

Evaluation of Fuzzy Intelligent Learning Systems

F. de Arriaga*¹, M. El Alami², A. Arriaga³

^{1,3} Universidad Politécnica de Madrid, E. T. Sup. de Ing. de Telecomunicación, 28040, Madrid, Spain

² Universidad Alfonso X El Sabio, Villanueva de la Cañada, (Madrid), Spain

The evaluation of Intelligent Learning Systems yielding complex models as formal representations of student's behavior is still an open problem. Standards so far do not provide suitable methods for the evaluation of competence. The paper discusses some guidelines according to the APA Standards spirit and apply them to the evaluation of FINANCE, a system derived from NEOCAMPUS2, a long effort project, devoted to coaching and the acceleration of the transfer of novices into experts. Results show its high reliability and important consequential and evidential validity.

Keywords evaluation; fuzzy logic; intelligent learning systems; multi-agent systems;

1. Introduction

Intelligent Tutoring Systems (ITS) have been introduced into the educational scene with the advent of SCHOLAR in 1972. Their use has increased during the past twenty years and also their functionality, robustness and emphasis on student's learning; this change is reflected in the transition from ITS to ILS (Intelligent Learning Systems) [1]. Multi-agent systems have collaborated quite a bit to the enhancement of ILS because their flexibility, learning and collaborative capabilities have made possible the adoption of sophisticated procedures and complex models, both educational and cognitive [2], [3]. The extensive use of the Internet has opened the possibility of web-based tutors and E-learning [4], [5]. However, the evaluation of those systems has not followed a parallel evolution.

An interesting experience is due to VanLehn [6] related to the evaluation of an assessment system. After considering that the 1985 Standards for Educational and Psychological Testing proposed by APA and other institutions do not include any provisions for performance assessment, he starts from Messick's principle of validity [7] and decomposes it into evidential validity, consequential validity and reliability. Iqbal and coworkers [8] analyse existing evaluation methods and classify twenty of them according to their possibilities for internal or external evaluation purposes and also according to the depth of the analysis carried out with the method. The same year APA, AERA and NCME publish the 1999 Standards for Educational and Psychological Testing; it is a revised version, also founded on the concepts of validity and reliability, it does not yet include appropriate methods for the evaluation of performance assessment in complex environments. Also Vanlehn and Niu [9] report on a sensitivity analysis of the student modeling module of the Andes ITS. For that purpose they varied parameter values and structural features of the assessor module. They used simulated students with a different randomly generated knowledge in order to account for the many different conditions required by the sensitivity analysis. They obtained measures of the overall and average accuracy of the assessment.

2. NEOCAMPUS2: A Fuzzy Intelligent Research System

Neocampus is a long-range effort fully devoted to do research on several problems related to a new generation of Intelligent Learning Systems, and to the subsequent construction of different spin-off real systems devoted to reinforce learning, and to accelerate the transfer of apprentices into experts.

The main objectives of NEOCAMPUS2, from the user's point of view are the following: a) to build a software environment for the construction of multi-agent intelligent systems able to be cloned if necessary, make their own decisions, learn from experience, cooperate among them, and communicate

