Aspects of Intelligent Systems Explanation.

Article · October 2013

DOI: 10.13189/ujca.2013.010204.

1 author:

Keith Darlington
London South Bank University
13 PUBLICATIONS **81** CITATIONS

Universal Journal of Control and Automation 1(2): 40-51, 2013 http://www.hrpub.org
DOI: 10.13189/ujca.2013.010204

Aspects of Intelligent Systems Explanation

Keith Darlington

Senior Lecturer in Knowledge Based Systems, BCIM Faculty, The Knowledge Based Systems Centre, London South Bank University, London, SE1 OAA
*Corresponding author: keithd@lsbu.ac.uk

Abstract Intelligent systems encompass a wide range of software technologies including heuristic and normative expert systems, case-based reasoning systems, and neural networks. This field has been augmented in recent years by Web-based applications, such as recommender systems and the semantic Web. The uses of explanation facilities have their roots in heuristic rule-based expert systems and have long been touted as an important adjunct in intelligent decision support systems. However, in recent years, their uses have been explored in many other intelligent system technologies - particularly those making an impact in e-commerce such as recommender systems. This paper shows how explanation facilities work with a range of symbolic intelligent techniques and, when carefully designed, provide a range of benefits. The paper also shows how, despite being more difficult to augment with non-symbolic technologies, hybrid methods predominantly using rule-extraction techniques have provided moderate success for explanation facilities in a range of ad-hoc applications.

Keywords Artificial Intelligence, Non-Symbolic Technologies, Rule-Extraction Techniques

1. Introduction

One of the assumed strengths of expert systems has been their explanatory capabilities. Expert systems, like any human expert, must be able to explain what they do and justify their actions in terms which are understandable to the user. Their role is considered very important in decision support expert systems. For example, according to [1], "One of the most important lessons of medical computing research is that expert-level decision making performance does not guarantee user acceptance". They describe many expert level systems comparable to that of a human expert that have gained only limited acceptance in the medical environment. Furthermore, according to [2] the ability to explain advice is the single most important feature of a computer based decision support system. They show that explanation can enhance the acceptability of expert systems

However, there was much scepticsm towards explanation facilities in the first generation of expert systems. The main reason for this was that the traditional paradigm for problem solving with expert systems was based upon an explicit model of the domain implemented using rules - mostly, shallow heuristic rules. The attempts to incorporate explanation facilities were first attempted with the heuristic rule-based expert system MYCIN during the late 1970's [3]. During this time, the potential for explanations became apparent because of the way that the explanation chain links problem with solution. This gave developers access to structures for explanation facilities which provided a free adjunct within the development tools available. Nevertheless, researchers soon discovered that, without further effort, they would not provide adequate quality to attract enough attention from users ([3]; [4]). The explanatory inadequacies of MYCIN were also evident in many other systems under development at that time, in that they had attracted little or no interest from end-users ([3]; [4]). For example, [5] point out that: "Some expert systems had been developed without any explanation component at all". They elaborate further by saying that in two cases – a route planning system and a manufacturing selection system – the client had stipulated that explanations were unnecessary and confusing. This reflected the general feeling towards explanation facilities at that time. They were perceived as being better suited to knowledge engineers - for validating system knowledge and testing - than for end-users of the system. For, in those early systems, the available explanations were little more than a trace of the detailed problem-solving steps - sometimes enhanced by canned text. Many other shortcomings of expert system explanation were identified by [6] when he tried to use the MYCIN medical diagnosis expert system for training of junior consultants. His research led to his formulation of an epistemological model that used three types of knowledge: these comprising trace, justification and strategic knowledge. Later, a fourth category called terminological knowledge was added which is a sub-category of justification knowledge.

Other researchers noted that some improvements in explanation facilities were possible by using other knowledge representational methods available at that time, such as frames and semantic networks ([7]; [8]). Furthermore, a range of other AI technologies, such as neural networks [9], case-based reasoning (CBR) [10], and e-commerce technologies, such as recommender systems [11]

could benefit from the provision of explanation facilities. This paper examines the potential for explanation facilities using a range of intelligent system techniques now in widespread use and some ways which describe how they have been incorporated in AI technologies. The paper begins with an appraisal of the empirical research completed to date in this subject and its general implications for designers of explanation facilities. The purpose of this paper is to show how explanation facilities work with a range of symbolic and non-symbolic techniques. This work is important because in recent years, a range of Web-based decision making applications have emerged which could be enhanced with explanation support.

2. Empirical Research and the Implications for Explanation Design

According to the Webster's New Collegiate dictionary, to explain is "to make plain or understandable", "To give the reasons for or cause of ", or "To show the logical developments or relationship of". This definition of an explanation emphasises the need for a mutual discussion to correct a misunderstanding or reconcile differences. From this definition, an explanation will involve two parties: one who delivers the explanation (a human expert, or a computer expert system); the other being the receptor of the explanation (called the user). The recipient (user) of an explanation could be an end-user of the system – such as a novice or expert user. Another category of user could be a system developer (knowledge engineer) who may use an expert system explanation facility during the development phase to verify the correctness of the knowledge base or structure of the knowledge base.

The early years of explanation research focussed heavily on the design of explanation prototypes for expert systems, with little consideration to understanding the user-needs or the theoretical case for their use. According to [12] "much of the work on explanation has been concerned with internal architectural issues, exploring complex architectures for producing more sophisticated explanations and explanatory dialogues. External issues concerning the relationship between explanation provision and the user-system task, has been under-emphasised". A unifying theory was recommended by [13] that provided the basis for judging the quality of an explanation. A great deal of empirical research followed that was encapsulated by the work of [14] who provided detailed research describing the factors affecting explanation usage linked to the theoretical foundations. The components of this unified theory are: the Cognitive Effort Perspective [15], the Production Paradox [16], Toulmin's theory [17], and the Adaptive Character of Thought Rational (ACT-R) theory of skill and expertise acquisition ([18]; [19]). The main implications for successful use of these theories relate to the access mechanism provided, explanation types available, explanation orientation, and the expertise level of the user.

