

다중 에이전트 기반 지식 탐사 및 문제 해결 프레임워크

강성희, 박승수
이화여자대학교 컴퓨터학과

Multi-Agent Knowledge Discovery and Problem Solving Framework

Sung Hee Kang, Seung Soo Park
Dept. of Computer Science and Engineering in Ewha Womans University

요 약

Decentralized 정보는 여러 도메인에 대한 heterogeneous한 독립적인 정보가 자율적으로 존재하며 이들 정보 간의 관계성을 고려한 전체에 대한 global view가 존재하지 않기 때문에 inter-domain에 대한 마이닝을 수행하는데 어려움이 있다.

본 연구에서는 intra-domain knowledge discovery, intra 및 inter-domain problem solving method라는 접근 방법으로, decentralized 데이터 환경에서 문제 해결에 필요한 정보 추출을 위한 데이터 tailoring과 분산 데이터에 대한 목표-지향 데이터마이닝(goal-oriented data-mining)을 통해 문제 해결을 위해 필요한 지식을 생성하고 이들 간의 관련 정보를 탐색하여 문제를 해결하는 프레임워크를 제안한다. 특히, 생성된 지식간의 협동 문제 처리를 위해서 멀티 에이전트 패러다임을 이용하기로 한다.

제안 프레임워크는 산재되어 있는 데이터들로부터 문제 해결에 필요한 유용한 지식 차원의 정보를 추출해내고 생성된 지식을 바탕으로 각 도메인 정보에 대한 개별적인 사용뿐 만 아니라 서로 cooperation을 통한 문제 해결을 지원함으로써, 개방된 분산 환경하에 decentralized 되어 있는 여러 도메인 정보를 보다 효율적으로 활용할 수 있는 새로운 형태의 문제 해결 방법이라고 할 수 있다.

1. 서론

컴퓨팅 및 네트워크 기술의 발달은 정보 형태의 변화를 야기하였으며 정보를 처리하는 새로운 컴퓨팅 기술을 요구하고 있다. 미래의 정보는 고도의 네트워크를 통해 분산(decentralized)되어 있는 대량(enormous)의 데이터들로 구성될 수 있고, 이질적(heterogeneous)이며 동적(dynamic)으로 변화하는 개방화(open)된 특성을 가지고 있다. 이러한 특성을 갖는 미래의 정보 환경에서 문제 해결에 유용한 지식을 발견하는 데이터마이닝(data mining)[2,6] 기술은 필수적인 기술일 것이다.

일반적으로 분산 정보에 대한 지식 발견의 대부분의 경우에는 지식 발견을 위한 관련된 모든 정보를 하나의 단일 데이터베이스로 모으고 이들 모아진 데이터(datawarehouse)에 대해 마이닝을 적용하는 방법을 사용한다. 그러나 이러한 방법은 변화되는 정보 특성을 반영하지 못하며 실제적으로 비현실적이거나 불가능한 여러 문제점을 가지고 있다. 분산 데이터마이닝(distributed data mining)[5]은 변화되는 정보의 특성에 대응하여 분산된 대용량의 정보로부터 새로운 지식을 발견할 수 있는 방법에 대한 관심에서

발생되었다고 할 수 있다. 분산 데이터마이닝은 분산 컴퓨팅 환경하에 있는 분산된 데이터베이스에서 일정한 패턴을 찾아내는 것으로 meta learning[1], stacking, collective data mining[4], distributed association rule learning 등의 대표적인 접근 방법이 있다. 그러나 기존의 분산 데이터마이닝 연구는 대부분이 수평 분할 데이터에 적용되거나 data reduction을 이용한 수직 분할 데이터에 적용되는 것에 국한되어 있기 때문에 inherently하게 decentralized되어 있는 다양한 특성을 갖는 정보 형태에 적용하기 어렵다는 제한점들을 가지고 있다.

decentralized 데이터 환경에서 지식 발견의 문제점은 대상 데이터가 여러 도메인에 대한 heterogeneous하고 독립적이며 자율적인 정보로서 존재하여 이들 정보간의 관계성을 고려한 전체에 대한 global view가 존재하지 않기 때문에 inter-domain에 대한 마이닝과 협동 작업(cooperative job)을 자동적으로 수행하기 어렵다는 점에서 발생한다.

