KDMAS: A Multi-Agent System for Knowledge Discovery via Planning

Li Jin and Keith Decker

Computer Science and Information Sciences Department
University of Delaware, Newark, DE 19716
{jin, decker}@cis.udel.edu

In the real world, there are some domain knowledge discovery problems that can be formulated into knowledge-based planning problems, such as chemical reaction process and biological pathway discovery problems (Khan et al. 2003). A view of these domain problems can be re-cast as a planning problem, such that initial and final states are known and processes can be captured as abstract operators that modify the environment. For example, approaching biological pathway discovery with an AI planning approach would mean that a valid plan that transfers the initial state into the goal state is a hypothetical pathway that prescribes the order of events that must occur to effect the goal state. We believe that AI planning technology can provide a modeling formalism for this task such that hypotheses can be generated, tested, queried and qualitatively simulated to improve the domain knowledge and rules.

Our current approach is to build toward a general multi-agent system for knowledge discovery (KDMAS) via planning. The plans produced are hypotheses capturing relevant qualitative information regarding domain knowledge. We will use the biological pathway domain as an application to present our approach, focusing on acquiring planning knowledge and hypothesis generation. As shown in Figure 1, KDMAS integrates:

- The domain knowledge, which is built up with the domain experts' aid, and can be translated into the planning formalism automatically by the planning formalism agent.
- Other domain knowledge from data that is retrieved from heterogeneous sources including human experts by information gathering multi agents and used for hypothesis generation.
- Computer interpretable plans capturing relevant qualitative information regarding the domain problems such as signal transduction pathways.
- Testable hypotheses regarding gaps in knowledge of the domain such as a biological pathway to drive future research.

- Confidence measures to calculate the confidence for the hypotheses based on the supporting data.
- Visualization tools.

Though the properties of general planning and knowledge-based planning have been well investigated (Weld 1999; Bonet & Geffner 2000), the hardest challenge we face is to extract the domain knowledge and its subsequent representation into the planning formalism. Our approach is to start with existing Semantic Web approaches based on the Resource Description Framework (RDF), for example the BioPax representation for biological pathway data (BioPAX 2005) which is implemented in OWL (W3C 2004). A general planning representation can be extracted from BioPAX and then the domain planning formalism can be generated from the domain Knowledge Base (KB) by modifying the general formalism with the domain experts' aid.

The second challenge we face is that the local domain knowledge base, created from semantic web resources, does not have complete information. Much relevant information is known to exist somewhere else. In this case, the planning engine engages a traditional multi-agent system that is responsible for gathering data from outside data sources. Figure 1 shows that the planner agent publishes its queries to a blackboard for more information. The manager agent collects queries from the blackboard and finds some agents responsible for answering the queries. In the pathway domain, larger pathways can be built by combining the individual pathways of parameterized modules.

We have initially focused on hierarchical task network (HTN) planning (Erol et al. 1994) because of the hierarchical nature of the underlying pathway information, and the hierarchical nature of the underlying biological process hypotheses. The third challenge of our approach is that many hypothetical plans can be generated for the same initial and final states while most of the plans do not have real-world counterparts. Therefore a hypothetical plan has to be ascribed a confidence measure based on the supporting data. Differential diagnosis (Hammond 1990; Roos 1993) can be used to propose laboratory experiments that differentiate between top hypotheses.

