THE EVALUATION OF EXPERT SYSTEMS FOR CONSTRUCTION PLANNING

Carlos T. Formoso, M.Sc.   Prof. Peter S. Brandon, M.Sc.

Quantity and Building Surveying Division
Department of Civil Engineering
University of Salford
Salford M5 4WT

ABSTRACT

A micro-computer based expert system in the field of construction planning of house building has been developed. This system encapsulates some of the knowledge experts in this field use for programming the production stage of low rise, traditional housing projects, and can be used by construction planners as a decision making supporting tool. The evaluation process devised for this system has focused on the validation of the planners' expertise model.

This paper reviews some of the fundamental concepts related to the problem of evaluating expert systems. Some of the techniques that have been used in the validation of such systems are described, and the main conditions and constraints concerning the application of knowledge engineering in the field of construction planning are discussed. A prescriptive approach for the validation of the developed system is proposed. This approach includes both qualitative validation techniques and the development of a method for the continual validation of the system during its life cycle.

KEYWORDS: Expert Systems, Construction Planning, House Building

1 INTRODUCTION

Although a lot of effort has been devoted to the tasks of designing and constructing expert systems, very little has been reported on the evaluation of such systems. The techniques that have been used for evaluating their performance are generally "ad hoc", informal and of dubious value (O'KEEFE et al., 1987).

Evaluation can be described as the process of accessing an expert system's overall quality. In the literature on conventional software quality measurement, the concept of software quality is manifested in a variety of characteristics, each characteristic being generally broken down into a larger set of attributes or criteria (WATTS, 1987). In this context, quality is to be understood as the degree of compliance (or non compliance) of the system with specified requirements, rather than a degree of excellence. Some of the most commonly mentioned quality characteristics are: usability, security (or integrity), efficiency, correctness, reliability, maintainability, testability, flexibility, re-usability, portability and inter-operability.

On the other hand, the literature in the field of expert system evaluation has mostly approached the problem of validating and verifying such systems. Validity is the degree at which the outcomes of the resulting expert system resembles the outcomes of the human expertise modelled in the knowledge base. Verification is a more specific concept, which applies to testing that the knowledge base is logically sound and complete (PREECE, 1989).

The primary objective of this research project is to investigate the applicability of expert system technology in the field of construction planning of house building rather than to develop a high performance commercial package. Therefore, the evaluation process concerning the expert system that has been developed will concentrate on the validity of the model of expertise produced.

2 DIFFICULTIES IN EVALUATING AN EXPERT SYSTEM

2.1 Evaluating computer software

The first basic difficulty in evaluating any computer software arises from the fact that testing a computer-code can never be absolute, except for very simple programs. The number of alternative combinations of inputs is normally so large that most systems cannot have their correctness scientifically tested.

A number of techniques have been proposed for measuring each of the quality characteristics mentioned above. Those techniques are either based on subjective rating methods performed by experts or on some formulas which consider aspects of computer programmes that are possible to quantify, such as number of lines of code, average length of sentences and words displayed to the user, etc. The main limitation of such techniques is that most of them have not been actually proved to be correlated to the characteristics they are supposed to measure (WATTS, 1987). Moreover, provided they are relative measures, their usefulness is restricted to comparing a number of alternative systems or comparing a system to an acknowledged gold standard.

In the particular case of expert systems, measuring the software quality is even more difficult because knowledge bases are usually built on the top of another software, an expert system shell or a knowledge engineering environment. Any attempt of measuring the quality of an expert system would have to consider the combination of the knowledge base with the programming tool.

An additional difficulty to the multi-dimensional approach to software quality is how to amalgamate the relevant characteristics in order to evaluate the system's overall quality. WATTS (1987) introduces six methods that can be used for getting such an overall measure, in which the relative significance of each characteristic is subjectively established. It may be difficult to get an agreement about the relative importance of each characteristic amongst the various people involved in the development of an expert system (developers, experts, sponsors, users, etc.), since they usually have different interests. For instance, while users generally are very concerned with the man-machine interface, developers probably give priority to model correctness.

2.2 Validating expert systems

Validation can range from formal to informal. A thoroughly formal validation consists of a process in which validation methods are specified in advance in terms of input domain specification, level of acceptance, when to use them, and relevant statistical techniques (if appropriate). It is widely accepted that validation of expert systems has been too informal (O'KEEFE et al., 1987). The major problems usually found in validating expert systems are:

(i) What to validate: validation may include final results, partial results, and the reasoning of the system. Validating the reasoning process must be carried out even if the system is apparently giving accurate results, otherwise the system may present problems of robustness, particularly if the knowledge base needs to be periodically expanded.