본 연구에서는 미래형 정보로서 개방화된 동적인 분산 환경하에 독립적으로 구성되어지는 decentralized된 여러 다양한 정보들을 보다 효율적으로 문제 해결에 활용하기 위해서 산재되어 있는 다양한 데이터들로부터 문제 해결에 필요한 유용한 지식 차원의 정보를 추출해내고 이들 독립적인 분산 정보를 융통성 있게 문제 해결에 적용하는 방법을 제안한다. 다각적으로 변화되는 분산된 다양한 정보로부터 유용한 지식을 생성하고 생성된 지식을 바탕으로 각 도메

* 본 논문은 KISTEP(98-NT-01-01-A-12)의 지원에 의해 수행되었음.

인 정보에 대한 개별적인 사용뿐 만 아니라 서로 협동을 통한 문제 해결을 지원할 수 있도록 intra-domain knowledge discovery, intra 및 inter-domain problem solving method라는 접근 방법을 사용한다. 다양한 데이터 풀에서 문제의 요구에 적당한 데이터베이스를 선택(데이터 tailoring)하고 이들 분산 데이터에 대해 목표-지향 decentralized 데이터마이닝을 통해 문제 해결을 위해 필요한 지식을 생성하고 이들 분산 지식간의 관련 정보(관련 attributes 사이의 chaining)를 탐색하여 분산 지식 간의 협동을 통한 문제를 해결하는 새로운 형태의 문제 해결 프레임워크를 제안한다. 이러한 과정에 다중 에이전트(multi-agent)[3,7] 패러다임을 적용하여 기계 생성 지식으로 구성되는 지식 기반 에이전트들이 서로 협동하여 inter-domain 간의 문제를 유연하게 해결할 수 있도록 한다.

2. 지식 모델

본 연구에서는 기계가 생성해 낸 분산된 지식을 관리하는 방법이 요구된다. 기계 생성 지식을 포함하여 일반적인 지식 관리를 위한 모델은 여러 가지 관점에서 구분되어 질 수 있다.

지식의 분산 여부에 따라 지식 관리를 두 가지 형태의 모델로 나누어 볼 수 있다. 1) knowledge-centralized model, 2) knowledge-decentralized(distributed) model.

분산 모델은 각 분산 지식의 통합 관리의 자치성 여부에 따라 1) control-centralized 모델, 2) control-decentralized 모델로 구분할 수 있고 이들을 두 가지가 복합된 형태를 취할 수도 있다. control-centralized model에서는 분산 지식을 관리하기 위한 central resource agent가 있어서 구성 분산 지식간의 상호작용을 담당한다.

분산 모델을 구성하는 지식 베이스의 특성에 따라 1) 협동 모델(cooperative model), 2) 경쟁 모델(competitive model), 3) 혼합 모델(hybrid model)로 나눌 수 있다. 이러한 협동, 경쟁 모델은 분산 지식 간의 inter-dependency에 따른 cooperation type(그림 1)에 의해 결정된다.

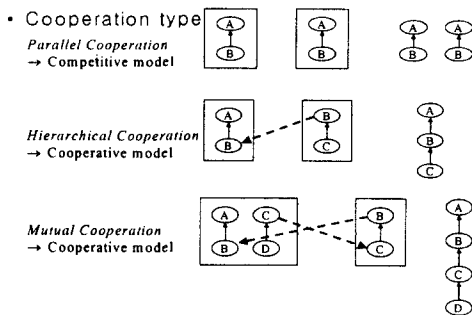


그림 2 분산 지식간의 cooperation type

본 연구에서는 knowledge-decentralized, control-centralized이며 hybrid한 knowledge model에서 데이터마이닝을 통해 생성된 지식을 문제 해결에 이용하고 에이전트 패러다임을 적용하여 다양하고 융통성있는 문제해결을 시도하는 프레임워크를 제안하는 것을 목적으로 한다.

3. 지식 탐사 및 문제 해결 프레임워크의 기본 설계

제안 프레임워크는 개방화되고 동적으로 변화하는 환경에서 다양한 도메인에 대한 decentralized data로부터 생성된 지식을 이용하여 다양한 도메인이 관련된 inter-domain 문제를 유연하게 해결할 수 있는 방법을 제공한다. 대상 데이터 환경에서는 데이터로부터 새로운 지식을 발견해내는 forward한 접근 방식의 기존 데이터마이닝을 그대로 적용하는 것이 어렵기 때문에 본 프레임워크는 목표-지향 데이터마이닝(goal-oriented data mining)을 이용한 backward한 접근 방식을 사용하기로 한다. 이것은 문제가 주어질 때, 이 문제를 해결하기 위한 데이터를 선택하고 선택된 데이터로부터 마이닝을 통해 문제 해결에 필요한 지식을 발견하고 이를 발견된 지식을 문제 해결에 이용하는 방법이다. 그림 2는 목표-지향 데이터마이닝을 적용한 지식 탐사 및 문제 해결 과정에 대한 전체적인 개요를 나타낸다.

