Gregor, S. and Yu, X. (2002). Exploring the explanatory capabilities of intelligent systems technologies. In V. Dimitrov and V. Korotkich ((Eds.). *Fuzzy Logic: A Framework for the New Millennium*. Physica-Verlag, 288-300. [Author's copy]

EXPLORING THE EXPLANATORY CAPABILITIES OF INTELLIGENT SYSTEM TECHNOLOGIES

Shirley Gregor

Xinghuo Yu

Faculty of Informatics and Communication
Central Queensland University
Rockhampton Qld 4702
Australia

s.gregor@cqu.edu.au

x.yu@cqu.edu.au

Abstract

Explanatory capabilities in expert systems can explain to their human users both the knowledge they contain and the reasoning processes they go through. Explanations in such systems, when suitably designed, have been shown to improve performance and learning and result in more positive user perceptions of a system. Justification-type explanations have been found to be particularly effective in leading to positive outcomes. In this paper we explore the use of explanatory capabilities in new-paradigm systems including neural nets, evolutionary computing and fuzzy logic. Explanations in neural net and evolutionary systems are relatively low-level at present, relying mainly on rule explication. There has been surprisingly little research on explanations from fuzzy systems. These systems, being similar to expert systems apart from the extended capacity to represent vague expert knowledge, can in principle provide explanations and justifications in a manner similar to conventional expert systems. Suggestions for providing explanations from the new-paradigm systems are given.

Introduction

Intelligent systems such as those incorporating neural nets, genetic algorithms and fuzzy logic differ from expert systems in that only rarely do the former possess an explanatory capability. Explanatory capabilities in expert systems can explain to their human users both the knowledge they contain and the reasoning processes they go through. Explanations in such systems, when suitably designed, have been shown to improve performance and learning and result in more positive user perceptions of a system (Gregor and Benbasat, 1999).

Many explanation facilities in expert systems provide simple reasoning traces of the path followed by the system to reach a conclusion, which allow answers to How? and Why? type questions from the user. It has been shown, however, that explanations that go beyond a simple trace mechanism, and conform to Toulmin's model of argumentation (Toulmin, 1958) (in that they provide adequate justification for the knowledge offered) should be more persuasive and lead to greater trust, agreement, satisfaction, and acceptance – of the explanation and possibly also of the system as a whole (Gregor and Benbasat, 1999). A justification-type explanation requires “deep” domain knowledge, causal knowledge or generally accepted rules or principles in the relevant field. Deep explanations can incorporate many different types of knowledge: analogies, cases, textbook knowledge, and so on. This deep knowledge is added to the explanation as a warrant or backing for the knowledge contained in the explanation.

The question that motivates this paper is how explanations, and justification-type explanations in particular, can be provided in intelligent systems such as those incorporating neural nets, genetic algorithms and fuzzy logic. There has been some work on explanations in such systems, though the explanations have been of a fairly primitive type. Exploration of desirable type of explanations for these systems is significant because of the positive effects that can flow from explanation use.

The paper proceeds by first outlining the types of explanations that have generally been used in intelligent systems to date and the newer-paradigm intelligent systems with which we are concerned. We then consider possible methods for offering explanations from these more recent forms of intelligent systems. It is suggested that further empirical study is needed.

Conceptual Background

In this section we first consider the nature of explanations as they appear in intelligent systems. We are considering explanations in the human-computer context, where a system provides an explanation to a human user in the course of a consultation. The purposes of the explanation can include: clarifying or making something understandable, justifying or convincing, or acting as an aid in

resolving misunderstandings or disagreements. Though many explanation capabilities have been modelled on explanations in the human-human context, it is recognized that explanations in the human-computer context may differ in important ways. It is not clear that humans expect the same type of explanations from computers as they do from other humans.

In addition, in this section we discuss the nature of intelligent systems in general, focussing on the newer types of systems in which explanation capabilities have not been much studied to date. The distinguishing feature of intelligent systems is that they commonly contain a knowledge component – a computerized version of human tacit and explicit knowledge. Such systems are based on the basic elements of artificial intelligence: knowledge representation, inference, and control (Hayes-Roth, 1997). Earlier forms of intelligent systems included “expert” or “knowledge-based” systems, where the knowledge in the system was held largely in the form of a rule-base of production rules of the form IF..THEN..ELSE. More recent paradigms in intelligent systems have included neural net, evolutionary and fuzzy logic systems.

