December 1999

A Theoretical Basis for the Assessment of Rule-Based System Reliability

Teresa Shaft
University of Oklahoma

Rose Gamble
University of Tulsa

A Theoretical Basis for the Assessment of Rule-Based System Reliability

Teresa M. Shaft, Division of Management Information Systems
University of Oklahoma
tshaft@ou.edu

Rose F. Gamble, Department of Mathematical & Computer Sciences
University of Tulsa
gamble@utulsa.edu

Introduction

Determining the reliability of KBS has become an important research area due to the application of KBSs to areas where misplaced confidence can cause large monetary losses or even loss of life. Developers need a set of criteria to evaluate KBS's reliability, i.e. *reliability criteria*. Using arguments from philosophy of science, we define three criteria, thus providing a theoretical basis for judging the reliability of a KBS. Previous researchers have argued for or against specific criteria, however there is little agreement on definitions and groupings of these criteria (Nazareth and Kennedy 1993). The lack of agreement is, in part, due to the lack of a theoretical foundation underpinning KBS reliability. A theoretical viewpoint provides a basis for understanding and generalizing results. Without a theory to explain results, findings may be idiosyncratic to a particular system. Worse, findings that may not be idiosyncratic are easier to dismiss as they lack a basis for generalizability.

We conceptualize KBS reliability as determining the criteria under which their results should be accepted. Hence, the issue of whether to consider a KBS reliable is similar to philosophy of science's concern with determining the criteria underwhich the results of scientific investigations should be accepted. By drawing an analogy between conducting scientific investigations and constructing KBSs, we rely upon philosophy of science arguments to establish KBS reliability criteria. Philosophers have debated the proper criteria by which scientific investigations should be evaluated longer than KBSs have existed. By adapting these criteria we gain the benefit of the philosophical debate.

Consistent with traditional software development, we distinguish between issues of verification and validation (V & V). Verification is the process of showing that the resulting KBS meets its specification, that one is building the product right. Validation questions if the right system was built. Establishing the validity of a KBS is often more difficult than for a conventional system (O'Keefe and O'Leary, 1993).

Criteria For Reliable Knowledge Based Systems

To establish the reliability criteria, we rely on criteria developed by Lakatos (1970, 1978) who built on the strengths of Popper (1968) and Kuhn's (1970) philosophies, while addressing criticisms of their work (Blaug 1980). Popper argued that theories must be specified in a falsifiable manner and subjected to rigorous testing. This approach was criticized as "naive falsificationism," viewing theories are separate entities that can be assessed independently and a single experiment can lead to acceptance or rejection of a theory. This approach can lead to a decline in knowledge, leading scientists to study ever smaller areas of inquiry of decreasing scientific content (Lakatos 1978). Kuhn argued that science is marked by the rise and fall of paradigms: "theories come to rise, not one at a time, but linked together in a more or less integrated network of ideas" (Blaug 1980, p. 137). Paradigms are overthrown in a revolutionary fashion whereby the burden of anomalies (unexplained results) weighs so heavily upon a paradigm that a shift occurs whereby a new paradigm becomes dominant. This view suffers from several criticisms including: the criteria for the overthrow of a theory are not scientific but sociological, a paradigm shift does not necessarily entail scientific progress (the growth of knowledge), and that paradigms do not shift via revolutions, but only after years of dedicated scientific work and the weight of much evidence (Lakatos, 1970).

Lakatos argued for a "sophisticated falsificationism" as part of a methodology of scientific research programmes. A scientific research programme (SRP) is disciplinary, typically worked on by many researchers joined by their conviction to the SRP. SRPs incorporate the concept that theories do not exist in isolation, "we propose a maze of theories, and Nature may shout INCONSISTENT"(Lakatos 1970, p. 130). This addresses a weakness of naive falsificationism, that theories can be assessed independently, while incorporating Kuhn's argument that groups of theories create a unified whole. Lakatos (1978) developed criteria for admitting a theory to a SRP. These criteria create a foundation for KBS reliability and Lakatos' dedication to the growth of knowledge ensures that the criteria avoid the pitfall of increasing reliability by decreasing content; i.e., creating very reliable KBS for tiny domains.

