1 Benchmarking of Multiagent Systems

Anja Zöller, Franz Rothlauf, Torsten O. Paulussen, Armin Heinzl
Universität Mannheim, Lehrstuhl für ABWL und Wirtschaftsinformatik, {zoeller | rothlauf | paulussen | heinzl}@uni-mannheim.de

Abstract. This chapter introduces benchmarking as a special form of evaluation and addresses problems and demands concerning the benchmarking of multiagent systems. It gives an overview of the evaluation concepts used in the German research program SPP 1083 for intelligent agents and realistic commercial application scenarios as well as examples for evaluation and benchmark studies for multiagent systems. The article provides basics for setting-up evaluation studies, regarding special concerns for the evaluation of multiagent systems. Moreover, the exemplary overview may serve as orientation for further evaluation and benchmarking of multiagent systems in realistic and commercial application scenarios.

1.1 Introduction

Before developers and users may utilize a new technology, such as multiagent systems, several questions need to be answered. For example: Under what circumstances is the use of a multiagent system be appropriate or superior to other systems? And, how can users adapt the system to their needs in the most efficient way? To answer these questions and to be able to make adequate decisions, it is necessary to execute evaluation studies and to provide the relevant information regarding the properties of multiagent systems.

Therefore, the aim of this article is to introduce the purpose of evaluation and the basic concepts for setting up evaluation studies. An important comparative form of evaluation is benchmarking. When setting up evaluation studies, special problems and demands concerning the benchmarking of multiagent systems have to be taken into account. Moreover, the concepts of these studies have to be adjusted to the field of application that is considered. Concerning realistic commercial application scenarios, different evaluation concepts have been developed within the German research program for intelligent agents and realistic commercial application scenarios, which provide an orientation for setting up respective evaluation and benchmark studies.
The chapter is structured as follows: the first part presents different aspects of evaluation and a basic model of evaluation. In the second part benchmarking is discussed in detail, addressing special issues concerning the assessment of multiagent systems. Finally, an overview of the evaluation concepts used in the German research program for intelligent agents and realistic commercial application scenarios is presented.

1.2 Evaluation and Benchmarking

The aim of evaluation is to show the applicability of an approach under certain constraints and to deliver decision support for choosing the best approach for a certain problem. One method to assess these questions is benchmarking, providing a comparative evaluation of different approaches. Depending on the questions under study, any kind of evaluation or benchmarking may have different forms and aspects [MaGr1993], as will be discussed in the next section.

However, regarding the evaluation of information systems in general, all these forms and aspects base on a collective model of evaluation which will be explained in Section 1.2.2.

1.2.1 Important Aspects of Evaluation

1.2.1.1 Point of Reference

Depending on the reference used, evaluation may be descriptive or comparative. Within descriptive evaluation, there is no other concrete system one refers to or compares with. Therefore, relevant performance measures are documented as information that may be used for comparison and evaluation of individual needs for a certain application. Descriptive evaluation is especially important in cases with a lack of adequate reference systems and may serve as reference for future comparative evaluation. Comparative evaluation may also be called benchmarking [HeHR 2004] [RePo1997]. It has a concrete reference system or value to compare with and is especially important when deciding between different approaches or assessing the applicability of two systems. The theoretic background of benchmarking will be discussed in more detail in Section 1.3 of this chapter.