2.1. The Access Mechanism

There are two main types of access to provision of explanation facilities - user-invoked and automatic. User-invoked explanations are explicitly requested by the user, and could be implemented by using hypertext or perhaps using language commands entered on the keyboard ([20]; [61]). Automatic explanations, on the other hand, are automatically provided as determined by the system [14]. Automatic explanations could take the form of intelligent explanations – that is, explanations that are given when deemed necessary or, alternatively, could take the form of embedded text – i.e., explanations that were embedded within the interface dialogue so that the user would always notice them. An experiment was conducted by [22] on the effectiveness of user-invoked versus embedded text explanations provision. Her experimental evaluation sought to discover which explanation provision mechanism would enhance learning the most when using a production-oriented scheduling expert system. Automatic explanations offered the greatest advantage in terms of learning. This result led [22] to conclude that the more difficult it was to access the explanations (I.e. user invoked), the more the subjects perceived the explanation component as a separate computerized tool—as opposed to a natural part of the human-computer interaction—and they, therefore, became less aware of its informational value.

An important theory having a bearing on the access mechanism is the Cognitive Effort Perspective [15] who suggests that expert system users will only invoke explanations, if the perceived benefits in accessing them are outweighed by the amount of mental effort in doing so. Another theory related to the access mechanism is the Production Paradox [16]. This refers to the conflict between work and learning so often prevalent in a working environment. Learning is inhibited by a lack of time, whilst working is inhibited by a lack of knowledge – something that could be improved by easy access to explanations. These theories suggest good reasons for the relative inattentiveness to explanation facilities in the absence of any specific trigger for their use. As a consequence, careful thought has to be given to the design of the access mechanism which will depend upon the user characteristics and their likely triggers to access them ([23]; [14]). Table 1 describes the triggers for explanation use for both expert and novice users. Explanations that require less cognitive effort to access will be used more and will be more effective with respect to performance, learning, or user perceptions.

Table 1. Triggers for explanation usage

User characteristics	Triggers for explanation and cognitive justification
Expert	Predominantly when a perceived incongruity (disagreement with system advice) occurs [20].
Novice	When one of the goals of the user is learning. ([23]; [14]) has shown that explanation usage increases when the user goals are learning rather than problem solving. Cognitive learning theory suggests that learning can be supported by explanations. When the user lacks the knowledge needed to contribute to the problem solving explanations will be used in such circumstances [23]. This knowledge could be terminological (such as when a question is being asked) or knowledge of some problem solving procedure.

Table 2. Definitions of Explanation Type Categories

Type of Knowledge	Description and Purpose	Illustration of question requesting explanations using such knowledge
Terminological knowledge Synonyms: definition knowledge	Knowledge of concepts and relationships of a domain that domain experts use to communicate with each other. In order for one to understand a domain, one must understand the terms used to describe the domain.	What is the definition of gross domestic product?
Justification knowledge Synonyms: Why, descriptive knowledge	"Textbook rudiments" which are required before one can solve problems. Justification knowledge provides abstract factual knowledge about a domain, typically represented declaratively.	Why is inflation dependent on the money supply?
Trace knowledge Synonyms: How, problem solving knowledge.	Knowledge about how tasks have, or are about to be accomplished.	How did you conclude that the patient has diabetes?
Strategic knowledge Synonyms: Control knowledge	Knowledge about the system's control behaviour and problem solving strategy.	Why do you need to know if the patient has ever had mumps?

2.2. Types of Explanation Used

An epistemological model was formulated by [6] that used three types of knowledge: comprising trace (problem solving), justification and strategic knowledge. Table 2 describes the explanation types that can be used in rendering explanation facilities. The surveys conducted by [24]; [20]; [25] and [60] show a propensity towards the use of justification explanations. This is strong evidence, and is cognitively underpinned by [17], who proposed a model of explanation which describes the nature and structure of explanation. Indeed, much of the empirical work completed on explanation has been grounded in this theory ([25]; [24]) and it is the popular choice pattern amongst trace (problem solving), justification, terminological and strategic explanations.

This is, perhaps, not surprising because justification explanations are more conducive to the underlying stages of Toulmin's theory. According to [25], justification is by far the most well used explanation type, and is particularly popular with novice users. Indeed, [20] show that novices use explanations more than trace or strategic explanations, but that expert's use trace as much as justification but more than strategic explanations. Furthermore, [23] found that users whose aim is learning are more likely to use justification knowledge, whereas users whose goals are problem solving are likely to use less justification knowledge and focus on trace or strategic knowledge.

Regarding trace explanations, [26] found that they are more likely to be used for debugging by knowledge engineers than by end users. If the user interface does not require interactive data, then the users need for trace explanations, could become relatively high because they want to know what data the system is currently using to generate conclusions [27]. This means that designers should try to ensure that the user interface is supportive by describing the sub-goals attained during the question phase of a consultation. Also, many researchers emphasise the need for case-specific rather than generic explanations [14]. Terminological explanations are generic whilst trace, justification, and strategic are all examples of case-specific explanations. Support for explanation types described above is now common in expert system development tools although, according to [28], developers often fail to incorporate terminological-type explanation functions. Examples of the uses for these different knowledge types are shown in table 2.

2.3. Explanation Orientation

The explanation component in expert systems would either provide post-advice explanations – called feedback – or explanations before advice – called feedforward. The latter provides the user with a means to find out why a question is being asked during a consultation (i.e., during the data input stage). Feedforward explanations would frequently take the form of a description of technical terms (terminological) to enable the user to answer the question(s) in a meaningful way. Feedforward explanations are general in that they are not dependent on any particular output case.

By contrast, feedback explanations are case-specific in that they will normally present a trace of the rules that were invoked during a consultation and display intermediate inferences in order to arrive at a particular conclusion. Feedback explanations provides the user with a record of problem solving action during a consultation to that the user can see how a conclusion was reached when the data has been completely input.

Very little empirical research has been completed on explanation orientation, apart from that of ([29]; [21]; [20]). Their findings suggest that feedforward explanations are seldom used by experts, but are often used by novice users for clarification or understanding the meaning of terms and definitions and/ or declarative knowledge to enable them to answer questions during the question input stage of a consultation. The work done by [24], show that feedback explanations are more likely to be used by experts than novices and novice users are more likely to use feedforward explanations, especially during the early stages of acquisition of expertise. This is consistent with Anderson's ACT-R [30] theory of learning since feedforward explanations provide for declarative learning through obtaining know-what information prior to procedural problem solving. Hence, embedded feedforward explanations could be important for novice users who may sometimes misunderstand, or make assumptions about the precise nature of a question being asked by the system [22] during a consultation with an expert system.