Explanatory Capabilities

When considering the types of explanations to be found in intelligent systems in general, we adopt the taxonomy proposed by Gregor and Benbasat (1999) that shows how explanation type can be considered in terms of three dimensions: basic content, human-computer interface, and responsiveness. Table 1 shows examples of typical explanations based on the first dimension, content of explanations. Other forms of explanations can be achieved by changing the presentation format (text-based or multimedia) or the provision mechanism (user-invoked, automatic or intelligent) of explanations. In this paper we will be considering primarily the first dimension, the content of the explanation. It is important to note, however, that the other dimensions of explanations are also important. In particular, the degree of effort a user needs to exert to access an explanation is likely to affect the efficacy of the explanation. Empirical evidence shows that explanations will not be used if the user has to exert “too much” effort to get the. Explanations should be provided automatically if this can be done relatively unobtrusively, or by hypertext links, and should be context-specific rather than generic (Gregor and Benbasat, 1999).

The classification of explanations into four different types by content (Table 1) reflects to some extent the historical development of explanation facilities. It includes the types of explanations found in most operational systems and expert system shells (Wick and Slagle, 1989b; Chandrasekaran et al., 1989; Wognum; 1990). Explanations types I, II and III in Table 1 are the three types proposed by Chandrasekaran et al. (1989): trace, justification and control. These explanation types have been used in a number of studies (Dhaliwal, 1993; Hsu, 1993; Mao, 1995; Ye, 1995). Most explanation facilities available in expert system shells (expert system building tools) are limited to the two reasoning trace queries (Type

I): *How* and *Why* (Wick and Slagle, 1989b). These queries were introduced in MYCIN, a system developed in the early 1970s for diagnosing infectious blood diseases (Shortliffe, 1976; Clancey, 1983).

Type II explanations provide justification for the explanation by including “deep” domain knowledge. These explanations have been found to be particularly convincing in a number of empirical studies. For example, Everett (1994) found that confidence, satisfaction and perceptions of usefulness, effectiveness and ease-of-use were all strongly affected by the presence of justification-type explanations. Frustration with the KBS also decreased when automatic justification-type explanations were given.

Table 1

Classification of explanations by content type
(from Gregor and Benbasat, 1999).

<p>Type I. Trace or line of reasoning Chandrasekaran et al.’s (1989) Type I explanations which explain why certain decisions were or were not made by reference to the data and rules used in a particular case. How and Why in MYCIN are this type.</p> <p>Question: Why do you conclude that a tax cut is appropriate? Explanation: Because a tax cut’s preconditions are high inflation and trade deficits, and current conditions include those factors</p> <p>Type II. Justification or support Chandrasekaran et al.’s (1989) Type II explanations which justify part of a reasoning process by linking it to the deep knowledge from which it was derived. These explanations were introduced in the Xplain system (Swartout, 1983).</p> <p>These explanations can be formed by attaching “deep” domain knowledge to portions of a procedure – for example, by attaching “see textbook, p. 36” to the preceding Type I explanation or a direct hypertext link to such text.</p> <p>Type III. Control or strategic Chandrasekaran et al.’s (1989) Type III explanations which explain the system’s control behaviour and problem solving strategy. These explanations were introduced in NEOMYCIN (Clancey, 1983).</p> <p>Question: Why aren’t you suggesting increased tariffs as a way of decreasing trade deficits? Explanation: Because that plan involves political costs. My strategy is to consider politically easier plans first.</p>

Type IV. Terminological

These explanations supply definitional or terminological information. They were distinguished by Swartout and Smoliar (1987).

Question: What is drug sensitivity?

Explanation: A drug sensitivity is an observable deviation that causes something dangerous that is also caused by the drug.

The effect of including justification explanations was so strong that it affected perceptions which should not have differed depending on their presence or absence, notably system ease of use. Although there was absolutely no difference in any aspect of ease of use, subjects receiving justification explanations rated the system significantly easier to use than did those not receiving justifications. All of the justification effects were significant regardless of invocation mode or explanation content treatment (Everett, 1994, p. 307).

Deep (justification) explanations can incorporate many different types of knowledge: analogies, cases, textbook knowledge and so on. The role of deep knowledge in explanations can be explicated further by considering the model of practical reasoning and argumentation provided by Toulmin (Toulmin, 1958; Toulmin et al., 1984). Toulmin's model has been used as a basis for constructing explanation capabilities (Miller and Larson, 1992; Wick and Slagle, 1989a) and in empirical work (Gregor, 1996; Ye, 1995; Ye and Johnson, 1995).

