

복수결론을 유도하는 지식획득이론

강병호, 박덕진
호서대학교 컴퓨터학부 전자계산학전공

Multiple Classification Ripple Down Rules

ByungHo Kang, DuckJin Park
Hoseo University, Computer Science

요 약

Ripple Down Rules(RDR)이론은 지식베이스시스템을 지식공학구축기술 또는 지식공학자의 도움 없이 특수분야전문가에 의해 효율적으로 유지보수, 구축되어진다. 특히 시스템의 운용 중 지식베이스의 수정을 효율적으로 처리할 수 있다 본 논문은 단일결론을 생성하는 RDR이론의 확장인 복수(複數)결론(multiple classification)을 유도하는 MCRDR이론에 대하여 설명한다 MCRDR은 복잡한 복수결론을 허락하면서 RDR이론의 최대 강건인 지식베이스의 간편한 유지보수 기능을 유지한다 MCRDR의 KA과정, 기초케이스 문제해결방법, 그리고 복수결론 추론문제에 대하여 논할 것이다 MCRDR시스템의 우수성을 모의전문가를 이용한 시스템 구축과 실험으로 증명해 보일 것이다. 이 실험을 통하여 복수결론을 지원하는 MCRDR이론이 단일결론을 지원하는 RDR이론을 통하여 효율적으로 증명하고, 또한 기존의 기계학습방법과의 차이점도 보여줄 것이다

1. 기존의 지식획득 방법과 지식베이스 유지보수의 문제점

기존의 전문가시스템에는 두 가지 문제점이 존재해 왔다. 첫째, 개발과정에서 KA(Knowledge Acquisition)작업을 수행하기 위하여 전문가로부터 전문지식을 파악하여 시스템에서 사용할 수 있는 지식으로 변환작업을 수행하는 지식공학자가 필수적이거나, 개발 팀의 누군가가 이 지식공학자의 역할을 수행해야 했다[1]. 그러나 소프트웨어의 개발과 특정분야의 전문지식을 고루 갖춘 지식공학자(knowledge engineer)를 확보하기 어렵고, 시스템의 개발분야가 바뀔 때마다 그에 맞는 지식공학자를 구한다는 것은 더욱 어려운 일이다[2]. 둘째, KB(Knowledge Base) 유지보수(maintenance)의 문제로, 대부분의 전문가시스템 개발방법은 시스템개발 초기단계에서 철저한 KA에 중점을 두고 있었으며, 시스템의 운용단계에서 지식의 유지보수역할은 강조하지 않았다. 이러한 접근방식은 전문지식의 빠른 변화에 능동적으로 대처하지 못하는 약점이 있다. 실질적으로 시스템의 사용기간이 길어질수록 KB의 크기는 기하급수적으로 증가하며, 지식의 상호연결 관계는 더욱 심화되기 마련이다[3].

위에서 열거한 이유들로 인하여, KA병목현상에 관한 연구에서는 최근 전문지식에 대한 인지학(認知學) 측면에서 새로운 해석이 시도되고 있다. 대표적인 것으로, 전문지식은 오랜 경험을 통하여 축적된다는 사실에 착안한, Case Based

Reasoning (CBR)[4]과 Bill Clancey등에 의하여 제창되고 있는 situated cognition[2]등이 그러한 예이다.

본 연구에서는 위의 두 가지 단점을 보완한 Garvan-ES1의 개발과 운영단계에서 밝혀진 전문가들의 지식행위 과정에 대한 분석을 바탕으로 정립한 Ripple Down Rules(RDR)이론을 토대로 새로운 이론을 소개하려 한다.

2. Ripple Down Rules (RDR)

기존의 KA방법은 직접적으로 지식공학자의 도움을 필요로 하기 때문에 많은 제약과 비용, 시간상의 어려움이 많았다. 또, 새로운 지식을 획득하려면 여러 단계를 거쳐 기존의 KB에 추가하고 다시 컴파일 해야 실질적인 KA가 이루어지는 것이다 이는 하나의 지식을 생성하기 위해 많은 비용과 시간이 소요되는 비효율성을 가지고 있다. 그러나, 기존의 KA방법과는 달리 RDR 시스템은 하나의 새로운 지식을 학습시킬 필요가 있을 때, 지식공학자를 배제하고 전문가가 새로운 지식을 KB에 직접 실시간으로 입력함으로써 지식의 변화에 빠르게 대응함으로써 많은 시간을 단축하고, 삽입된 새로운 지식은 기존의 KB에 단지 새로운 rule(지식)을 적당한 자리에 추가하여 KA가 쉽게 이루어진다. 또 KB에 지식을 입력하는 전문가와 KB가 같은 공간대를 유지하기 때문에 전문가는 시스템의 오류(잘못된 판단의 이유)를 쉽게 판단할 수 있는 장

점 또한 가지고 있다[5] 그러나 RDR이론의 위와 같은 장점에도 불구하고 한계를 가지고 있다.