2.4. The Level of Expertise

Most of the studies conducted on the level of expertise focus on two levels: expert and novice. However, this is frequently an over-simplification. For example, in the healthcare domain, user types could include expert clinicians, junior clinicians, nurses, medical researchers, administrators and patients. The work of [14] lends strong support to the view that explanation facilities influences novices and experts' judgements and supports Anderson's ACT-R) theory of learning [30]. This is a cognitive learning theory which describes skill and expertise acquisition and describes the transition from novice learning to expert, based on the theory that expertise is formed by a three stage process. Studies show little difference in the amount of explanations accessed by novices and experts ([25]; [31]).

2.5. The Potential Benefits of Explanation Facilities

The main benefits arising from the use of explanation facilities are user perceptions and user performance. Many studies show that these benefits are likely to follow when explanation facilities are well-designed. For example, according to [21], explanations have become a core component of most knowledge-based systems (KBS) designed for use by professional decision makers. A common finding in most of these studies show that the inclusion of explanation helps improve user-acceptance in expert systems, although the conclusions in a small number of studies run counter to these findings. For example, [60] found that the inclusion of justification explanations had a profound impact on user acceptance of the system, whilst [32] found that the presence of explanations made little difference to user performance or user acceptance of a KBS. However, there is a substantial amount of empirical research in this field ([14]; [25]; [24]; [20]; [21]) that demonstrates several benefits accruing from the use of explanation in expert systems for both novice and experts. Regarding novice users, the studies by both [32] and [23] found a positive relationship between frequency of novice use of explanations and problem solving performance – both in terms of accuracy of the quality of the decisions made and time taken to make decisions. It has also been shown by [24] that feedback explanation facilities can be of benefit to expert decision making – leading to greater adherence to the system recommendation.

User perceptions are an important determinant that will clearly have a bearing on user intentions for future use. The perceptions arising from the use of explanation facilities include trust, user satisfaction, user acceptance and persuasiveness about the quality of the expert system. Many believe that trust is likely to be the crucial perception that will affect the use of expert system explanations. According to [33] "trust in a system is developed not only by the quality of its results but also by a clear description of how they are derived. This is especially true when first working with an expert system". Furthermore, [34] has shown that there is also a gain from what he calls the explanation effect of expert system generated explanations. Most of the empirical research conducted in recent years using real laboratory expert systems - not simulations - and show overwhelmingly that user acceptance is very high when explanations are used and provided. The studies completed by [14], [29], [24], and [25] show that explanation improves user belief in systems, and can lead to greater accuracy in the ensuing decision making and reduce time in decision making.

However, for explanation facilities to be effective, designers must give careful consideration to the access mechanism, and try to ensure that explanation facilities provided deliver the correct explanation type, mainly justification, the orientation, and expertise level of likely users of the system. With regard to the access mechanism, all the techniques described in the following sections would support automatic explanation through embedded text since this function could be obtained via a pop up window through the user-interface. The other issues will be considered in the following sections.

3. Explanation Using Symbolic Expert System Techniques

Symbolic expert systems operate by using explicitly stored symbols that are logically manipulated during an expert system execution. For example, in first order logic,

the logical manipulations could be entailment relations. The premises and conclusions may represent propositions and be written symbolically with Boolean (true or false) outcomes resulting in the derivation of conclusions from premises. For example, consider the proposition “Cardiff is in Wales”. Let us represent this proposition by the symbol A. Consider also the proposition “Wales is in the UK” represented by the symbol B. Then if the statement “Cardiff is in the UK” is represented by the symbol C, this conclusion can be derived as true. I.e., if A and B is true then this implies that C is also true. This process is explicit and therefore, amenable to explanation with regard to the manner in which conclusions are derived from premises.

Because these symbols are explicitly stored, it is relatively easy – compared with non-symbolic expert systems – to incorporate explanation facilities. The classification of symbolic expert systems is often divided into model-based – which contain a model of the concepts and their relationship with other concepts in the subject domain – and non-model based. Examples of model based representational schemes in common use include rules, object based and Bayesian systems. Examples of non-model based representations include case-based reasoning. The former provide directed inference that make them more amenable to feedforward explanations, unlike non-model based representations, like case-based reasoning, which are not well suited to feedforward explanations.

3.1. Rule-Based Systems

Explanation facilities have their roots in rule-based expert systems because of the natural way that chaining of rules during a consultation can produce a rule-trace. Most of the early research in explanation facilities involved rule-based expert systems. This is not surprising for production rules were the predominant means for representing knowledge in the first generation expert systems that were developed during the late 1970s and early 1980s. The explanation content provided in the first wave of rule-based expert systems would have been predominantly based on “problem solving” knowledge. This will often have amounted to no more than a capability to provide a trace of rules that were invoked during a consultation ([28]; [33]). A rule trace is a record of the system’s run time rule invocation history. In the medical expert system MYCIN [3], the user can request *why* a question is being asked during a consultation – known as feedforward – and the explanation that follows will trace the chain of rules to see what higher goals the system is attempting to achieve. The user can also ask *how* a conclusion was reached – known as feedback – and the explanation that follows will trace the rules fired during the consultation to see what sub goals were satisfied to arrive at the conclusion. Rule trace methods offer some advantages in that they are an integral component of the inference engine. Rule traces are also useful during the testing and debugging phase; this is because the explanations provided by a rule trace reflect the code directly; therefore consistency between code and explanation is guaranteed – hence their popularity with knowledge engineers.

