PERGAMON

Expert Systems with Applications 18 (2000) 1–16

Expert Systems with Applications

www.elsevier.com/locate/eswa

Validation of intelligent systems: a critical study and a tool

E. Mosqueira-Rey, V. Moret-Bonillo*

Laboratory for Research and Development in Artificial Intelligence (LIDIA), Department of Computer Science, Campus de Elviña, University of A Coruña 15071, A Coruña, Spain

Abstract

One of the most important phases in the methodology for the development of intelligent systems is that corresponding to the evaluation of the performance of the implemented product. This process is popularly known as verification and validation (V&V). The majority of tools designed to support the V&V process are preferentially directed at verification in detriment to validation, and limited to an analysis of the internal structures of the system. The authors of this article propose a methodology for the development of a results-oriented validation, and a tool (SHIVA) is presented which facilitates the fulfilment of the tasks included in the methodology, whilst covering quantitative as well as heuristic aspects. The result is an intelligent tool for the validation of intelligent systems.

Keywords: Artificial intelligence; Intelligent systems; Validation of intelligent systems; Validation tools

1. Introduction

For many years now, software engineering has been concerned with the development of methods and techniques for the definition, construction and maintenance of quality software. Classical methodologies are not, however, entirely suitable to knowledge engineering. In the words of Morris (1985), the world of artificial intelligence does not quite fit into the grey areas of software engineering.

One of the development methodologies which has had most repercussions in the field of intelligent systems is the spiral methodology, proposed by Boehm (1988). This methodology is recommended by several authors, among them Lee and O'Keefe (1994), Noblett and Jones (1991), and Cardeñosa et al. (1991), because it permits the inclusion of concepts such as incremental development and fast prototyping, fundamental in the development of an intelligent system. Fig. 1 illustrates an example of the spiral methodology, taken from Lee and O'Keefe (1994).

In the spiral methodology the final stages of each development cycle are given over to the testing of the quality of the developed product. These stages are commonly known as verification and validation, or simply V&V.

Verification refers, according to Boehm (1981), to building the system right. When it is a question of intelligent systems, this definition is described as 'testing that the system has no errors and complies with its initial specifications'.

Validation on the other hand, and again according to Boehm (1981), refers to building the right system, and this concept expressed in terms of intelligent systems implies testing that the output of the system is correct and complies with the needs and requirements of the user.

As a follow-up to validation, many authors include one or several additional phases—commonly grouped together under the term *evaluation*—whereby aspects that go beyond the validity of the final solutions are analysed. Evaluation thus is an endeavour to analyse aspects such as utility, robustness, velocity, efficiency, extension possibilities, ease of use, credibility, etc.

The validation phase may be viewed from two different perspectives:

- *Results-oriented validation* (Lee and O'Keefe, 1994) compares the performance of the system with an expected performance (provided by a standard reference or by human experts) to ensure that the system reaches an acceptable performance level.
- *Usage-oriented validation* goes beyond the correction of the results obtained by the system and concentrates on matters referring to the man–machine interaction. This type of validation is generally referred to in the literature in terms of 'assessment' (O'Leary, 1987) and there are authors such as Liebowitz (1986), who include it within the evaluation phase.

Results-oriented validation is normally a prerequisite for

* Corresponding author. Tel.: +34-981-167000. Ext. 1248; fax: +34-981-167160.

E-mail address: civmoret@udc.es (V. Moret-Bonillo)