A SRP has two major components: a *hard core* and a *protective belt* (Lakatos, 1970). The hard core forms the basis of the SRP and is considered irrefutable because it is typically too abstract and imprecise to be tested explicitly. The protective belt is comprised of theories derived from the hard core. The SRP is furthered by deriving and testing theories to form the protective belt. The criteria for evaluating scientific investigations focuses on the theories offered by the SRP, specifically the *acceptability* of a theory (Lakatos 1978). There are three levels of acceptability. *Acceptability$_1$* assesses the "boldness" of a theory; it must entail some "novel factual hypothesis" (Lakatos 1978, p. 170). *Acceptability$_2$* evaluates the evidence for a theory; bold theories, having met *acceptability$_1$*, undergo severe tests to determine if they are corroborated by evidence. *Acceptability$_3$* appraises the future performance of a theory; its "fitness to survive." Theories that meet the three criteria are included in the protective belt.

Acceptability can be applied to a KBS by drawing an analogy between a KBS and a SRP. The KBS's specification is analogous to the hard core of a SRP. They serve similar purposes, to define the problem space (domain) of interest. A typical specification is similar to the hard core in that it is also abstract and must be refined. The specification is the foundation from which rules are derived and included in the knowledge base (KB), just as the hard core is the foundation from which theories are derived and included in the protective belt. The rules that comprise the KB are analogous to the theories contained in the protective belt of a SRP; i.e., a rule in a KB is equivalent to a theory among the "maze of theories" put forth by a SRP. The analogy between a KBS and a SRP focuses on the relationship between the rules of the KB and the theories of the protective belt. This analogy provides a theoretical justification for the KBS community's focus on rules, as opposed to the inference engine, etc. to assess V & V.

The analogy between a KBS and a SRP extends to include verification and validation. Verification, in the context of a KBS, is a matter of establishing that the rules faithfully represent the specification. With respect to a SRP, verification questions how well the theories cover the hard core. Validation asks if the right system was specified, questioning the reliability of the specification. With respect to a SRP, validation questions the veracity of the hard core. The hard core is *irrefutable* because its veracity cannot be assessed directly (Lakatos 1970). Instead, the veracity of the hard core is assessed indirectly and the three forms of acceptability aid in the assessment.

Acceptability$_1$: Non-redundancy

Acceptability$_1$ appraises the "boldness" of a theory. A theory must have *excess content* over other theories, offering an explanation for phenomena not explained by other theories within the protective belt of the SRP. "[O]ne cannot decide whether a theory is bold by examining the theory in isolation" (Lakatos, 1978:171), instead a theory must be examined in the context of the other theories. Boldness is a verification issue because only the theories that comprise the protective belt are questioned. For KBSs, boldness implies that each rule must contain content that is not contained in other rules; i.e., non-redundant. Non-redundancy is a verification issue because it considers only the rules, not the specification. Four forms of redundancy can be identified: duplication, subsumption, unnecessary IF, and chained redundancy. Redundancy has been considered undesirable due to the complications that may arise during development and maintenance, e.g., the effects associated with altering or deleting only one instance of a rule in a set of redundant rules. From a theoretical standpoint, redundant rules add no knowledge and should not be included.

Acceptability$_2$: Consistency

Acceptability$_2$ addresses the *corroboration* of bold theories, i.e., theories that met the criterion of *acceptability$_1$*. A theory is corroborated if it "entails some novel facts" (Lakatos 1978, p. 174), i.e., is the empirical evidence consistent or inconsistent with the theory? The key issue for KBSs is *consistency*. *Acceptability$_2$* has two interpretations. The first asks if the theory has been corroborated, is it consistent with evidence. This is a verification issue concerning the protective belt, not the hard core. The second interpretation asks if the theory moves the SRP nearer to the truth, accurately portraying the true state of nature, a validation issue addressing the hard core of the SRP.

Consistent with *acceptability$_2$*, we define two aspects of consistency: *non-conflicting rules* (a verification issue) and *accuracy* (a validation issue). Non-conflicting rules are evaluated by considering the entire contents of the KB. Conflicts among rules exist when more than one rule can succeed, but with contradictory consequences. This state is undesirable in a KBS just as it is undesirable to construct a protective belt containing theories that predict contradictory states of nature to be *true* under identical circumstances. Conflicts includes three forms: *direct*, *chained*, and *complex*. Accuracy, the second interpretation of *acceptability$_2$*, is a validation issue and concerns whether or not the KBS reflects the problem space. A KBS that does not reflect the problem space would not be considered reliable, just as a SRP that does not reflect nature would not be acceptable. To assess accuracy, one could argue that it is necessary to gather evidence to support or refute every portion of the specification. Such an argument ignores the process used to develop the specification, i.e., interviews with human experts, probing written documents, etc. We argue that if the knowledge encompassed by the specification was acquirable, then there is reason to believe that it reflects the domain. This line of reasoning assumes that human experts, written source documents, etc., do not intentionally deceive. However, some inaccurate knowledge may be specified. Hence, knowledge engineers should consult multiple sources for knowledge to expose possible inaccurate information. To paraphrase Lakatos: we propose a maze of knowledge and the analysis may shout inconsistent. Accuracy cannot be assessed directly, but is addressed indirectly through careful development of the specification.

