A DECISION-THEORETIC TREATMENT OF IMPRECISE COMPUTATION

John Yen and Swaminathan Natarajan*

Department of Computer Science, Texas A&amp;M University, College Station, Texas 77843-3112, USA

* Xerox Corporation, Henrietta, New York, USA

ABSTRACT

Imprecise computation has been suggested as a promising model of real-time computing in order to deal with timing constraints imposed by the environment. However, the theoretical foundation of the technique has not been fully explored. To address this, we propose a decision-theoretic foundation of imprecise computation. The main benefit of such a treatment is that it enables the qualitative assumptions underlying imprecise computation techniques to be explicitly stated in a formal way. The theoretical foundation laid out in this paper, hence, will not only enable the justification of using imprecise computation techniques for a real-time application, but will also facilitate the development of extended techniques for more complex real-time systems.

1 INTRODUCTION

A real-time system is a computer system that operates under externally imposed resource constraints which must be satisfied if correct operation is to be achieved. The resource constraints imposed by the environment raises requirements and issues that are not usually addressed in traditional systems. In particular, neither producing correct result nor being "fast" is in itself sufficient to be real-time. Real-time systems must be able to perform their computations while satisfying resource constraints in a predictable fashion.

1While time is clearly the crucial resource in a real-time system, it is also important to consider the needs of other resources such as memory and network bandwidth, since a shortage
The real-time system we build to solve these problems must meet the timing requirements inherent in the problem. These timing requirements may include either hard or soft deadlines. For instance, picking up an object off a conveyor belt may involve a hard deadline if the object may fall off the belt and be damaged if not picked up in time. On the other hand, a monitoring system may have a soft deadline - the earlier the fault is detected, the less damage it will cause.

An approach that has been proposed for obtaining the desired predictability is the notion of imprecise computation [20, 22] (a similar notion, anytime algorithms [6], has also been developed in the real-time AI community). The idea of imprecise computation is that if we do not have sufficient resources to complete a perfect computation, we perform only a part of it and produce some approximate results. The focus is then on maximizing the quality of this partial result and ensuring that it meets certain acceptability criteria. Work in this area has been in the general context of real-time systems, and on designing algorithms which can produce useful partial results.

The main objective of this paper is to explore the theoretical foundation of imprecision computation. Investigating the theoretical foundation of imprecision computation is important for two reasons. First, it can improve our understanding about the applicability of the technique. Second, it can help us to identify potential extensions to the technique.

In this paper, we first review the basics of imprecise computation. We then present a decision-theoretic interpretation of the technique. Based on such a theoretical foundation, we discuss imprecise computation techniques applied to hierarchical task structures in AI systems. Issues raised by other extensions to imprecise computation are also discussed based on the decision-theoretic framework. Finally, we compare imprecise computation with other decision-theoretic approaches to resource-bounded computing and summarize its benefits.

of any of these may affect the time needed. Therefore we refer to general computational resources rather than time specifically.