(ii) What to validate against: a gold standard is needed, which the results and the reasoning of the system can be compared to. This gold standard can be human expert performance or data from the real world. In some fields, the only gold standard available is the human expert performance, since the cost of obtaining data from the real world is very high (WEISS & KULIKOWSKI, 1984). In such cases it may be difficult to know how the system performs in relation to the real world, since there might not be an adequate measure of the quality of human expertise. However, it is not fair to expect an expert system to have a very high performance if it encapsulates the knowledge of human experts who may occasionally give wrong solutions.

(iii) What to validate with: the number of test cases used in the validation process obviously affects the level of confidence that a system can reach. The main issue is not the number of test cases, but their coverage, i.e. how well they reflect the input domain (O'KEEFE et al., 1987). If enough historic cases are not available for the validation, it may be possible to use a number of hypothetical test cases created by experts. The difficulties in using hypothetical cases is that they might not represent a well-stratified sample of possible cases, and the experts are unlikely to spend as much time and effort on them as on real problems.

(iv) When to validate: the validation process should be continual, beginning with the system design, extending in an informal way through the early stages of development, and becoming increasingly formal as the system begins to achieve a real-world implementation.

(v) How to control the cost of validation: validation can be time consuming and expensive. It is difficult to establish exactly when to stop it. The value of validation depends on the value of the system to its users and on the risk involved in using a poor validated system.

(vi) How to control bias: there are two main types of bias. The first one relates to the experts involved in validation who might have bias against (or for) results produced by computer programmes. Such bias can be controlled by using blinded validation, in which the experts are not able to distinguish which results were produced by the computer and which ones were produced by human experts. The other kind of bias relates to the difficulties that the development team (developers and experts) might have in validating their own system, once they are very much involved in the project. This problem can be minimized by having an independent team of experts for the validation stage.

(vii) How to cope with complex results: even considering that an adequate gold standard is available, validating an expert system might not be easy if its results cannot be easily classified as correct or incorrect. If a system produces a piece of text from the concatenation of several statements as a conclusion, it may be difficult to break that text in a number of firm endpoints that can be compared to a gold standard (WEISS & KULIKOWSKI, 1984).

(viii) How to cope with disagreements between experts: one of the problems that knowledge engineers have been faced during the development of expert systems is the fact that different experts may disagree about a particular piece of knowledge. It is not always clear if the correct solution is one that a human expert would give, one that a group of experts agree upon or one that represents an ideal solution (GASCHNIG et al., 1983).

Even the results of a rigorous validation must be considered very cautiously. GASCHNIG et al.(1983) emphasize that the process of acceptance testing cannot be considered complete until the expert system is actually used routinely, in the capacity for which it was designed.

3 CURRENT APPROACHES FOR THE VALIDATION OF EXPERT SYSTEMS

Validation methods can be either qualitative or quantitative. Quantitative validation employs subjective comparisons of performance, while quantitative validation employs statistical techniques to compare expert system performance to a gold standard (O'KEEFE et al., 1987). Qualitative validation does not mean informal validation. It is possible to develop a highly informal qualitative validation. O'KEEFE et al. (1987) list seven common qualitative approaches to validation:

(i) Face validation: the system performance is subjectively compared to the human experts' by the developers, users, or people knowledgeable about the application domain. The results obtained from an expert system are compared to a prescribed acceptable performance range, for a given set of test cases.

(ii) Predictive validation: the system is used in some historical cases and its results are compared to corresponding results - either known results or those obtained from human experts.

(iii) Turing tests: similar to the previous ones, except by the fact that both human expert and system performance are compared without knowing the subject performer's identity.

(iv) Field tests: a prototypical expert system is placed in the field and performance errors are corrected as they occur. It is only possible in non critical applications.

(v) Sensitivity analysis: the expert system's inputs are changed over some range of interest and the effect upon system performance is observed. It is especially useful when few or no test cases are available.

(vi) Visual interaction: a visual animation of the expert system task which allows human experts to interact, altering parameters as desired, is provided. In essence, it is simply an environment for other validation methods.