The model provides a basis for the examination of practical reasoning and argumentation, as distinct from formal logic. The model distinguishes the following different parts of an argument:

- claims - the assertions or conclusions that are put forward for acceptance,
- grounds - the factual data that is the foundation for the argument,
- warrants - the justification for moving from the grounds to the claims (examples are rules of thumb and laws of nature),
- backing - the authorisation for the warrant (an example is a legal statute),
- qualifiers - phrases expressing the degree of certainty placed on a claim,
- possible rebuttals - the extraordinary or exceptional circumstances that might undermine the force of the argument.

Arguments that are strong and well-founded are thought to be convincing, while others that are weak or baseless are unconvincing. The model can be applied to explanations in intelligent systems. A rule-trace explanation which has a rule with data premise, certainty factor and conclusion corresponds to the grounds, qualifier and claim in Toulmin's model. In justification-type explanations a warranty and possibly a backing will also be distinguished. Explanations that conform to Toulmin's model should be more persuasive because they contain the elements which are present in convincing human-human arguments. Thus, they should lead to greater trust, agreement, satisfaction and acceptance. The challenge, which is considered in this paper, is how justification-type explanations can be provided in recent forms of intelligent systems.

New-paradigm intelligent systems

It is comparatively easy to provide Type I explanations from expert and knowledge-based systems in the form of a rule-trace of rules that have fired in the course of a consultation. In this paper, we are concerned with how explanations can be formed in more recent systems, the distinguishing features of which are discussed below.

Artificial Neural Networks

An artificial neural network is a computational structure that is inspired by observed processes in natural networks of biological neurons in the brain (Klir and Yuan, 1995). A neural network consists of a mass of neurons, each of which is a simple computational unit, yet the intersections between the neurons emulate the enormous learning capability of the brain in learning patterns from examples. This learning is done by adjusting the so-called weights that represent the interconnection strength of neurons, according to certain learning algorithms.

In general, the learning can be supervised or unsupervised. In a supervised learning algorithm, learning is guided by specifying, for each training input pattern, the class to which the pattern is supposed to belong. That is, the desired response of the network to each training input pattern and its comparison with the actual output of the network are used in the learning algorithm for appropriate adjustments of the weights. These adjustments, whose purpose is to minimize the difference between the desired and actual outputs, are made incrementally. This is essential for facilitating a convergence to a solution (specific values of weights) in which patterns in the training set are recognized with high fidelity. Once a network converges to a solution, it is then capable of classifying each unknown input pattern with other patterns that are close to it in terms of the same distinguishing features. In an unsupervised learning algorithm, the network forms its own classification of patterns. The classification is based on commonalities in certain features of input patterns. This requires that a neural network implementing an unsupervised learning algorithm be able to identify common features across the range of input patterns.

To date a considerable amount of research has been done in the development of fast and efficient neural learning algorithms. It is noted that interpretation of results from neural networks has been problematic as the learning mechanism inside the network, apart from outside mathematical analysis, is not well understood. Some effort has been directed at eliciting knowledge in a form understood by humans by using for example, clustering algorithms, case studies and sensitivity analysis. It is recognized that using neural networks alone in real-life applications may be problematic because of lack of transparency and limited understandability by humans.

Evolutionary Computation

The evolutionary computation paradigm attempts to mimic the processes observed in natural evolution and utilize them for solving a wide range of optimization problems. Techniques in evolutionary computation include genetic algorithms, genetic programming, evolutionary algorithms and strategies. In general, evolutionary computation performs directed random searches using mutation and crossover operations through a given set of alternatives (population of solutions) with the aim of finding the best alternative (solution) with respect to given criteria of goodness. These criteria are required to be expressed in terms of an objective function, which is usually referred to as a fitness function. Generations of evolution of solution populations using evolutionary computation result in best solutions.

The evolutionary computation paradigm has provided an important alternative to optimization problems. The idea of evolving populations of solutions to achieve optimal solutions is appealing and offers an advantage in dealing with problems which are hard to solve using deterministic mathematical optimization methods such as gradient descent and similar approaches, and linear and nonlinear programming. This paradigm, although mentioned decades ago, has only recently shown its power due to the revolutionary advances in computer technology.

Fuzzy Systems

A fuzzy system is any system whose variables (or, at least, some of them) range over states that are fuzzy sets. For each variable, the fuzzy sets are defined on some relevant universal set, which is often an interval of real numbers. In this special but important case, the fuzzy sets are fuzzy numbers, and the associated variables are linguistic variables. Representing states of variables by fuzzy sets is a way of quantifying the variables. Due to the finite resolution of any measuring instrument, appropriate quantification, whose coarseness reflects the limited measurement resolution, is inevitable whenever a variable represents a real world attribute.