There are, however, several shortcomings with rule traces ([6], [33], and [3]). Explanations based on rule traces always reflect the current structure of the knowledge base: hence, a poorly organised rule base, for instance, with many premises per rule, could destroy the transparency of explanations. Moreover, a rule base may have references to internal procedures – for example, the execution of mathematical calculations – which could make explanations difficult to understand. Some of the problems described above could be curtailed by ensuring that the knowledge base is well designed with the transparency of the trace explanation in mind. For example, internal procedures, such as calculations could be broken down into a series of explainable steps, or by re-organising the knowledge base in a more appropriate form. A rule trace explanation can only link problem and solution: i.e., present the chain of reasoning constructed by the system using problem solving knowledge. However, the trace can only reconstruct from what knowledge is contained in the knowledge base – this would normally be problem solving knowledge. If the builder has not included the justification knowledge in the rule-base, then the system will not be able to justify the existence of the knowledge? For example, if the system contains a rule of the form" A and B imply C." then it cannot explain why this rule exists - beyond repeating the rule encoded in the knowledge base. This means that the justification knowledge – identified in the last section as being the most popular choice with all users – is difficult to implement with rule-based systems because the rules would only store the problem-solving knowledge unless there were a separate layer of rules added to the system which incorporated such knowledge ([33]; [6]). Another problem with, rules is that all the knowledge is stored in a uniform “If ... Then” format and this makes it difficult to separate the knowledge types described earlier – rules can contain strategic and problem solving knowledge. This would be costly to implement and maintain for the builder. Meta-rules were proposed by [6] to store the justification and strategic knowledge, whilst other researchers considered frames ([7]; [8]). Rules would also be suited to feedforward inference because their inferences are chained mechanisms and therefore, conducive to explanation at any point during a consultation.

3.2. Frame or Object-Based Systems

Frame based expert systems [35] store knowledge in a hierarchy of objects, rather like conventional object oriented programs – called frames – and allows data to be passed between frames via inheritance. Frames store items of data connected with them by slots – the slots can be seen as fillers and may be input during run-time by a user, or contain a default value which would override the inherited value or may use demons which are attached to the slots of the frames and performs specific functions by processing or firing an associated set of production rules.

Frames can support all knowledge types for explanation facilities ([7]; [8]). For example, a logical partitioning of the rule set can be provided because production rules can be attached to the slots by demons. A partial explanation as to the purpose of a rule can therefore be given by examining the slot information to which it is attached providing scope for justification explanations. Furthermore, the hierarchical structure of a frame based system identifies the relationships among objects and therefore allows for an explanatory description of an object by identifying the sub-frames attached to it. This provides scope for justification and terminological explanations. Also, inheritance can be used to support explanation through descendant descriptions. That is, the explanation for a frame inheriting the characteristics of a parent frame would contain similar descriptions from inherited slot values. To apply a different slot value to inherited frame from a parent frame a default can be used for the slot value and this default would be explicitly available for explanation. Finally, in frame-based systems, the problem solving knowledge can be separated from the domain knowledge. This facilitates provision of strategic knowledge for such knowledge would be explicitly available through procedural programming. One of the earliest examples of a frame based system that recognized these characteristics was a system called PUFF [8]. PUFF was an expert system that was used for the interpretation of pulmonary function data. It was also commercially successful in that it became a working tool in the pulmonary physiology lab of a large hospital. Lambert and Ringland [7] also examined frame based explanations in the finance domain through a project called Paraflex. This project was concerned with forms of representation particularly suited to this domain. The knowledge base structure displayed in Paraflex has two elements: an inheritance network representing entities in the domain, and problem solving contexts which capture important reasoning. Node and link type diagrams are used to enable a user to follow one 'path' through the knowledge base at a time. The inheritance network displays the relationship between entities in the knowledge base. The use of this frame based representation enables generalities to be made explicit. Thus ensuring the knowledge is available for explanation.

3.3. Case-Based Reasoning

Cased-based expert systems solve new problems by adapting solutions that were used to solve old problems. Case-based reasoning (CBR) systems have in the last decade been used very successfully in certain types of expert systems, such as Help Desk applications [36], finance and the medical domain ([37]; [38], and [39]).

Explanation in CBR can be achieved at a basic level by describing the retrieved case – called knowledge-light. Nevertheless, such explanations can be powerful because a retrieved case is a case-specific description of an actual case. Knowledge-light explanations can emphasise differences in features between a retrieved case and a query case. [10 make a case for knowledge-light explanations for CBR systems in a medical expert system for bronchiolitis treatment. Their system uses precedent cases enhanced with justification text. This justification text is generated so that supporting and non-supporting features can be displayed for the explanation case. This system is further improved by what they call the Fortiori argument. This means that if a decision is appropriate in a retrieved case, then it is even more appropriate in another case where the symptoms supporting the course of action are more corroborated in the retrieved case. Justification type explanations – the most popular choice pattern for both novices and experts alike – are difficult to achieve with knowledge-light applications, because of their limitations in referring to variations on a retrieved case. However, knowledge-light applications could be adequate for experts because being expert would suggest that they could draw inferences themselves about causality from the difference in cases by virtue of their expert knowledge. Knowledge-light applications cannot provide feedforward explanations because they can only compare with other cases in the library when all the features for a case is supplied from the user. This makes knowledge-light CBR applications difficult for novices to use.

Knowledge-intensive CBR systems include rule-based or other representations to perform this role. There are many examples in use. DIRAS [40] is an example of a knowledge intensive CBR system that is used for the diagnosis and treatment of diabetes. DIRAS infers the risk of particular types of complication using a technique called Lazy Induction of Descriptions (LID). It selects the most discriminating feature which is then added to the description. The system then checks to see if all the remaining cases have the same risk level for the complication. This process continues until the set of remaining cases have the same risk level. DIRAS can then produce a report that can be used to manage diabetic patients which includes the most discriminatory values used during the consultation. This report can be understood by non-expert physicians such as patients themselves – meaning that knowledge-intensive applications can be suited to novices to some extent.

Little empirical evidence is available comparing rule-based and case-based explanations, apart from the work completed by [40] who show that, from a relatively small sample, case-based are preferred to rule-based explanations. However, they acknowledge that the results given for their chosen domain may be different in domains where the subjects have differing insights into the causal mechanisms. Nevertheless, as [37] argues, what differentiates CBR from other similar ideas in model-based reasoning is the concreteness of the cases and that may be the reason for their popularity in explanation.