3. RDR의 한계

현재 상용화되고 있는 RDR의 한가지 문제점은 오직 하나의 결론을 만들어 내고 조작하는 것이다[2]. 예를 들어 환자가 여러 개의 독립적인 병을 가지고 있다고 하자. 현재 RDR에 의해서 구현된 PEIRS는 이러한 경우에 여러 가지 질병을 묶어 새로운 한가지의 질병으로 분류한다[6]. 그러나 이는 KA작업량을 기하급수적으로 증가시킬 수 있고, KB의 내부구조가 비효율적이면서 상당한 양의 지식이 중복 저장된다. 이러한 문제는 서브도메인의 정확한 선택과 복수결론의 사용으로 해결할 수도 있다. 그러나, RDR만으로 이를 해결하려면 도메인을 나누고, 나눈 각 도메인마다 독립적인 KBS (Knowledge Based System)를 구축하는 것이다. 그러나 도메인을 많은 하위도메인으로 나누기가 쉽지 않고, 작업전반에 걸쳐 많은 문제점이 발생할 수도 있다. 본 논문의 의도는 단일결론의 단점을 보완하여 복수결론을 다루는 방법을 개발하는데 있다.

4. Multiple Classification Ripple Down Rules(MCRDR)

MCRDR의 목적은 RDR의 기본 전략과 장점을 유지하면서 여러 개의 독립적인 결론을 다루는 것이다. RDR과 같이 MCRDR은 테스트케이스에 대한 결론이 잘못되었다면, 올바른 결론을 유도한 다른 기초(基礎)케이스(cornerstone case)와의 차이를 이용하여 잘못된 결론을 바로잡을 수 있다. 즉, 시스템의 도움으로 rule의 구성요소를 만들 수 있어 시스템의 유지보수가 한결 수월해진다.

4.1 추론(Inference)

MCRDR의 KB구조는 각 노드(node)를 rule로 표현하는 n-ary 트리로 구성되어 있고, KB의 첫 번째 레벨에서 모든 rule들을 평가한다. 그리고 다음 레벨로 이동하여 각 rule들을 평가한다. 이렇게 하위 노드까지 평가해 나간다면 프로세스는 더 이상 평가할 지식 노드가 없거나 전문가에 의해서 입력된 케이스에 의해 만족되는 rule이 더 이상 없을 때 동작을 멈춘다. 이렇게 하여 여러 경로를 가진 복수결론을 도출할 수 있다. 각 경로를 통하여 각각의 복수결론을 도출할 수 있다. 이렇게 평가하여 나온 결론이 전문가의 객관적인 판단으로 결론을 바꿀 필요성을 느낄 때 KA가 발생한다.

4.2 MCRDR의 KA

케이스의 결론이 잘못 됐거나 결론이 존재하지 않을 때 KA가 요구된다. KA는 새 가지 프로세스를 수행한다. 첫째, 시스템은 전문가로부터 올바른 결론을 요구한다. 둘째, 시스템은 새로운 rule의 위치를 결정한다. 셋째, 시스템은 전문가로부터 새로운 rule의 조건을 요구하고, 이 새로운 rule을 KB에 추가한다

① 새로운 결론(classification)의 요구 : 새로운 결론을 얻기 위해서 전문가가 새로운 결론을 진술할 필요성이 있다 예로 만약 주어진 어떤 문제에 대하여 class2, class5, class6을 결

론으로써 시스템이 만들어 낸다면 전문가는 class6을 바꿀 필요가 없고 class2와 class5는 제거되고, class7과 class9는 추가되어야 한다고 결정할 수 있다.