Fuzzy systems provide an alternative representation framework to express problems that can not easily be described using deterministic and probabilistic

mathematical models. The idea of allowing vagueness in expression, which is deeply rooted in human's activities, provides a powerful tool to represent a wider range of practical problems which contain uncertainties and ambiguities.

Integration of Intelligent Systems

As described above, the three major paradigms offer different advantages. Integration of these paradigms gives rise to powerful tools for solving difficult practical problems. Major research efforts in intelligent systems have been aimed at the development of hybrid systems that incorporate these three paradigms, taking advantage of their different strengths. Many results have been obtained in the integration of intelligent systems paradigms, including:

- using fuzzy systems as a representation framework and evolutionary computation as an optimization tool to achieve the optimal model which can model complex systems with minimal fuzzy rules and best fuzzy membership functions.
- using fuzzy systems as a representation framework and neural networks as a learning mechanism to build the fuzzy rule base and optimize fuzzy membership functions.
- using fuzzy systems as a representation framework, neural networks as a learning mechanism, and evolutionary computation as an optimization tool to develop optimal fuzzy models.

In conclusion, it should be noted that neural networks and evolutionary computation are about a process that enables learning and optimization while fuzzy logic is a representation tool.

Explanations from new-paradigm Intelligent Systems

In the previous section we discussed how explanation capabilities in general have been developed, primarily in work with expert and knowledge-based systems. We also outlined paradigms for intelligent systems that have evolved more recently. Explanation issues in these newer paradigms (neural networks, evolutionary and fuzzy systems) have not been well explored. In the following sections we advance ideas on how these explanations could be formed.

Explanations in Neural Networks and Evolutionary Computing

Neural networks do not provide explanations directly. However, knowledge elicitation from trained neural networks could provide an explanation. In the example of a question such as “why do you conclude that a tax cut is appropriate”, it is assumed that the neural network has already been well trained to map:

- from given input patterns to output patterns for the supervised learning case,
- from given input patterns to a few clustered groups for the unsupervised learning case.

However, it is not possible to directly provide an explanation to suggest that “because a tax cut’s preconditions are high inflation and trade deficits, and current conditions include those factors etc”. Recent research indicates that it is possible to acquire the above explanation by analysing the distribution of weights in the neural networks to conclude which attributes contribute more significantly than others. This area of research is referred as “classification rule mining” or “sensitivity analysis” (Rahman, Yu and Srinivasan, 1999) .

Klimasauskas (1991, p. 16) describes sensitivity analysis and how it can be applied to ‘explaining’ the thinking in a neural network.

Sensitivity analysis looks at each of the inputs and determines how much and what kind of impact a small change in inputs would have on outputs. If a small change in a particular input causes a drastic change in a particular output, that input might be considered one of the key factors in the current outcome.

From sensitivity analysis an explanation of some form could be attempted, by providing an explanation of why particular input factors are related to the output (Table 2). In terms of Toulmin’s model we have the grounds for the argument readily available but the warrant that shows how the system moved from the grounds to the claims is limited, showing only the relative importance of factors. Thus, a limited form of a Type 1 explanation is possible. The backing for the warrants is even more problematic, because the system learned the method by itself, in a manner that is not transparent, and does not rely on generally held rules or laws. A very limited form of backing can be given by giving the reasons the important factors were included as input to the neural net training (limited Type II). Thus, the limited form of explanation that is possible is reliant on prior domain knowledge – the exact nature of the relationships among the input and output variables is not known and cannot be explained.

Additional mechanisms for explanations have been proposed for hybrid systems and genetic-algorithm systems. A number of authors propose the integration of expert systems and neural nets to provide explanations (Osyk and Vijayaraman, 1995; Yoon, Guimaraes, and Swales, 1994). Goonatilake (1996) shows how a genetic-algorithm insurance system can give explanations by showing the rule set that is generated for an optimal solution.

The types of explanations possible from the hybrid systems are similar to the Type I explanations that are normally possible from expert systems but again it is

difficult to produce Type II justification explanations because the “rules” have been learned by the system from matching input patterns to output patterns, and are not dependent on generally accepted principles or knowledge (Table 3). Again, as with the sensitivity analysis, it is possible to show why particular factors that have been included in the analysis are relevant – but not the details of the underlying causal relationships between these factors and a conclusion.