3.4. Explanations for Handling Uncertainty

Uncertainty can be represented in expert systems using both numeric and non-numeric techniques and fuzzy logic. However, if there are multiple conclusions reached by

several reasoning paths to explain, the task quickly becomes difficult [41]. This is one of the reasons why fuzzy logic systems, are despite the fact that they use the same kind of rules as expert systems, difficult to apply for explanation purposes. Numeric approaches attempt inference with uncertainty mainly by using probabilistic measures, whereas non-numerical techniques attempt to do inference about uncertainty by maintaining information about the source of the uncertainty. Examples of the later include Non Monotonic Logics and Cohen's Theory of Endorsements ([39]; [42]) – which attempts to give endorsements to data and rules. For example, a rule which has worked well repeatedly could be endorsed as reliable or as important when true. Rules would then be triggered at least in part of their endorsements. However, most practice systems use numerical methods for handling uncertainty – the two most common techniques being certainty factors and Bayesian inference.

4.5. Explanation using Certainty Factors

Certainty factors provide a simple numerical model at the rule trace level, so with regard to explanation, certainty factor values could be included with content explanations [43]. However, numerical values may not have much meaning to some recipients of explanation. One approach to improving this to achieve explanation would be to map numerical values from the certainty factor range to linguistic terms which can be more easily understood by the recipient of the explanation.

4.6. Explanation in Bayesian Networks

The use of Bayesian inference was first used successfully in the expert system called PROSPECTOR [47], which was used in the evaluation of the mineral potential of a geological site. This system uses subjective probability theory including Bayes theorem supplemented by certainty factors and fuzzy logic. However, explanations using Bayesian inference are difficult to implement in that their reasoning methods follow a normative mathematical approach, using the formulae applied in Bayes theorem. Non-mathematically inclined users may have difficulty in understanding explanations based on this technique [43]. However, [43] describe a range of methods for improving content and interaction explanation using Bayesian inference. They consider different approaches to explanation of evidence, explanation of the model, and explanation of reasoning. In Bayesian systems, an explanation of the model may be appropriate from the builder but may transcend the knowledge of a domain expert, because of the mathematical knowledge required to support the model. They acknowledge that many of these methods have not been implemented in prototype applications but also note that these explanation methods provide very little quality interaction with the user and are difficult for experts and novices alike. Other researchers have examined explanation in Bayesian Networks. For example, [44] developed an approach that considered the most relevant explanation as the best explanation for the given evidence. There are many medical Web expert systems that use Bayesian inference, although from the point of view of explanation, will simply rank likely outcomes according to probabilities [45].

Bayes theorem has also been used to support other AI representational methods for improving explanation. For example, [46] uses a CBR system called ProCon, which adapts using Bayesian inference. He argues that although displaying the most similar case enhances explanation, it can sometimes present contradictory evidence against the result. This transpires when some of the features of the most similar case oppose the result. To incorporate the weights of these features he uses Bayesian inference to update the probabilities of an outcome class being true (defined as hypothesis in Bayes theorem) against the presence or absence of features (defined as evidence in Bayes theorem).

4. Explanation Using Non-Symbolic Expert System Techniques

4.1. Neural Networks

Neural networks provide an example of a non-symbolic AI paradigm that does not use explicit knowledge stored as rules of operation. Neural networks (NNs) work by learning from the use of large amounts of training data. Implicit knowledge is encoded in numeric parameters – called weights – and distributed all over the system. This implicit knowledge is encoded as tens of thousands of numerical input weights that are calibrated by the software as each example is read by the software. The complexity of these calculations could not be understood in explanatory format by human beings. This means that neural networks are not naturally conducive to the generation of explanation structures. Hence, despite the success of neural networks in many areas such as predictability, their widespread acceptance has been impeded by their inability to explain [9].

Nonetheless, despite the black box nature of neural networks, a number of research models have been devised to incorporate explanation. Most of these approaches have been hybrid combining two or more techniques. For example, [48] use a decompositional approach for the extraction of propositional rules from feed-forward neural networks of binary threshold units. Their method decomposes the network into single units, and then extract's rules to describe a unit's behaviour. This is done using a suitable search tree which allows the pruning of the search space. Other approaches have combined genetic algorithms with neural networks to generate limited explanation [49]. The applicability of genetic algorithms is generally confined to optimisation and planning problems for it is more of a heuristic search method rather than a technique that can be readily applied to mimicking expert system knowledge.

Nevertheless, it has been combined well in improving explanation facilities in neural networks [26]. Another approach adopted by [50] attempt to develop an explanation facility in a system called InMES by using a rule extraction method to generate rules from a trained neural network. This system is used for mark up estimation in building projects. Rules are used to validate rules used to check the input data. Rules are also used here to convert rules from the NN to estimate the mark up decision. Their mark-up approach offers the advantage of capturing experiential tacit knowledge and delivering an explanation facility which uses the rules extracted from the rule extraction algorithm. The researchers state that the explanations generated by these methods when combined with rules are justification type explanations because they justify output given a certain combination of input attributes and values from large historical datasets. However, they do not provide a justification rationale in the same way that model-based reasoning does because the generation on the rule-sets are generated from runtime examples. Moreover, these extraction methods are unlikely to support feedforward explanations either because they retrospectively extract rule-sets from the NN. Rule induction is a machine learning technique that is well suited to generating rule-sets from examples facilitating explanation. Rule induction generates formal rules from examples using algorithms – the two most common being ID3 and C5, [51] – enabling the data to be extracted into rule format. For example, [52] developed a Web based medical expert system with a self training heuristic rule induction algorithm that performs self-training to improve diagnosis.

5. Web Based Intelligent Applications

During the last decade, there has been a proliferation of Web-based applications offering the potential for explanation. Examples include recommender systems, the semantic Web and XML.

5.1. Explanation for Recommender Systems

Recommender systems are Internet-based tools, designed to help sites to help consumers find products to purchase [53]. Recommender systems are being used by an ever-increasing number of E-commerce Web sites. Recommender systems are different from traditional expert systems in that they do not only use product knowledge provided by experts but also subjective judgements from other consumers to guide consumers to locating products they will like. This means that explanation content differs from that of expert systems which rely on knowledge from domain experts. There are two main approaches to the implementation of recommender systems: these being collaborative filtering (CF) systems and content based recommender systems. CF systems attempt to imitate collaboration among users for sharing recommendations: these recommendations are computed by identifying a set of similar users according to their user profiles and then recommending products that are highly rated by similar users [54] – similar to the way that users might share interests through word of mouth.