② Rule의 위치지정 : 시스템이 지식베이스의 오류를 수정하기 위하여 새로운 rule을 추가할 경우 추가될 rule의 위치는 표1.과 같은 방법으로 결정된다 여기서 stopping rule이란 일반 rule과 같으나 결론부분이 없는 특별한 rule을 의미한다. 이 stopping rule의 기능은 틀린 결론의 경로를 다른 rule에 영향을 주지 않고 제거하는 것이다. 새로운 rule의 위치지정은 KB의 진화와 유지보수를 용이하게 하는 핵심적인 요소이다 상위 레벨의 rule 추가는 rule의 사용빈도를 높여 도메인 지식의 빠른 확장이 가능할 수 있으나 빠른 만큼 오류율도 높아진다. 반대로 하위레벨 rule 추가는 rule의 사용빈도를 제

틀린 결론(classification)	KB를 수정하기 위하여
틀린 결론을 제거한다.	현재의 경로 끝에 stopping rule을 추가한다
틀린 결론을 새로운 결론으로 대체한다	새로운 rule을 추가하기 위하여 경로의 끝에 rule을 추가
새로운 독립된 결론을 추가한다.	새로운 결론을 추가하기 위하여 최상위 레벨에 rule을 추가

표1 지식베이스에 새로운 rule을 추가하기 위한 3가지 방법 한하는 것이므로 도메인지식의 빠른 확장은 기대할 수 없으나 적용 rule의 오류율은 낮아질 것이다.

③ rule의 조건(condition) : Rule의 위치가 지정되었으면 rule의 조건을 부여해야 한다. rule의 조건을 전문가가 직접 결정해 줄 수도 있으나, 중복된 조건을 피하여 최적의 조건을 부여하기 위해서는 시스템의 도움이 필요하다. 이를 위해 대차(對差)리스트(differenc list)와 기초케이스 개념이 요구된다 대차리스트의 식은 (case A - case B) U (case B - case A)로 표현되고, 기초케이스는 한 개의 rule만을 놓고 생각할 때 해당 rule을 만들게 한 케이스(KA event를 발생하도록 원인 제공을 한 케이스)로 정의할 수 있다. 조건을 구하는 프로세스는 다음과 같다

- ㉠ ②에 의해 지정된 위치에 존재하는 rule의 기초케이스 리스트 중 한 개를 로드한다
- ㉡ 한 개의 기초케이스를 선택하여 테스트케이스와 비교하여 대차리스트를 만든다
- ㉢ 생성한 대차리스트를 가지고 나머지 기초케이스를 만족하는지 테스트하여 만족하는 기초케이스만 분류한다
- ㉣ 기초케이스가 없을 때까지 ㉠~㉢을 반복 실행한다.

이렇게 얻은 최종 대차리스트 중에 전문가가 적당한 조건을 선택하면 이 선택된 조건이 새로운 rule의 조건이 된다 이 테스트케이스는 새로 만들어진 rule의 기초케이스가 되고, 각 경로의 rule도 테스트케이스를 자신의 기초케이스로 기억하고 있어야 한다 이러한 반복작업을 거쳐서 새로운 rule의 조건이 결정되었다. MCRDR은 하나의 rule에 대하여 여러 개의 기초케이스를 가지고 있다 반면에 RDR은 한 rule에 한 개의 기초케이스를 가진다. 이렇게 위의 세 가지 프로세스를

거치면 KA작업은 끝나게 된다

5. MCRDR의 성능실험

우리는 지금까지 MCRDR의 기능과 우수성, KB의 구조, 그리고 KA에 대하여 논하였다. 실제로 인간전문가를 대신한 모의실험전문가(simulated expert)라는 전문가시스템을 사용하여 다른 전문가시스템과 비교하여(Compton, Preston et al. 1994) MCRDR의 우수성을 실험으로 증명해 보일 것이다[7]. 우리가 실험할 것은 KBS의 성능을 알아보는 test case 추적(inference)능력, KBS의 크기, 그리고 KA기술의 복잡성 문제에 대하여 실험할 것이다.

5.1 실험

우리는 비교평가대상을 크게 세 종류로 나누었다

① MCRDR : 다시 지식수준을 C(clever), M(moderate), S(stupid) 셋으로 나누어 자세히 평가하였다.

② INDUCT/RDR : 기계학습(Machine Learning)의 한 종류인 C4.5로 생성한 KB를 RDR KB구조인 2진트리로 구성된 KB를 평가하였다[8].