Acceptability$_3$: Viability

Acceptability$_3$ appraises the future performance of a theory, specifically its *trustworthiness* and its *fitness to survive* (Lakatos, 1978). The concern is the future usefulness of a SRP; its ability to predict scientific phenomena. With respect to KBSs we are concerned with the performance of the KBS, i.e., its ability to generate useful answers; will it be a *viable* system? Recall that a primary concern is the growth of knowledge, which can only be accomplished by considering all forms of acceptability. Appraising *acceptability$_3$* without assessing the other forms of acceptability could lead to accepting theories with great "total evidential support" but less content than earlier theories (Lakatos 1978), leading to the degeneration of an SRP and its ultimate demise. Similarly, assessing *viability* without considering *non-redundancy* and *consistency* could lead to the development of KBs that contain rules with great "total evidential support," yet in total do not address the breath of the problem space. Such systems would give accurate answers within a very narrow problem space. However, too small a problem space does not protect one from costly errors, specifically errors of omission. Hence we would not label such KBSs reliable. Viability has two aspects, *completeness* and *coverage*. The KBS should completely address the problem space as defined by the specification, a verification issue. Coverage asks if the KBS addresses the true problem space, a validation issue.

A complete KBS addresses the breadth of the domain defined by the specification, containing rules to move from an initial state to a goal state. If the system cannot eventually achieve a goal state from a legal initial state, the rules are incomplete. Violations of completeness occur due to *circularities* or *gaps* (missing rules, unreachable clauses, or deadend clauses). Coverage, the second interpretation of viability, assesses a KBSs "fitness to survive," its long-term usefulness which concerns the relationship of the KBS to the real world domain. The primary concern is if the KBS covers the breadth of the problem space, a validation issue. Assessing this criterion requires a clear understanding of a KBS's intended purpose. This problem is similar to the problem encountered during requirements analysis in traditional software development. Without a thorough understanding of the KBS's requirements, however, the knowledge engineer can only hope to build a KBS that meets the needs of the user

Conclusions

Three criteria for KBS reliability were developed using arguments from philosophy of science: non-redundancy, consistency and viability. Although the reliability criteria are based on theoretical arguments from philosophy of science, they are generally consistent with experience. Through trial and error, researchers have discovered characteristics that enhance reliability. Elsewhere we have demonstrated the application of these criterion to KBS development (Gamble & Shaft, 1996). There are significant efforts to develop tools to assess V & V of KBSs, particularly for verification (cf. Zlatareva & Preece, 1993). However, that the lack of consensus regarding definitions and criteria may have hindered progress in their development as these tools often focus on a single criterion (e.g., redundancy). The introduction of theoretical arguments provides a rigorous basis to assess reliability and a foundation for future V & V tools.

References

Blaug, M. "Kuhn versus Lakatos, or paradigms versus research programmes in the history of economics," in *Paradigms and Revolutions,* G. Gutting (ed.), University of Notre Dame Press, Notre Dame, 1980.

Gamble, R. and T. M. Shaft "Eliminating Redundancy, Conflict, and Incompleteness in Knowledge-Based Systems." *International Journal of Software Engineering and Knowledge Engineering,* 6(4), 1996, 673-697.

Kuhn, T. *The Structure of Scientific Revolutions, Second Edition,* The University of Chicago Press, 1970.

Lakatos, I. "Falsification and the methodology of scientific research programs," in *Criticism and the Growth of Knowledge,* I. Lakatos and A. Musgrave (eds.), Cambridge University Press, Cambridge, 1970, pp. 91-196.

Lakatos, I. *Mathematics, Science, and Epistemology,* Cambridge University Press, Cambridge, 1978.

Nazareth, D.L., and Kennedy, M.H. "Knowledge based system verification and validation: The evolution of a discipline," *International Journal of Expert Systems* (6:2), 1993, pp. 143-162.

O'Keefe, R.M., and O'Leary, D.E. "Expert system verification and validation: A survey and tutorial," *AI Review* (7), 1993, pp. 3-42.

Popper, K. *The Logic of Scientific Discovery,* Harper and Row, New York, 1968.

Zlatareva, N. and Preece, A. " An efficient logical framework for knowledge-based systems verification," *International Journal of Expert Systems,* 7(3), 1993.