Content based recommender systems, on the other hand, recommend products that are based on the items content, rather than other user ratings. For example, if a customer has bought, or shown an interest in CD's recorded by the rock band, called the Manic Street Preachers, then other items known to have been recorded but assumed not to be owned by this customer may be recommended. The retailing website Amazon uses content based recommendations. However, in practice, these methods would be combined to provide a hybrid solution. [11] and [55] have described one major problem associated with CF systems: their computations are often based on incomplete data – this may result from an insufficient data sample from which CF recommendations are based, and as a result, this means that recommendations are mostly correct but occasionally wrong. Explanation facilities can be of benefit in such systems because they can assist the user in either detecting or estimating the likelihood of errors in the recommendation. [11] and [53] identify many benefits which are similar to those identified of benefit to KBS explanation facilities described in section 2. These include greater user acceptance of the recommender system as a decision making aid. [55] has examined content based explanations and lists transparency, trustworthiness, user performance (i.e., time to make a decision and quality of choice), user satisfaction, persuasiveness, the degree to which a user would be convinced to buy as likely benefits of explanation facilities for recommender systems.

However, the algorithmic methods described have not claimed much success to date, and [46] has advocated the use of CBR to implement recommender based feedforward explanations. His method allows for recommendations to be applied to a travel domain using a system called Top Case and shows how recommender explanations can be made with incomplete queries by drawing upon user user-preferences with respect to attributes not mentioned in the query which could affect the outcome. He also shows how the relevance to any question that the user is asked can be explained in terms of its ability to discriminate between competing cases. Both collaborative filtering and content based explanations are justification type explanations because, unlike expert systems, they are not based on expert knowledge and mostly reflect the opinions of their peers or themselves.

5.2. Semantic Web Explanation

Explanation could be particularly important with regard to the Semantic Web because those abilities can substantially affect its usability and acceptance [61]. The goals of the Semantic Web include the provision of semantics-based applications that could act as knowledgeable assistants for end users by making HTML documents machine readable - the current Web mostly renders pages that are only readable

by human. Some of the technologies used to do this draw upon AI techniques described earlier in this paper. The Semantic Web is a W3C specification [10]. The base-layer is XML and it provides the syntax for content structure for the implementation of higher layers. Some of these higher layers of the semantic web architecture contains layers for ontology's (domain specifications), and descriptive logic (ideas migrated from the use of frame-based representations). These layers enhance the possibility of disseminating justification knowledge for explanation on the Semantic Web, despite the additional complexities arising from the massive increase in scale of the domain itself.

From the research completed on explanation facilities for the Semantic Web to date [56], considers requirements for explanation for the Semantic Web particularly with regard to attribution of sources. They propose an Inference Web to deal with opaque query answers and a description of its main components – this includes a registry that contains details of information sources, along with a proof specification and explanation browser. More recently, [58] also acknowledge the need for explanation facilities to support the Semantic Web. They propose a method of presenting explanations using a hybrid interface consisting of tree-based, graphical and text/logic mechanisms. They emphasis the need for transparency and this means that explanation is important in such applications. Given the enormous size of cyberspace, of particular concern here is not only the source of the information (provenance) but how such information could be manipulated along with the trustworthiness of the methods. The importance of the interaction in delivering explanations for the semantic Web has been considered by ([57]; [62]). They propose an explanation infrastructure which would include an access to browsing ad visualization tools through such a generation of Web browsers geared for explanation. The Semantic Web can also benefit recommender system explanations because it could identify URL similarities in finding neighbours from Semantic searches. The semantic Web is a model-based architecture and this means that it should be capable of providing justification and feedforward explanations and support the needs of expert and novice alike with some effort.

5.3. XML and Explanation

The Web provides an excellent delivery mechanism for expert systems in that high-quality knowledge can be distributed at low cost to users. However, the expert systems that were built for stand-alone computers required complex software tools that worked with specific hardware and software combinations – such technologies would make dissemination on the Web very difficult. XML is a descriptive meta-language which provides a means of structuring data – the structure is contained in the mark-up which can be defined by the developer. XML not only represents one of the layers of the semantic Web but also because it is hardware and software independent – making it very interoperable for use on the Web. Also, XML is a generalised mark-up language which could facilitate extraction and distribution from opaque AI techniques to transparent rule-based techniques which are more amenable to explanation in general applications. For instance, the rules extracted from the system described by [50] could be codified in Rule ML for distribution on the Web. Rule ML is an XML application covering all aspects of Web rules and their interoperation (www.ruleml.org). Moreover, [10] describe CBML, an XML-based Case Mark-Up Language to facilitate such integration. They describe benefits in terms of extensibility, ease of reuse and interoperability. The language allows for the formal definition of the structure of cases that are completely independent of the application code. This means that the structure and definition of cases to be described and modified easily. The CBML language would also allow cases to be exchanged between heterogeneous CBR systems. Standard XML technologies could also assist with batch-oriented Web data extraction from HTM [59]. They use a web-based infrastructure for explanation which could be useful in the development of problem solvers in general.

6. Conclusions and Recommendations

This paper has shown that a strong case can be made for the provision of explanation facilities in intelligent decision making systems providing that designers consider the factors discussed in section two of this paper – particularly regarding the provision of justification explanations. Rule-based reasoning supports specific rule-trace problem-solving explanations well, but not justification or strategic explanations without much additional effort by the builder. Object-based methods are well suited to implementing justification explanations through inheritance. CBR explanations also offer much scope because a retrieved case is a specific description of an actual case and specific explanations are preferred to general explanations. However, justification explanations are more difficult to implement for CBR applications without augmentation with other KBS techniques – such as rules. Development time would clearly increase in such systems as a result of the increased knowledge acquisition effort. In either case, it could be argued that citing an actual case is justification knowledge – particularly for an expert user. This is one of many reasons why CBR offers much scope for explanation facilities, because CBR can be used in such a way that it invites cooperative problem solving through participation with the user by interpretation of inconsistencies that may arise in accepting a recommended case. CBR is also specific, adaptable and combine well with other technologies like decision trees algorithms as described below. Thus, rules could be generated to improve the guidance rules for the adaptation process.

