variable on the estimated time. Concerning the linear model LM2, although at first sight it appears that the duration of dock works decreases as the variable A30 value increases it was not the case. The variable A30 influences the model only if it has the value zero. In that case, the estimated time is for 2.34 days shorter. That is logical because in that case the renewal of gaskets on propeller shaft is not a part of the project.

5.2. The comparison with the expert estimates

A comparison of values obtained by the chosen model tree, and those estimated by the expert, is given in Fig. 3. The comparison showed that 89.7% of the results lay within the acceptable bounds of ±2–3 days of deviation or less, assessed by the domain expert as acceptable, while only 10.3% lay outside. The average deviation of the model compared to the expert’s estimate is 1.83 days. On the other hand, it is known that for small datasets the performance of the model could deviate sharply.

The greatest deviations are obtained in five examples, where total quantity of steel was greater than 12,500 kg (A35), and HP washing surface (A1) greater than 4980 m². Although it seemed that in such cases the predictions of the model have to be taken with reserve, further analysis revealed that it is not the case. On the contrary, for these five cases it was found that expert estimates are not consistent with expert estimates given for similar projects. In other words, it was revealed that model estimates for these five examples appeared to be closer to the expert estimates for the similar cases, than values recorded as expert estimates for these instances. The conclusion confirmed

Table 3
A model tree for predicting the duration of repair works in dock

<table>
<thead>
<tr>
<th>Pruned training model tree:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A35 ≤ 11,300: LM1</td>
<td></td>
</tr>
<tr>
<td>A35 &gt; 11,300:</td>
<td></td>
</tr>
<tr>
<td>A35 ≤ 12,200: LM2</td>
<td></td>
</tr>
<tr>
<td>A35 &gt; 12,200:</td>
<td></td>
</tr>
<tr>
<td>A35 ≤ 12,500: LM3</td>
<td></td>
</tr>
<tr>
<td>A35 &gt; 12,500:</td>
<td></td>
</tr>
<tr>
<td>A1 ≤ 4,980: LM4</td>
<td></td>
</tr>
<tr>
<td>A1 &gt; 4,980: LM5</td>
<td></td>
</tr>
</tbody>
</table>

Model at the leaves:
Unsmoothed (simple):
LM1: TIME = 7.02 + 5.36E−4A5+12 + 6.46E−4A6+13 + 0.228A8+15 + 2.14E−4A35
LM2: TIME = 10.5 − 2.34A30 = 0
LM3: TIME = 12.3 − 7.78E−4A1 + 0.00443A3+10 + 11.5A30 = 0
LM4: TIME = 10.6 + 1.14E−4A35
LM5: TIME = 28

Number of rules: 5
Time taken to build model: 0.55 s

LM, linear model; A35, total quantity of steel (kg); A1, HP washing (m²); A5+12, surface preparation Sa 2/Sa 1.5, underwater part and freeboard (m²); A6+13, surface preparation Sa 2.5, underwater part and freeboard (m²); A8+15, coating F/C, underwater part and freeboard (no. of coatings); A30, renewal of gaskets; A3+10, scrapping.

Fig. 3. The comparison of the expert’s estimates and values forecasted by the chosen model with the five cases containing inconsistencies in expert estimates included.
by the expert was that true estimates for these five cases were not recorded, thus not included into the dataset. In order to get the job, shipyards sometimes give even unrealistic offers to the ship-owners. Also, they could estimate delivery times having in mind the need for additional manpower. For these reasons, for real-life applications it is always important to have notes considering the presumptions of the estimate (e.g. normal conditions or additional manpower anticipated, etc.), as well as data on the estimate itself (e.g. initials of the expert, the date of the estimate, etc.), included into records.