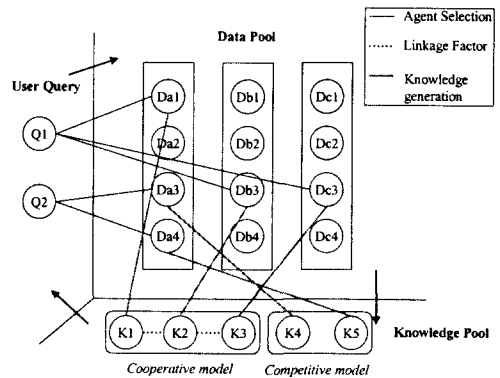


그림 3 지식 탐사 및 문제 해결 프레임워크

본 연구에서는 다양한 도메인 생성 지식에 대한 여러 경우의 유연한 문제 해결을 가능하게 하도록 에이전트 패러다임을 적용한다. 가상 시스템(그림 3)은 로컬한 데이터베이스로부터 마이닝을 수행하고 결과로서 생성된 지식을 도메인 지식으로 갖는 지식 베이스 기반 시스템으로서의 역할을 하는 에이전트들과 문제 해결을 위한 데이터베이스를 선택하고 생성된 지식 간의 linkage factor를 탐색하며 문제 해결에 필요한 에이전트들 간의 수행을 조절하는 등 복잡한 마이닝을 통한 문제 해결 과정을 통제하고 자동화할 수 있는 조정자 레벨의 에이전트인 facilitator(그림 4)로 구성된다.

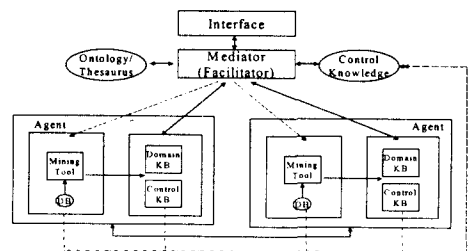


그림 4 전체 시스템 구성

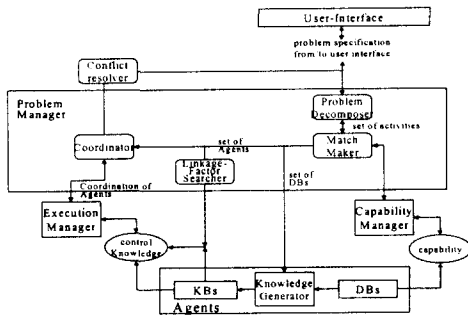


그림 5 facilitator의 구조

4. 프레임워크의 세부 과정

전체 프로세싱 과정은 질의에 필요한 관련 지식을 발견하는 과정과 해당 질의에 대한 문제 해결 과정으로 그림 5와 같이 진행된다.

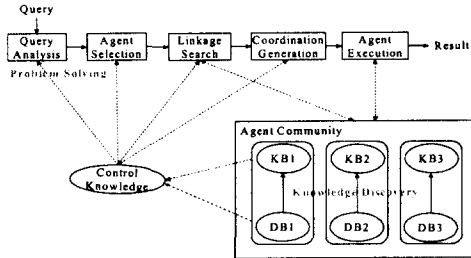


그림 6 전체 프로세싱 과정

4.1 문제 관련 데이터베이스 선택 문제

전체 데이터 풀(data pool)은 다양한 schema를 갖는 분산된 데이터베이스로 구성되며, 대상 데이터 형태는 관계형 데이터베이스(relational database)라고 가정한다. 사용자의 질의(query)가 주어지면 우선 이를 해결하기 위해 필요한 지식을 생성해낼 수 있는 데이터베이스를 선택하는 것이 필요한데 질의 분석에 의한 known, unknown attributes에 대해 이들을 cover할 수 있는 coverable databases를 결정한다.