This paper has shown that rule-based expert systems along with object-based and CBR representations are better suited to explanation than non symbolic techniques – such as neural

networks. But several hybrid system approaches have been described in this paper that provide scope for explanation for black-box and other techniques that takes a variety of combinations, such as CBR with rule-based techniques; genetic algorithms with neural networks; and others, to produce moderately successful explanation facilities by data extraction methods. Other combinations, such as decision-tree or rule-based combined with CBR to create knowledge-intensive CBR; Bayes and CBR using the ProCon system, or even CBR and recommender explanation techniques as described in the Top Case example. These examples provide limited evidence to suggest that designers should consider hybrid solutions to improving explanation – assuming that the implementation costs are not too excessive. Empirical research should be conducted to assess possible benefits of explanation facilities in such applications. For, despite the difficulties in implementing justification and feedforward type explanations, these techniques could still provide many benefits - to experts in particular. Moreover, further research would need to consider whether general methods could be used to determine where and when data extraction methods would or would not be appropriate for such techniques.

This paper has extolled the benefits of XML as a means of distributing heterogeneous data on the Web. Further research could examine the viability of XML as a vehicle for various forms of extraction to facilitate explanation on the Web. Future work in this field could consider the viability, through empirical research of this approach when applied to other emerging technologies, such as e-commerce recommender systems. In the case of recommender systems, algorithms that generate explanation facilities have not met with much success and therefore, rule extraction may provide more suitable algorithmic methods.

REFERENCES

[1] Langlotz, C., and Shortliffe, E. Adapting a Consultation System to Critique User Plans. In Coombs, M. (Editor). Developments in Expert Systems. Academic Press, London. 1984.

[2] Teach, R. L., and Shortliffe, E. H. An analysis of physician attitudes regarding computer-based clinical consultation systems. Computers and biomedical research, an international journal, 14(6), 542–558. 1981.

[3] Shortliffe, E. H. Computer-based Medical Consultations: M YCIN, Elsevier Computer Science Library, New York, 1976.

[4] Rogers, Y. What, When How - Explanation facilities. Open University Document. Dept. of Computing, Milton Keynes, 1989.

[5] Berry, D. C., and Broadbent, D. E. "Explanation and Verbalization in a Computer-assisted Search Task," Quarterly Journal of Experimental Psychology (39A), pp. 585-609. 1987.

[6] Clancey, W. J. "The Epistemology of a Rule-based Expert System: A Framework for Explanation," Artificial Intelligence. 20:3, pp. 215-251, 1983.

[7] Lambert, S. and Ringland, G. Representing knowledge in financial expert systems. SERC publication 1989.

[8] Aikens, J. S. PUFF: An expert system for interpretation of pulmonary function data, Computers and Biomedical Research, Vol 16, pp. 199 - 208, 1983.

[9] Baesens, B., Setiono, R., Mues, C., and Vanthienen, J.: Using neural network rule extraction and decision tables for credit risk evaluation. Management Science 49(3), 312–329. 2003.

[10] Doyle, L., Hayes, C., and Cunningham, P. Representing Cases for CBR in XML 2004.

[11] Herlocker, J. L., Konstan, J.A. and Riedl, J. Explaining collaborative filtering recommendations. In Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, pages 241-250. ACM Press, 2000.

[12] Brezillon, P. (editor). Proceedings of the ECAI-92 Workshop W15 "Improving the Use of Knowledge-based Systems with Explanations," Institute Blaise Pascal, Paris, June 1992.

[13] Johnson, P., and Johnson, H. Explanation Facilities and Interactive Systems. Proceedings of the International Workshop on Intelligent User Interfaces, pp 159-166. New York. ACM. 1993.

[14] Gregor, S., and Benbasat, I. Explanations from intelligent systems: theoretical foundations and implications for practice. Management Information Systems Quarterly, 23, pp497-530. 1999.

[15] Payne, J. W., Bettman, J. R., and Johnson, E. J. The Adaptive Decision Maker, Cambridge University Press, Cambridge, England, 1993.

[16] Carroll, J. M., and Rosson, M. B. "Paradox of the Active User," in Interfacing Thought, J. M. Carroll (ed.), pp. 81-111, 1987.

[17] Toulmin, S. The Uses of Argument, Cambridge University Press, Cambridge, England, 1958.

[18] Anderson, J. R. Cognitive Psychology and Its Implications, 3rd ed., W. H. Freeman, New York, 1990.

[19] Ausubel, D. "Learning as Constructing Meaning," in New Directions in Educational Psychology, N. Entwhistle (ed.), Falmer Press, London, pp. 71-82,1985.

[20] Dhaliwal, J. S. An Experimental Investigation of the Use of Explanations Provided by Knowledge-based Systems, Unpublished Doctoral Dissertation, University of British Columbia, 1993.

[21] Mao, J., Benbasat, I., “The Use of Explanations in Knowledge-Based Systems: Cognitive Perspectives and a Process-Tracing Analysis”, Journal of Management Information Systems, Vol. 17(2), pp153-180, 2000.

[22] Moffitt, K. E. An Empirical Test of Expert System Explanation Effect on Incidental Learning and Decision-making, Unpublished Doctoral Dissertation, Arizona State University, 1989.

[23] Gregor, S. D. Explanations from Knowledge-based Systems for Human Learning and Problem Solving, Unpublished

Doctoral Dissertation, University of Queensland, Brisbane, Australia, 1996.

[24] Arnold, V., Clark, N., Collier, P.A., Leech, S. A., and Sutton, S.G. The Differential Use and Effect of Knowledge-Based System Explanations. MIS Quarterly, Vol 30, No 1. 2006.

[25] Ye, L. R., and Johnson, P. E. "The Impact of Explanation Facilities on User Acceptance of Expert System Advice," MIS Quarterly, pp. 157-172, June 1995.

[26] Eberhart, R. C. The role of genetic algorithms in neural network query-based learning and explanation facilities. In International Workshop on Combinations of Genetic Algorithms and Neural Networks (COGANN-92), pages 169-183, 1995.

[27] Ye, L. R. "Value of Explanations in Expert Systems for Auditing: An Experimental Investigation," Expert Systems with Applications (9:4), pp 543-556, 1995.

[28] Nakatsu, R.T. Explanatory Power of Intelligent Systems: A Research Framework. Decision Support in an Uncertain and Complex World: The IFIP TC8/WG8.3 International Conference. 2004.

[29] Gregor, S. D. Explanations from knowledge-based systems and cooperative problem solving: an empirical study. Int. J. Human-Computer Studies, 54, pp 81-105. 2001.