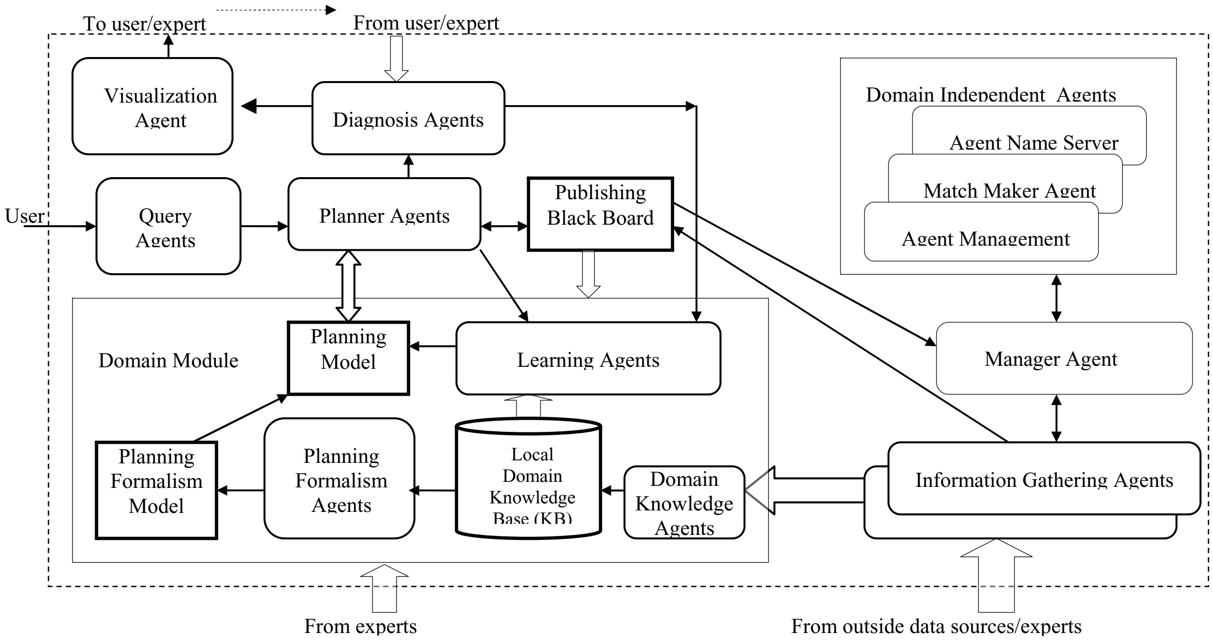

Figure1. Knowledge-based planning multi-agent system

The visualization of the diagnosis and the selected hypothetical plans is also a challenge in our approach. Besides a general semantic web visualization tool for the knowledge base, some domain specific visualization tools should be embedded into visualization module.

Real-world planning problems call for domain knowledge-based planning (Wilkins & DesJardins 2001). Realistic domain knowledge-based planners also have to deal with plan explanation and repairing plan failures. Because our approach is built on an OWL domain ontology, the integrated ontology set and reasoning mechanism can be used to support simple learning mechanisms. Large numbers of planning experiences can be used to predict plan failure and to modify the planning domain model, such as to create and modify task decomposition rules and to build macro actions to improve planning search efficiency. So any plan failure will lead to planning module modification or knowledge discovery by hypothesis generation. Realistic domains (e.g. biology) may have dozens of parallel actions, so multi-agent planning will also have to be considered. The multi-agent system in Figure 1 will use Biomas (Decker et al. 2002) for information retrieval. In additional to the domain-independent agents, some agents have to be developed specifically for the domain task. The planner agent is created to wrap the planning system and publish its queries to blackboard for more information. The manager agent is responsible for finding some agent to gather more information for the planner. Diagnosis agents analyze the failure plans and successful plans by comparing the hypotheses with the real-world counterparts. And learning agents modify the planning model by diagnosis.

Acknowledgements

The authors wish to thank Dr. Carl J. Schmidt and Dr. Karl V. Steiner for various discussions and many valuable suggestions.

References

BioPAX: Biological Pathways Exchange Language Level 2, version 1.0, 2005, http://www.biopax.org

Bonet, B. & Geffner, H. 2000. Planning with incomplete information as heuristic search in belief space. In *Proc. of the Intl. Conf. on Aut. Planning & Scheduling*. 52–61, Breckenridge, CO: AAAI press

Decker, K., Khan, S., Schmidt, C., Situ, G. Makkena, R., and Michaud, D. 2002. Biomas: A multi-agent system for genomic annotation. *Intl. J. of Coop. Info. Sys.* 11:265-292

Erol, K., Nau, D., and Hendler, J. 1994. HTN Planning: Complexity and Expressivity. In *Proc. AAAI-94*, 1123-1128. Seattle, WA: AAAI press

Hammond, K. 1990. Explaining and Repairing Plans That Fail. *Artificial. Inteligence* 45(1-2): 173-228

Khan, S., Gillis, W., Schmidt, C., and Decker, K. 2003. A Multi-Agent System-driven AI Planning Approach to Biological Pathway Discovery, In *Proc. of Intl. Conf. on Automated Planning & Scheduling*. 246-255, Trento, Italy: AAAI press

Roos. N. 1993. Efficient model-based diagnosis. *Intelligent System Engineering*, pages 107–118

W3C. 2004. OWL: OWL Web Ontology Language. http://www.w3.org/TR/owl-features/

Weld, D. 1999. Recent advances in AI planning. *AI Magazine* 20(2):93-123.

Wilkins D. and DesJardins M. 2001. A call for knowledge-based planning, *AI Magazine*, vol. 22, no. 1