0957-4174/00/$ - see front matter
PII: S0957-4174(99)00045-7

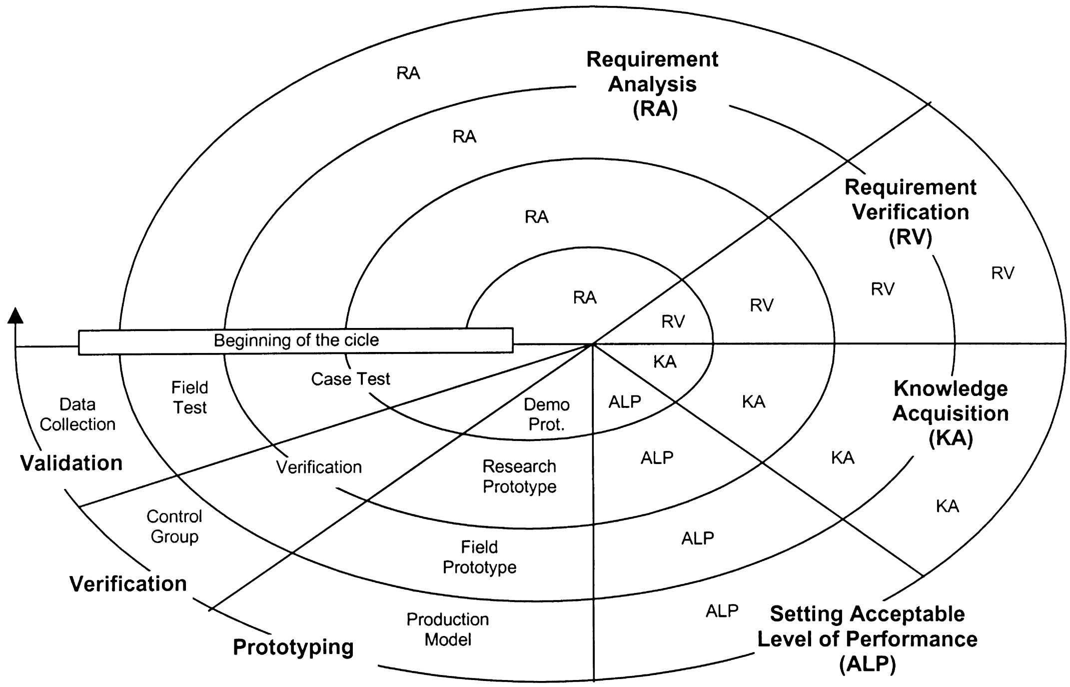

Fig. 1. Spiral development methodology for intelligent systems (Lee and O'Keefe model).

the implementation of a validation oriented to use. Thus, if the system fails to render a satisfactory performance (or fails to give indications that performance will improve in the future on making improvements in its development), then aspects concerning its utilisation are irrelevant. For this reason, in this article the study of results-oriented validation, which henceforth shall be referred to simply as 'validation', is the central issue.

The complexity of the V&V processes has motivated endeavours to automate these phases, although success in this respect has been mixed. Thus, while considerable advances have been made in the matter of verification—with the construction of commercial shells (that partially eliminate the need to carry out a verification of the inference engine) and specific tools such as EVA (Stachowitz and Combs, 1987), CHECK (Nguyen et al., 1987) and COVER (Preece et al., 1992) that facilitate the verification of knowledge bases—the validation of intelligent systems continues to be a poorly structured field of research, in which many ad hoc approaches have been developed, but where an overall vision of developments in the field is still lacking.

In a recent study, Murrel and Plant (1997) analysed the principal verification and validation tools that were discussed in the bibliography between the years 1985 and 1995, and from this study it is clear that the majority of the tools described carry out verification or refinement tasks. Refinement is an intermediate phase between verification and validation, concerned with applying 'white box' tests to the system. Among the different knowledge refinement tools deserving mention are KVAT (Mengshoel, 1993), SEEK (Politakis, 1985), SEEK2 (Ginsberg and Weiss, 1985) and DIVER (Zlatareva, 1998).

In this article, we wish to move up a step on the pyramid of behaviour analysis by carrying out a validation in which the intelligent system is treated as a 'black box', whilst concentrating on a results-oriented validation.

It is evident that an automation of the validation phase brings with it many advantages, since it permits the construction of systems with a guarantee of quality and more easily. Nonetheless, before proceeding to the automation of this phase it is necessary to identify and organise the tasks that must be carried out and to construct a validation methodology that indicates the steps to be followed in the process.

For this reason the objectives of this research are: (1) to identify and study the different processes related to validation; (2) to construct a methodology that characterises the validation process and that indicates the task to be carried out at each step of the way; and finally (3) to construct a computational tool that automates, as far as possible, the distinct phases of the methodology.

2. The validation of intelligent systems

A study of validation indicates that there does not exist a global classification of the problems to be resolved, nor is there a clear relationship between these problems and the techniques destined to resolve them (López et al., 1990). Gupta (1993) points out that—from among the current most important validation problems—those that stand out