Table 2: Example explanation from neural net system following sensitivity analysis

Claim (decision):	Deny credit application.
Grounds (data):	Income = \$45,000 p.a. Expenses = \$500 per week Home-owner = no Existing debts = \$5,000
Warrants (rules):	Credit was denied primarily because: <ul style="list-style-type: none"> • income was too low • expenses are too high.
Backing (justification):	Income level has been found to be related to ability to repay loans. The level of expenses has been found to be related to defaults on loans.

Explanations in Fuzzy Systems

Fuzzy systems in some sense are akin to traditional expert systems hence inheriting many features from them. However they are different in that although they share the same rule format, IF-THEN,

- the attributes have far more complex representation power as they represent functions or even function of functions;
- both linguistically and numerically formatted data can be interchangeably processed and shared;
- they possess more representational power as one IF..THEN rule may represent a meta-rule which refers to sub-rules, data base, mathematical models etc.

It may help when considering the relationship between fuzzy systems and rule-based systems to use as an analogy the difference between rules and principles as

distinguished in the area of law. For example, Kosko (1992) considers that the distinction between the artificial intelligence judge and the fuzzy judge reduces to the distinction between rules and principle. Rules apply "in an all-or-none fashion." Principles "have a dimension that rules do not -- the dimension of weight or importance" and the court "cites principles as its justification for adopting and applying a new rule" (Dworkin, 1977). Rules generally outnumber principles. Principles guide while rules specify.

Table 3: Example explanation from hybrid neural net/expert system

(adapted from Yoon et al., 1994)

Claim (decision):	Classify firm as "well-performing"
Grounds (data):	Output-measure A = .7 Output-measure B = .5 (Output measures show stock performance of the firm)
Warrants (rules):	The firm was classified as "well-performing" because <ul style="list-style-type: none"> • Output-measure A > Output-measure B and • Output-measure A > .5
Backing (justification):	Output-measure A has been found to be a reasonable indicator of future stock prices. The relationship between Output-measure A and Output-measure B has been found to be related to future viability of firms.
Qualifier:	Conclusion certainty is definitive.

In principle, explanations from fuzzy systems can be generated in a manner similar to that in knowledge-based systems (Table 4). The difference is that some of the rules given as warrants may contain "fuzzy" concepts. For example, in Table 4, the establishment of "risk tolerance" could have involved a fuzzy reasoning process, and the outcome is a fuzzy value "low" rather than a precise or numerical value.

Conclusions

The conclusion is that explanations in genetic algorithm and neural net systems are relatively unsophisticated at present. The few existing mechanisms rely mainly on rule explication. The possibility exists for providing justifications for the inclusion of the factors in genetic algorithm and neural nets that are found to be important in arriving at outcomes. It would also be possible to explore the use of prior "cases", whose salient features explicate the patterns found by neural nets

and evolutionary systems in explanation functions. This possibility has not been considered in detail here, but appears worthy of consideration.

Fuzzy logic as a representation paradigm can be used to represent expert knowledge. Fuzzy (expert) systems are inherently similar to conventional expert systems except for their extended capacity in representing vague expert knowledge. Hence mechanisms of explanations such as justification-type explanations for conventional expert systems are applicable. Surprisingly there has been little research done in this area so far.

We suggest further empirical work constructing explanatory capabilities for new-paradigm intelligent systems using the methods suggested here. These systems could then be tested empirically to investigate whether the justification-type explanations of the type proposed here adds to their believability and trust with human users, as has been found with rule-based systems.

Table 4: Example explanation from fuzzy system

Claim (decision):	Recommend investment \$50,000 in Term Deposit.
Grounds (data):	Client details: <ul style="list-style-type: none"> • Value of assets = \$200,000 • Current annual income = \$25,000 • Amount available for investment = \$30,000 • Age = 68 • Preference = maximize capital
Warrants (rules):	The client was classed as “low risk tolerance” on the basis of age and preferences. Term Deposit was recommended because: <ul style="list-style-type: none"> • Term Deposit is low risk investment • Client has low risk tolerance • Amount available < \$100,000.
Backing (justification):	Term Deposits are recommended where the client has low risk tolerance and amount to be invested is relatively small. Risk tolerance decision is based on age, income level and preferences for capital versus income. Older clients on fixed incomes are generally classed as low risk.