All this demonstrates how additional justification of expert opinion could be critical. It also demonstrates how a relatively small sample of historical cases could contribute significantly to this. In the case of employing such a basic concept as systematic recording of relevant data into expertly designed databases (preferably within quality systems), the higher estimating accuracy of machine learning predictors could be achieved, and be of great benefit to the profitability of ship repair service. For these reasons, the mechanisms as suggested within the DQC approach and stated in the introductory part of the paper are so important to be recognised and applied in quality systems. The feedback, on the other hand, although also important, for this particular problem has to be applied with caution. Differences between estimated and realised times could be caused by various reasons, for example, bad weather conditions and/or additional works of different complexity, and do not automatically indicate inaccuracy in expert estimate. Because of that, it is important again to have a set of possible reasons carefully classified and codified for standardised and consistent recording into database.

6. Discussion and conclusions

Increasingly, differences in a firm’s performance are attributed to tacit knowledge (e.g. [27,28]). According to Simon [29], the reason why experts on a given subject can solve a problem more readily than novices is that the experts have in mind a pattern born on experience, which they can overlay on a particular problem and use to quickly detect a solution. On the other hand, the uncertainty associated with humans increases the need for knowledge formalisation. Rarity of real experts increases such need too. For example, in the examined shipyard after one expert was retired, only one capable of making reliable estimates on the problem explored remained. Consequently, the reliability of quality as well as of business systems, among others, lies in formalising as many as possible of patterns born on experience.

In this paper an attempt to formalise one such pattern from the ship repair database is presented. The works required on the enquiries received by the shipyard have been analysed, and the related ship repair data model was created. The methodology of learning from examples was employed, and several models were induced eliciting and representing the implicit knowledge from the database. Finally, the model tree directly usable for estimating the possible repair time was chosen. The experiments confirmed that total quantity of steel within the renewal of steel on shell plating, and shell plating treatment are the most important attributes, appearing in all generated models. On the other hand, HP washing surface revealed as a more informative parameter than it is usually thought. The greatest improvements in results accuracy on unseen cases evaluated by 10-fold cross-validation were obtained using different learning algorithms. On the other hand, varying the datasets influenced the achieved performances the least. That clearly confirmed Witten and Frank’s [19] thesis that in many practical situations perhaps the overwhelming majority of attributes are irrelevant or redundant. However, their statement about negative effect of irrelevant attributes on most machine learning schemes was not confirmed in this application—the results remain stable regardless of the discarded attributes, particularly when accuracy was evaluated on unseen cases. On the training data, these results were more dependent on the dataset used.

The experiments also demonstrated that machine learning methods can offer an advantage over approaches based on linear and network analysis that are usually employed for the time estimate problem in shipyards, such as, e.g. Gant charts or activity networks, because they do not need any prior assumptions or knowledge about the relationships between the variables. Also, of 21 attributes used for learning, the machine-learning algorithm identified model with as few as seven attributes in the chosen tree. The comparison of the induced model predictions with the expert estimates, and analysis of the instances with the greatest deviation showed that such inconsistencies do not always have to indicate bad performance of the induced learner. They could indicate inconsistency in human reasoning, or in data available. In this study it is demonstrated how data mining could reveal such inconsistencies. For this reason, the checking of the results obtained using data mining techniques has to be always addressed in two directions—i.e. in (1) direction of checking accuracy of the learner, and (2) in checking consistency of predictions used for learning. Once the input data model was created and data standardised and cleansed, the estimate structures were obtained relatively quickly, although the process was not trivial, and required a high share of domain modelling knowledge and reasoning. In case of additional records on new projects, such models can be improved, of course, when estimating presumptions are not significantly changed.

On the other hand, at least 10–15 years of experience are needed for graduated engineer or talented technician to be able to give reliable estimates. This includes supervised work with the experienced mentor. Inaccuracies of predictions of insufficiently experienced human predictors could exceed 30, and sometimes even 50% or higher. For example, in ship repair community it is well known the case for which difficulties in accurate delivery time estimate and tons of steel needed caused serious problems to one of renowned ship repair yards in Croatia. The fact that the shipyard has already had the standard ISO 9001 certificate did not prevent the final disaster. This is also consistent with Massow and Siksnes-Pedersen [30] statement that shipbuilding as an extremely complex process requires special methods and tools for order processing.
How overconfidence in forecasts based on expert judgement can be risky is also discussed in Armstrong [31]. In the handbook he edited, an overview of principles of forecasting, as well as methods of reducing the impact of inconsistency and bias in judgemental forecasting are also described. Many of the concepts discussed, such as careful identification of the most important causal forces (attributes in this case study), accurate records, use of models—especially computerised models, are also included in the DQC approach, and employed in this study.