(vii) Sub-system validation: the system is decomposed into sub-systems, which are individually validated using some of the methods above. Such method usually makes validation easier, since sub-systems are less complex and more manageable than the whole system, making error detection less time-consuming. Also, sub-system validation can be carried out along the several stages of development, before the whole system is completed. Its main limitation is that a successful validation of sub-systems does not necessarily imply that the whole system is validated.

Very few systems have been submitted to a complete formal validation so far. The most common approach has been to show a system to experts and to ask them if they agree with the conclusions for a number of test cases. Very little has been reported on the validation of expert systems in the construction field, probably because only few of them have reached an operational stage.

4 VALIDATING AN EXPERT SYSTEM FOR CONSTRUCTION PLANNING

4.1 Description of the system

HOUSE PLANNER is a knowledge based framework that encapsulates the expertise construction planners use for the strategic planning of house building projects. This framework was primarily designed to be used as a decision making support system for planners, during the early stages of the construction process, such as feasibility, design and tendering. The system proposes a construction programme, based on information such as general aspects of the project, site conditions, approximate design dimensions, and the specification of some key components. At several points of the session the user is asked to confirm or overwrite some of the key parameters of the programme. HOUSE PLANNER is able to cope with missing information, and its main outputs are a general programme for the production stage, a plan of the main milestones related to each dwelling, and the schedules of the most critical resources. A panel of five experts, from three different construction companies, provided the expertise for the system. See FORMOSO & BRANDON (1990) for a more detailed description of HOUSE PLANNER.

The practice of incremental development was adopted due to the exploratory character of the study. It means that the process of establishing requirements, designing, building and testing the system was carried out in several stages, rather than attempting to solve the whole problem at once. The experience obtained in implementing every part was systematically used for planning subsequent stages of work. The development of a feasibility prototype early in the study, based on the expertise of only one expert, played a key role in getting the interest of more contributors.

4.2 Problems in the validation process

Construction planning is a very complex task that involves a large number of different activities, being usually subjected to a number of conflicting constraints concerned with time, cost, space or availability of resources. One of the main difficulties in choosing a single gold standard for testing the performance of an expert system that generates construction programmes is the fact that a unique solution does not exist for most real projects. Construction planning involves a great deal of uncertainty, caused by several factors such as the lack of knowledge about the variability of gangs' performance, unexpected interference from external sources (inclement weather, strikes, shortage of resources, etc.), and lack of control over sub-contracted gangs. Additionally, planning at the early stages of the construction process usually has to be based on incomplete information about the location of the job, the site conditions and the design. These facts turn the optimisation approach, used in other engineering fields, largely ineffective in the construction practice. Generally, construction planners search for a feasible arrangement of actions for the production stage of a project, rather than an optimum one.

Considering that there is an infinite number of feasible arrangement of actions for any real construction project, it is very unlikely that the construction plans generated by different human experts for the same project can be identical. For the same reason, it is not reasonable to expect that a programme generated by an expert system that encapsulates the expertise from a number of practitioners should be identical to a chosen gold standard.

The use of quantitative techniques on the validation process in this particular study was not possible due to the lack of representative test cases. The number of historical cases available was not very large - only fifteen were considered usable. These were restricted to human generated programmes, since none of the companies has kept systematically in their files complete information about the way production actually happens on site. The use of hypothetical test cases was also impractical because the amount of time experts could dispense to the validation process was not enough for creating and analysing in detail a large number of cases.

4.3 Approach adopted in this study

The validation process involved in the development of HOUSE PLANNER was divided in three stages: early validation, acceptance validation and continual validation. Early validation consists of a number of testing procedures adopted in the early stages of development. As a result of the incremental approach adopted in this study, early validation was carried out in a very informal way, using the following procedures:

(i) Static validation of rules and frames: the content of some rules and frames was checked by human experts. Unfortunately, the small number of experts involved in the knowledge acquisition process did not allow an independent validation to be carried out during the early stages of development.

(ii) Sub-system validation: all sub-systems or parts of the system, that were subsequently built, had their correctness checked using structural test cases. Structural test cases can be defined as sets of input data that test the logic of specific segments of code (HETZEL, 1984).

Acceptance validation consists of checking whether the system has reached a reasonable level of quality at the end of the development stage. It has been carried out since the first full version of the system was completed. A number of qualitative validation methods have been chosen, rather than a single one, in order to provide an analysis as formal, as independent, and as exhaustive as possible. They can be summarized as follows:

(i) Predictive validation: information about fifteen past real projects have been used for analysing the correctness of the programmes proposed by the system. They were compared to the ones generated by human experts. The main discrepancies were identified and taken to further discussion with the panel of experts. Seventy seven different aspects of a construction programme were selected for these comparisons, although not all of them are relevant for every project.