References

- Chandrasekaran, B., Tanner, M. C., and Josephson, J. R. (1989). Explaining control strategies in problem solving. *IEEE Expert*, Spring.
- Clancey, W. J. (1983). The epistemology of a rule-based expert system - A

- framework for explanation. *Artificial Intelligence*, 20(3), 215-251.
- Dhaliwal, J. S. (1993). *An experimental investigation of the use of explanations provided by knowledge-based systems*. Unpublished doctoral dissertation, University of British Columbia.
- Dworkin, R.M., (1977). *Taking rights seriously*, Harvard University Press, Cambridge, MA.
- Everett, A. M. (1994). *An empirical investigation of the effect of variations in expert system explanation presentation on users' acquisition of expertise and perceptions of the system*. (Doctoral dissertation, University of Nebraska). University Microfilms, Inc. order number AAC 9510968.
- Goonatilake, S. (1996). Risk assessment using intelligent systems, *Insurance Systems Bulletin*, 11(10), April, p. 2.
- Gregor, S. D. (1996). *Explanations from knowledge-based systems for human learning and problem solving*, Unpublished Ph.D. thesis, University of Queensland, Brisbane.
- Gregor, S. and Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *Management Information Systems Quarterly*, 23(4), 497-530.
- Hayes-Roth, F. (1997). Artificial intelligence: What works and what doesn't', *AI Magazine*, Summer, 99-113.
- Hsu, K. (1993). *The effects of cognitive styles and interface designs on expert systems usage: an assessment of knowledge transfer*. (Doctoral dissertation, Memphis State University). University Microfilms, Inc. order number AAC 9402986.
- Klimasauskas, C. (1991). Neural nets tell why. *Dr. Dobb's Journal*, April, 16-24.
- Klir, G. J. and Yuan, B. (1995). *Fuzzy sets and fuzzy logic: Theory and applications*, Prentice Hall.
- Kosko, B. (1992). *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence*, Prentice-Hall.
- Mao, J. (1995). *An Experimental Study of the Use and Effects of Hypertext Based Explanations in Knowledge-Based Systems*. Unpublished PhD. Dissertation, University of British Columbia.
- Miller C. A., and Larson, R. (1992). An explanatory and "argumentative" interface for a model-based diagnostic system. *UIST'92*, Monterey, CA.
- Osyk, B.A. and Vijayaraman, B. S. (1995). Integrating expert systems and neural nets: Exploring the boundaries of AI. *Information Systems Management*, 12(2), 47-55.
- Rahman, M., Yu, X., and Srinivasan, B. (1999). A neural network based approach

- for fast mining characteristic rules. In N. Foo (Ed.), *Lecture Notes in Artificial Intelligence*, Springer-Verlag, 36-47.
- Shortliffe, E.H. (1976). *Computer-based medical consultations: MYCIN*, Elsevier Computer Science Library, New York.
- Swartout, W. R. (1983). What kind of expert should a system be? XPLAIN: a system for creating and explaining expert consulting programs. *Artificial Intelligence*, 21, 285-325.
- Swartout, W. R., and Smoliar, S. W. (1987). On making expert systems more like experts. *Expert Systems*, 4(3), 196-207.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Toulmin, S., Rieke, R., and Janik, A. (1984). *An introduction to reasoning*. New York: Macmillan.
- Wick, M. R., and Slagle, J. R. (1989a). The partitioned support network for expert system justification. *IEEE transactions on systems, man, and cybernetics*, 19(3), 528-535.
- Wick, M. R., and Slagle, J. R. (1989b). An explanation facility for today's expert systems. *IEEE Expert*, Spring, 26-35.
- Wognum, P. M. (1990). *Explanation of automated reasoning: How and why?* Ph.D. thesis, University of Twente, Netherlands.
- Yager P R and Zadeh L A, (1994). *Fuzzy Sets, Neural Networks and Soft Computing*, van Nostrand Reinhold.
- Ye, L. R. (1990). *User requirements for explanation in expert systems*. (Doctoral dissertation, Graduate School, University of Minnesota). University Microfilms International, order number 9100995.
- Ye, L. R. (1995). Value of explanations in expert systems for auditing: an experimental investigation. *Expert Systems with Applications*, 9, 4, 543-556.
- Ye, L. R., and Johnson, P. E. (1995). The impact of explanation facilities on user acceptance of expert system advice. *MIS Quarterly*, June, 157-172.
- Yoon, Y., Guimaraes, T., and Swales, G. (1994). Integrating artificial neural networks with rule-based expert systems. *Decision Support Systems*, 11, 497-507.