Of course, all these do not mean that conventional quality management principles have to be put apart. As shown in [4], it only means that all these concepts, as well as other concepts relevant for quality management that are still to be identified, need to be re-validated and put together, leading to a new, more sophisticated and complete quality management philosophy. The quality management community, as well as quality standardisation and award bodies need to recognise that need. The criteria for knowledge formalisation are stated in [4]. As it is suggested, besides now already established functions like quality managers and engineers, in designing, development and maintenance of quality systems knowledge engineers should be anticipated as well.

Acknowledgments

The authors wish to thank the Faculty of Organization and Informatics Varazdin, University of Zagreb, for providing partial funding to support this research. They also owe thanks to the shipyard, and special thanks to Mr. Drago Brzac, the domain expert for his contribution to this research. Thanks also to Mr. Christian Massow from Logimatic, Germany, for his useful suggestions on Section 2, as well as to two anonymous referees who helped us to clarify the positioning of the presented research in relation to the current quality management approaches.

References

[3] S. Al-Hawamdeh, Knowledge management: re-thinking information management and facing the challenge of managing tacit knowledge, Informa-
[22] J.R. Quinlan, Learning with continuous classes, in: Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence, World Scientif-

Alira Srdocˇ graduated in Naval Architecture at Technical Faculty, University of Rijeka, Croatia, where she also received her MSc degree. Since now, she has worked in two shipyards, ‘Viktor Lenac’ and ‘3.MAJ’, both in Rijeka, Croatia, in the area of planning, as well as information systems design and implementation. In collaboration with the Artificial Intelligence Laboratory, Jozef Stefan Institute, Ljubljana, Slovenia, she has also developed some industrial knowledge-based systems. Currently she is...
project manager in ‘3.MAJ’, as well as the final year PhD student in Quality Systems at the Department of Control and Manufacturing Systems at the Faculty of Mechanical Engineering, University of Ljubljana, Slovenia. Her research interest is focused on issues related to management, particularly quality and knowledge management, and expert knowledge modelling. She has published several research and professional papers, the most important of which is A quality management model based on the ‘deep quality concept’, published in International Journal of Quality and Reliability Management in 2005.

Ivan Bratko is professor of computer science at Faculty of Computer and Information Science of Ljubljana University, Slovenia. He heads the Artificial Intelligence Laboratory and is also associated with J. Stefan Institute, Slovenia. He has worked as visiting professor at various universities worldwide. His research interests include machine learning, qualitative modelling, computer game playing, and applications of artificial intelligence in biomedicine, ecological modelling and systems control. He has published over 200 research papers, as well as several books, the best known being Prolog Programming for Artificial Intelligence (third ed., Addison-Wesley 2001).

Alojzij Sluga is an associate professor of manufacturing engineering at Faculty of Mechanical Engineering, University of Ljubljana, Slovenia. He heads the Laboratory for Manufacturing Cybernetics and Experimentation. He has worked with different industries in areas of manufacturing technology and computer integrated manufacturing. His current research interest includes enterprise modeling, networked organisations, quality systems and technology. He has published over 100 research papers. He has been scientific responsible for the University of Ljubljana of the EUREKA project TRUST ‘Use of Regional Potentials with Respect to Problem-Solving Processes in Production’, and of the Network of Excellence VRLKCiP ‘Virtual Research Lab for a Knowledge Community in Production’ funded by the European Community. He is a member of SATENA (Slovenian Academic Society for Technology and Natural Science), and associate member of the CIRP (International Academy for Production Engineering).