4.2 데이터마이닝을 통한 문제 관련 지식 생성

질의에 대해 선택된 각 로컬 데이터베이스들에 대해 parallel하게 데이터마이닝을 통한 지식을 생성한다. 사용되는 마이닝 알고리즘(mining algorithm)은 규칙 연역법(rule induction)이나 결정 트리(decision tree)와 같은 기계 학습(machine learning) 기법을 사용하며 생성된 지식의 형태는 명제식 규칙(propositional statement)을 사용하는 CNF(Conjunctive Normal Form)이다. 각 로컬 마이닝은 질의에서의 known attributes가 생성 규칙의 LHS에 포함되고 unknown attribute가 생성 규칙의 RHS에 존재하는 모든 규칙을 찾게 되며 발견된 지식들이 문제 해결을 위해 사용될 수 있는 부분 지식으로서 역할을 하게 된다.

4.3 분산 지식 간의 협동을 위한 연결 정보(linkage factor) 탐색

분산된 마이닝 결과 지식이 inter-domain 문제 해결을 위해 협동하려면 이들을 연결해 줄 공통의 정보(linkage factor)가 존재해야 하고 이를 찾아내는 것이 필요하다. 분산된 규칙 간의 rule chaining을 탐색하여 전체 질의에 대한 global한 rule inference graph를 생성한다.

4.4 에이전트 간의 수행 조정(coordination)

의존적인 분산 지식 베이스간의 결합을 통한 문제 해결을 유연하게 제공하기 위해서 에이전트 간의 협동을 통한 문제 해결이 필요하며 구성 에이전트 간 수행을 조정하는 coordination 과정이 요구된다. 에이전트 간 수행 조정은 rule inference graph를 agent coordination graph로 변경하여 질의에 대한 결과를 생성하기 위해 어떠한 제한조건(constraints)을 만족하는 에이전트 간 수행 순서 및 방법을 결정하는 과정이다.

5. 결론 및 향후 연구과제

본 연구에서는 대상 데이터가 inherently 분산되어 있는 데이터들로서 전체 구성 데이터들에 대한 data view가 없는 환경에서 다양한 도메인이 연루된 문제를 해결하기 위한 방법으로 분산 데이터 마이닝을 통해 로컬한 지식을 발견하고 이들 지식의 cooperative한 작업을 수행할 수 있도록 다중 에이전트 패러다임을 적용하는 지식 탐색을 이용하는 문제 해결 프레임워크를 제안하였다. 각 부분 생성 지식을 서로 연결할 수 있는 linkage factor를 탐색하여 문제를 해결할 수 있는 global한 rule chaining을 찾아내고 분산 지식 간의 문제 해결을 각 도메인 지식을 갖는 에이전트 간의 협동을 통해서 수행한다. 이는 decentralized 환경에서 지식 발견을 통한 새로운 형태의 문제 해결 방법을 제공한다.

또한 이러한 연구 결과는 다른 다중 에이전트의 응용에 활용될 뿐만 아니라 WWW contents에 대한 지식 발견(knowledge discovery)에 사용되어질 수 있을 것이라 기대된다.

제안 방법은 분산 마이닝과 다중 에이전트 패러다임을 병합함으로써 각 local에서의 autonomy를 보장하면서 parallelism을 통한 성능 향상과 modularity를 통한 extensibility, flexibility, robustness, 그리고 reusability 등의 장점을 가진다.

6. 참고문헌(References)

- [1] Chan, P., & Stolfo, S., Toward parallel and distributed learning by meta-learning, In *In Working Notes AAAI Work...*, Knowledge Discovery in Databases, AAAI, 1993
- [2] Fayyad, Usama M., et al., *Advances in Knowledge Discovery and Data Mining*, AAAI Press, 1996
- [3] Jennings, N. R., Sycara, K., et al., "A Roadmap of Agent Research and Development", *Autonomous Agents and Multi-Agent Systems*, Vol. 1, P275-306, 1998
- [4] Kargupta, H., Park, B., et al., *Collective data mining: a new perspective toward distributed data mining*, Technical Report, Washington State University
- [5] Kargupta, H., *Research On Distributed Data Mining*, URL http://www.eecs.wsu.edu/~hillol/ddm_research.html
- [6] The Data Mine, <http://www.cs.bham.ac.uk/~anp/TheDataMine.html>
- [7] Wooldridge, M., Jennings, N. R., "Intelligent Agents: Theory and Practice", *Knowledge Engineering Review*, Vol 10, No. 2, pp115-152, 1995