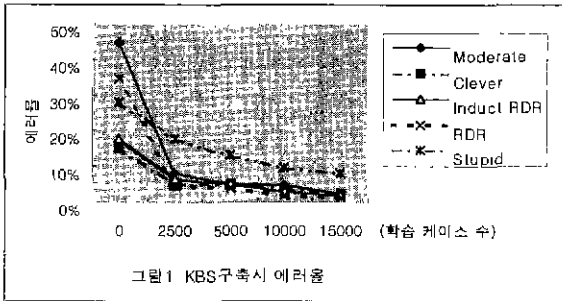


그림1 KBS구축시 어려움

③ RDR : 지식수준을 C(clever)로 설정하여 평가하였다

우리가 데이터로 사용한 GARVAN-ES1은 21,822개의 감산선 케이스로 되어있다. 이 케이스는 자연적 데이터들의 오랜 기간동안 축적된 병원데이터이기 때문에 모의전문가를 이용하여 구축한 시스템은 시간을 뛰어넘어 케이스에 의해 극복할 수 있는 실질적으로 오랜 시간 유지보수 되어온 시스템과 유사하게 구축된 것이다. 각각 비어있는 KB를 가지고 실험을 시작하고, 총 21,822개의 케이스 중 처음 case1에서 15,000까지 각각을 테스트한다 마지막 6,822개의 케이스는 test case로 사용하였다. 먼저 시스템의 추적능력과 KBS의 크기는 그림1과 2에서 잘 보여주고 있다 MCRDR에서 중요 고려사항은 KA과정에서 rule의 조건을 구하기 위하여 전문가가 확인

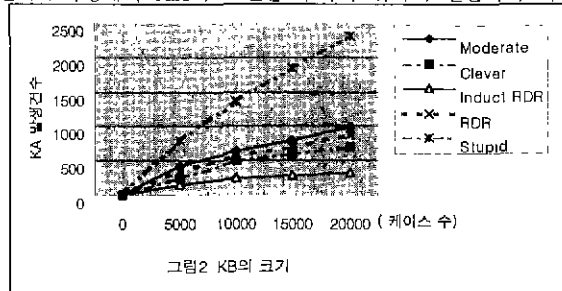


그림2 KB의 크기

해야 하는 기초케이스의 개수이다. rule의 생성과정에서 전문가가 너무 많은 기초케이스를 확인하게 되는 경우 시스템에 대한 전문가의 흥미를 반감시켜 시스템구축이 불가능할 수도 있다. 그러나 실험결과 M(MCRDR)인 경우에 평균 비교횟수가 2.25개, 최악의 경우는 13개로 나타났다. 이 정도의 숫자는 인간에 의하여 충분히 처리할 수 있는 숫자이다 그러므로 기초케이스의 개수도 문제가 되지 않는다.

6. 결론

MCRDR 전문가시스템은 기계학습에 의하여 구축된 전문가시스템, 그리고 RDR 전문가시스템에 대하여 진본적인 성능이 유사 또는 우수한 것으로 나타났다. 따라서 MCRDR 이론은 지식공학자를 배제하고 KA가 실시간으로 이루어진다는 측면에선 기계학습이론 보다 우수하고, 복수결론을 다룬다는 측면에선 RDR보다 우수하다. 이로써 우리는 MCRDR이론을 이용한 전문가시스템의 효율적인 구축을 입증하였다.

7. 참고문헌

[1] Design Problem Solving: a Task Analysis AI Magazine, 1990. 11(Winter): p. 59 - 71[11] Compton, P., Horn, K., Quinlan, J R., Lazarus, L., and Ho, K. Maintaining an Expert System, in Application of Expert Systems, J. R. Quinlan, Editor. Addison Wesley London (1989) pp. 336-385.

[2] Chandrasekaran, B., Design Problem Solving. a Task Analysis AI Magazine, 1990. 11(Winter) p. 59 - 71

[3] O'Keefe, R.M and D.E. O'Leary, Expert system verification and validation: a survey and tutorial Artificial Intelligence Review, 1993 7(1): p. 3 - 42.

[4] Kang, B.H. and P. Compton, A Maintenance Approach to Case Based Reasoning, in Advances in Case-Based Reasoning :EWCBR-94, M. Keane, J.-P. Haton, andM. Manago, Editors. 1994, Springer Paris, France. p.226-239

[5] T. Menzies : Maintaining Procedural knowledge . Ripple Down Functions In : A. Adams and L. stirling (eds.) : AI '92, Proceedings of the 5th Australian joint conference on artificial intelligence Hobart, World Scientific, Singapore 1992, pp. 335-342

[6] G. Edwards, P. Compton, R. Malor, A. Srinivasan and L. Lazarus : PEIRS : a pathologist maintained expert system for the interpretation of chemical pathology reports. Pathology 25, 27-34(1993)

[7] 박 덕진, 이 기돈, 김 우성, 장 병호 : 지식획득이론 과 모의실험에 대한 평가 : '97정보과학회 추계학술발 표 논문집 (373-379)

[8] J. R. Quinlan, P. J. Compton, K. A. Horn and L. Lazarus : Inductive knowledge acquisition : A case study. In (eds.) : Applications of Expert Systems. London : Addison Wesley 1987, pp.159-173.