Intelligent Query Answering

Zbigniew W. Ras
University of North Carolina, Charlotte, USA

Agnieszka Dardzinska
Bialystok Technical University, Poland

INTRODUCTION

One way to make Query Answering System (QAS) intelligent is to assume a hierarchical structure of its attributes. Such systems have been investigated by (Cuppens & Demolombe, 1988), (Gal & Minker, 1988), (Gaasterland et al., 1992) and they are called cooperative. Any attribute value listed in a query, submitted to cooperative QAS, is seen as a node of the tree representing that attribute. If QAS retrieves no objects supporting query q, from a queried information system S, then any attribute value listed in q can be generalized and the same the number of objects supporting q in S can increase. In cooperative systems, these generalizations are controlled either by users (Gal & Minker, 1988), or by knowledge discovery techniques (Muslea, 2004).

If QAS for S collaborates and exchanges knowledge with other systems, then it is also called intelligent. In papers (Ras & Dardzinska, 2004, 2006), a guided process of rules extraction and their goal-oriented exchange among systems is proposed. These rules define foreign attribute values for S and they are used to construct new attributes and/or impute null or hidden values of attributes in S. By enlarging the set of attributes from which queries for S can be built and by reducing the incompleteness of S, we not only enlarge the set of queries which QAS can successfully handle but also we increase the overall number of retrieved objects.

So, QAS based on knowledge discovery has two classical scenarios which need to be considered:

- **System is standalone and incomplete.**

Classification rules are extracted and used to predict what values should replace null values before any query is answered.

- **System is distributed with autonomous sites (including site S). User needs to retrieve objects from S satisfying query q containing nonlocal attributes for S.**

We search for definitions of these non-local attributes at remote sites for S and use them to approximate q (Ras & Zytkow, 2000), (Ras & Dardzinska, 2004, 2006).

The goal of this article is to provide foundations and basic results for knowledge-discovery based QAS.

BACKGROUND

Modern query answering systems area of research is related to enhancements of query-answering systems into intelligent systems. The emphasis is on problems in users posing queries and systems producing answers. This becomes more and more relevant as the amount of information available from local or distributed information sources increases. We need systems not only easy to use but also intelligent in handling the users' needs. A query-answering system often replaces human with expertise in the domain of interest, thus it is important, from the user's point of view, to compare the system and the human expert as alternative means for accessing information.

A knowledge system is defined as an information system S coupled with a knowledge base KB which is simplified in (Ras & Zytkow, 2000), (Ras & Dardzinska, 2004, 2006) to a set of rules treated as definitions of attribute values. If information system is distributed with autonomous sites, these rules can be extracted either locally from S (query was submitted to S) or from its remote sites. The initial alphabet of QAS associated with S contains all values of attributes in *S*, called local, and all decision values used in rules from KB. When KB is updated (new rules are added or some deleted), the alphabet for the local query answering system is automatically changed. It is often assumed that knowledge bases for all sites are initially empty. Collaborative information system (Ras & Dardzinska, 2004, 2006) learns rules describing values of incomplete attributes and attributes classified as foreign for its site called a client. These rules can be extracted at any site but their condition part should use, if possible, only terms which can be processed by the query answering system associated with the client. When the time progresses more and more rules can be added to the local knowledge base which means that some attribute values (decision parts of rules) foreign for the client are also added to its local alphabet. The choice of which site should be contacted first, in search for definitions of foreign attribute values, is mainly based on the number of attribute values common for the client and server sites. The solution to this problem is given in (Ras & Dardzinska, 2006).

MAIN THRUST

The technology dimension will be explored to help clarify the meaning of intelligent query answering based on knowledge discovery and chase.

Intelligent Query Answering for Standalone Information System

QAS for an information system is concerned with identifying all objects in the system satisfying a given description. For example an information system might contain information about students in a class and classify them using four attributes of "hair color", "eye color", "gender" and "size". A simple query might be to find all students with brown hair and blue eyes. When information system is incomplete, students having brown hair and unknown eye color can be handled by either including or excluding them from the answer to the query. In the first case we talk about optimistic approach to query evaluation while in the second case we talk about pessimistic approach. Another option to handle such a query would be to discover rules for eye color in terms of the attributes hair color, gender, and size. These rules could then be applied to students with unknown eye color to generate values that could be used in answering the query. Consider that in our example one of the generated rules said:

(hair, brown) $\wedge$ (size, medium) $\rightarrow$ (eye,brown).

Thus, if one of the students having brown hair and medium size has no value for eye color, then the query answering system should not include this student in the list of students with brown hair and blue eyes. Attributes hair color and size are classification attributes and eye color is the decision attribute.

We are also interested in how to use this strategy to build intelligent QAS for incomplete information systems. If query is submitted to information system S, the first step of QAS is to make S as complete as possible. The approach proposed in (Dardzinska & Ras, 2005) is to use not only functional dependencies to chase S (Atzeni & DeAntonellis, 1992) but also use rules discovered from a complete subsystem of S to do the chasing.

In the first step, intelligent QAS identifies all incomplete attributes used in a query. An attribute is incomplete in S if there is an object in S with incomplete information on this attribute. The values of all incomplete attributes are treated as concepts to be learned (in a form of rules) from S.

Incomplete information in S is replaced by new data provided by Chase algorithm based on these rules. When the process of removing incomplete vales in the local information system is completed, QAS finds the answer to query in a usual way.

Intelligent Query Answering for Distributed Autonomous Information Systems

Semantic inconsistencies are due to different interpretations of attributes and their values among sites (for instance one site can interpret the concept "young" differently than other sites). Different interpretations are also due to the way each site is handling null values. Null value replacement by values suggested either by statistical or knowledge discovery methods is quite common before user query is processed by QAS.

Ontology (Guarino, 1998), (Van Heijst et al., 1997) is a set of terms of a particular information domain and the relationships among them. Currently, there is a great deal of interest in the development of ontologies to facilitate knowledge sharing among information systems.

Ontologies and inter-ontology relationships between them are created by experts in corresponding domain,