(ii) Independent criticism: the system has been submitted to the general criticism of construction planners that have not participated of the knowledge acquisition process. Some additional historical cases have been used for testing the system to a greater extent. Also, some quantity surveyors that have been involved in cost planning of house building projects have been invited to examine the system, in order to access its potential as a consultancy system for people that have only general knowledge about construction planning.

(iii) Hypothetical test cases: the knowledge base robustness have been tested using a number of test cases that reflect extreme conditions to which the system may be submitted.

(iv) Sensitivity analysis: a number of sensitivity tests have been carried out in order to test how the system's reasoning reacts to subtle changes in a number of important variables, such as project size, availability of resources, and design changes.

Although several adjustments have been made in the system during the acceptance validation, it seems that the main contribution of this type of validation has been to gain a greater appreciation of the structure and limits of the expertise modelled by the system. Moreover, testing the system in real projects has highlighted some aspects of the problem about which the experts involved in the study disagree, as well as has indicated gaps in the domain expertise.

Continual validation, or validation during the system's working life, is particularly relevant for the development of HOUSE PLANNER for two reasons. Firstly, the system seems to be suitable for field validation. Provided it has been designed for supporting planners' decision making, and that users have some control over the planning approach used, the system can be used experimentally in real situations, without causing any serious trouble. Secondly, the presence of several rules-of-thumb in the knowledge base, and the fact that some disagreement was found amongst experts indicate that a fine-tuning of the system is periodically needed, before it is used in a new context.

Although a run-time version of the system has already been used experimentally in two construction companies, testing formally the system during its working life is outside the scope of this research, due to the limited time available. However, a method for continual validation of the knowledge base will be proposed at the end of the research, based mainly on the analysis resulting from early validation and acceptance validation. An additional module of the system will be developed in order to introduce some automation in this method. This module will contain expertise about the way the whole system works, and will be able to track down the section of the knowledge base that leads to unsatisfactory solutions.

5 CONCLUSIONS

Although a number of techniques have been proposed for the validation of expert systems, very little has been reported on the formal validation of real systems. One of the main difficulties for validating expert systems for construction planning is the fact that several alternative feasible solutions exist for each problem, being practically impossible to establish a unique gold standard solution which the performance of an expert system can be compared to.

In this particular study, the validation process has been also constrained by the small number of representative historical cases available, and by the limited amount of time domain experts have been able to dedicate to the analysis of hypothetical test cases, discarding any kind of quantitative analysis. From the qualitative techniques available, none of them seemed to be adequate if applied individually. The adopted approach was to combine a number of those techniques, in order to provide an analysis as formal, as independent, and as exhaustive as possible. This multi-technique approach not only brought up the system to an acceptable level of performance, but also provided some understanding on the structure and limitations of the system, that was very useful for establishing the basis for a continual method of validation.

ACKNOWLEDGEMENTS: Thanks to Alan Gillies and to Stuart Moralee for their very helpful comments on this paper, and to the Brazilian Government for the financial support.

BIBLIOGRAPHY

(1) FORMOSO, C.T. & BRANDON, P.S. **Using expert systems for strategic decision making in house building.** Salford, University of Salford, 1989. (to be published in the International Symposium on Building Economics and Construction Management, in Sydney, Australia, March 1990).

(2) GASCHNIG, J. et al. Evaluation of expert systems. In: HAYES-ROTH, F.; WATERMAN, D.A. & LENAT, D.B.(Eds.) **Building expert systems.** Reading, Massachusetts, Addison-Wesley, 1983.

(3) HETZEL, W. **A complete guide to software testing.** Wellesley, Massachusetts, Information Sciences, 1984.

(4) O'KEEFE, R.M. et al. Validating expert system performance. **IEEE Expert:** pp.81-90, Winter 1987.

(5) PREECE, A. Verification of rule-based expert systems in wide domains. In: SHADBOLT, N. **Research and development of expert systems IV.** Cambridge, Univ. of Cambridge, 1989.

(6) WATTS, R. **Measuring software quality.** Manchester, National Computer Centre, 1987.

(7) WEISS, S.M. & KULIKOWSKI, C.A. **A practical guide to designing expert systems.** London, Chapman & Hall, 1983.