[30] Anderson, J. R. "Acquisition of Cognitive Style," Psychological Review (89:4), pp. 369-406. 1982.

[31] Hsu, K. C., and Steinbart, P.J. Factors Affecting the Use of Different Types of Explanations Provided by Expert Systems. Advances in Accounting Information Systems, Volume 5, pp 215-241. 1997.

[32] Eining, M., and Dorr, P. B. "The Impact of Expert System Usage on Experiential Learning in an Auditing Setting," Journal of Information Systems (5:1) , pp.1-16, 1991.

[33] Swartout, W. R. "What Kind of Expert Should a System Be? XPLAIN: A System for Creating and Explaining Expert Consulting Programs," Artificial Intelligence (21), pp 285-325, 1983.

[34] Swinney, L. "The Explanation Facility and the Explanation Effect," Expert Systems with Applications (9:4), pp 557-567, 1995.

[35] Minsky, M. - A framework for representing knowledge. In P. Winston Ed. The Psychology of computer vision. Mc Grew Hill 1975.

[36] Dearden, A.M., and Bridge, D. G. Choosing a Knowledge Based System to Support a Help Desk. Knowledge Engineering Review, 8(3), pp201 - 222. 1993.

[37] Kolodner, J. L. Case-based Reasoning, Morgan Kaufmann. 1993.

[38] Yoon-Joo, P. Byung-Chun, K. And See-Hak, C. New Knowledge Extraction Techniques Using Probability for Case-Based Reasoning: Application to Medical Diagnosis. Expert Systems. Vol. 23, No. 1. February 2006,

[39] Darlington, K. The Essence of Expert Systems. Published by Prentice Hall 2000, ISBN 0-13-022774-9. 2000.

[40] Cunningham P., Doyle D., and Loughrey J. “An Evaluation of the Usefulness of Case-Based Explanation”, Proceedings of the 5th International Conference on Case-Based Reasoning (ICCBR 2003), Springer, pp122-130, 2003.

[41] Framling, K., and Graillot, D. Extracting Explanations from Neural Networks. 1998

[42] Giarratano, J. C., and Riley, G. D. Expert Systems: Principles and Programming PWS Kent. 1992.

[43] Lacave, C. F., and Dıez, J. A review of explanation methods for Bayesian networks. Knowledge Engineering Review, 17:107–127, 2002.

[44] Yuan, C., Lim, H., and Tsai-Ching, L. Most Relevant Explanations in Bayesian Networks. Journal of Artificial Intelligence Research 42 (2011) 309-352. 2011.

[45] Bramer, M. Knowledge Web: A Public Domain Expert System Delivery Environment. IEEE. 2003.

[46] McSherry, D. Explanation in Recommender Systems. Artif. Intell. Rev. 24(2): 179-197. 2005.

[47] Duda, R. O., and Reboh, R. AI and Decision Making: The PROSPECTOR Experience. Artificial Intelligence Applications for Business. W. Reitman. Norwood, NJ, Ablex: 111-147. 1984.

[48] Bader , S. Hölldobler , S. and Mayer-Eichberger, V. Extracting Propositional Rules from Feed-forward Neural Networks — A New Decompositional Approach 2000.

[49] Eberhart, R. C., and Dobbins, R. W. Designing neural network explanation facilities using genetic algorithms, IEEE, 1991.

[50] Li, H., and Love, P.E.D. Combining rule-based expert systems and artificial neural networks for mark-up estimation. Construction Management and Economics, 17, pp 169-176. 1999.

[51] Quinlan, J. R. Induction of Decision Trees. Mach. Learn. 1, 1 pp 81-106. 1986.

[52] Chorbev, I., Mihajlov, D., and Jolevski, I. Web Based Medical Expert System with a Self Training Heuristic Rule Induction Algorithm. First International Conference on Advances in Databases, Knowledge, and Data Applications. Guadeloupe, France. 2009.

[53] .Ricci, D. R. Fesenmaier, M. Mirzadeh, H. Rumetshofer, E. Schaumlechner, A. Venturini, K. W. Wöber and A. Zins, "DIETORECS: A Case-Based Travel Advisory System" in Destination Recommendation Systems: Behavioural Foundations and Applications, CABI Publishing, June 2006

[54] Koren, J., and Bell, P. Advances in Collaborative Filtering," in Ricci et al. (editors.) Recommender Systems Handbook, Chapter 5, 145-186. 2011.

[55] Tintarev, N., and Masthoff, J. A Survey of Explanations in Recommender Systems. In G Uchyigit (Ed), Workshop on Recommender Systems and Intelligent User Interfaces associated with ICDE07, Instabul, Turkey, @IEEE. 2007

[56] McGuinness, D., and Pinhiero da Silva, P., Inference Web: Portable Explanations for the Web. Stanford Knowledge Systems AI Laboratory. 2003.

[57] McGuinness, D., Ding, L., Glass, A., Chang, C., Zeng, H., and Furtado, V. Explanation Interfaces for the Semantic Web: Issues and Models. International Conference on

Intelligent User Interfaces (IUI'06). 2005.

[58] Kontopoulos, E., Bassiliades, N., and Antoniou, G. Visualizing Semantic Web proofs of defeasible logic in the DR-DEVICE system. Knowledge Based Systems. Vol 24, Issue 3. pp 406-419. 2011.

[59] Everett, A. M. An Empirical Investigation of the Effect of Variations in Expert System Explanation Presentation on Users' Acquisition of Expertise and Perceptions of the System, Unpublished Doctoral Dissertation, University of Nebraska, 1994.

[60] Roth-Berghofer, T., Forcher, B. Improving Understandability of Semantic Search Explanations. Int. J. Knowledge Engineering and Data Mining, Vol. 1, No. 3, 2011

[61] Mao, J., and Benbasat, I. "The Effects of Hypertext-based Explanations in Knowledge-based Systems on Explanation Use and Knowledge Transfer," Working Paper 96-MIS-001, Faculty of Commerce, University of British Columbia, 1996.

[62] Hasan, R., Gandon, F. Explanation in the Semantic Web: a survey of the state of the art. Research Report. N° 7974. INRIA Sophia Antipolis Méditerranée. 2012.