

사례기반추론과 유전자 알고리즘을 결합한 지식경영 방법론에 관한 연구: 신용평가문제를 중심으로

이전창* · 신경식**

요 약

최근 기업의 경쟁력 강화를 위하여 기업내의 지식을 중요한 자원으로 인식하고 활용하는 지식경영의 필요성이 강력히 대두되고 있다. 이러한 지식경영의 주요 활동을 지원할 구체적인 방법론으로 정보기술의 활용 방안이 다각도로 제시되고 있으나, 실제적인 연구는 아직 초보단계에 있다고 하겠다.

본 연구에서는 지식의 생성, 저장, 그리고 추출 및 활용이라는 지식경영의 주요 과제를 효과적으로 해결하는 방안으로써 인공지능기법인 사례기반추론과 유전자 알고리즘을 이용한 통합방법론을 제시한다. 본 연구에서 제시하고 있는 방법론은 생성된 지식의 표현, 저장, 그리고 추출에 사례기반추론기법을 활용하였다는 점 이외에 다음과 같은 두 가지 특징을 가지고 있다. 첫째로는, 해결하고자 하는 문제에 가장 적절한 과거 지식이 추출되도록 함으로써 활용 효과를 높일 수 있도록 하였다는 점이다. 둘째로는, 환경의 변화를 반영할 수 있는 방안을 제시하고 있다는 점이다. 본 인공지능 통합방법론은 신용평가부서의 지식관리모형을 통해 검증해 본 결과 그 효과가 입증되었다.

*) 성균관대학교 경영학과

***) 이화여자대학교 경영학과

1. 서론

최근 우리나라 기업들은 사회적 환경의 변화에 따라 심각한 생산성 저하현상에 부딪치게 되었다. 내부적으로는 고임금 저생산성 현상에 당면하게 되었고, 밖으로는 저렴한 노동력을 앞세운 개발도상국의 위협과 첨단기술로 경쟁력을 강화한 선진국 사이에 끼여 이른바 "넛크래커 현상(Nutcracker Phenomenon)"에 봉착하게 된 것이다. 이러한 현상은 1997년 12월 13일 이른바 제 2의 국치로 불리는 IMF 체제하의 강제적인 경제질서를 강요하기에 이르르게 되었고, 그 결과 우리나라 기업들은 획기적인 생산성 향상 아니면 파산을 선택하여야 하는 급박한 상황에 몰리게 되었다. 이러한 불리한 내적, 외적 요인을 극복할 수 있는 새로운 경영개념으로서 "지식경영(knowledge management)"이라는 개념이 대두되기에 이르렀다.

지식경영 개념의 등장과 관련된 제반 경영환경을 좀더 검토하여 보자. 첫째, 세계 시장이 점점 하나로 단일화 되어가면서 기업들은 글로벌 스탠더드 (global standard)를 선택하지 않으면 치열한 세계시장내에서 경쟁력 있는 기업으로 남기가 어려워지고 있다. 둘째로는, 기업들이 세계화 되어가면서 점점 같은 기업이라도 각 경영활동 부서가 지리적으로 멀리 떨어져 있는 경우가 많아지고 있다. 셋째로는 인터넷 등과 같은 새로운 정보통신 환경이 일반화되어 가면서 이를 이용하여 경쟁력을 극대화해야 하는 당위성이 등장하게 되었다. 이와 같은 시장판도의 급격한 변화는 기업들로 하여금 경영 핵심능력을 보다 체계화하여 이를 조직지식 (organization memory 또는 organization knowledge) 차원으로 발전시킬 필요성을 느끼게 하였다 [26].

이러한 지식경영의 주요활동을 구현하기 위한 과거 연구는 크게 두 가지 방향으로 정리할 수가 있다. 즉, 하나는 학습조직(learning organization)의 관점으로 보는 연구방향이 있고, 둘째는 정보기술 활용의 관점으로 보는 연구방향이 있다. 학습조직은 경쟁이 치열한 시장 내에서 경쟁우위를 확보하기 위한 조직론이다. 따라서, 이러한 관점으로 지식경영을 보면 결국은 기업 구성원인 개개인의 지식을 조직지식으로 변환되는 메카니즘이 확보된 조직이 학습조직이라는 개념이 성립된다. 과거 Porter [30]와 Wiseman [39]의 연구에서는 정보기술을 전략적으로 적용하면 조직의 구조가 경쟁우위 확보에 유리한 방향으로 재편될 수 있음이 제시되었다. 그러나, Mason [23]의 연구에서는 이들의 연구가 지나치게 경쟁적인 개념만을 부각하여 경쟁이라는 개념을 너무 협의적으로만 해석한다는 점을 지적하였다. 따라서, Mason [23]은 조직을 환경변화에 적응하는 하나의 살아있는 학습조직의 개념을 경쟁우위 확보방안으로 제안하였다.

이러한 학습조직의 개념은 이미 50년대 후반과 60년대 중반부터 몇몇 학자들에 의하여 제시되었지만 [11, 3], 그 후 여러 가지 유형의 초기 지식경영의 연구의 기초를 제공하였다. 이를 살펴보면, 지능형 조직의 개념 [22, 29, 24], 생물학적 적응능력을 갖는 조직 [15, 17], 조직학습의 개념 [13, 9, 21, 36, 33, 16] 등에 널리 적용되었다. 특히, Stata [36]와 Senge [33]는 조직지식 (organizational knowledge) 또는 조직기억(organizational memory)이라는 지식경영의 핵심개념을 설명하는데 이러한 학습조직의 개념을 활용하였다. 따라서, 학습조직의 개념은 지식경영에 있어서 가장 근본적인 개념인 학습성을 설명하는 중요한 원리를 제공하게 되었고, 지식경영에 있어서 조직지식의 역할과 중요성을 일깨우는 계기가 되었다.

한편, 정보기술 활용의 관점으로 지식경영을 보는 연구는 최근 90년도에 들어오면서 발표되기 시작하였다. 특히, 정보기술이 눈부시게 발전하게 되면서 정보기술 활용이 없이는 경쟁우위를 확보한다는 것이 사실상 불가능하게 되었다. Stein & Zwass [37]는 조직지식 자체가 확보되는 것만으로는 경쟁력 있는 지식경영이 어렵고 오히려 정보기술 기법에 입각한 조직지식 지원시스템이 구축될 필요성이 있음을 주장하였다. Minch [25]도 지식경영을 도입하는 조직이 갖추어야 할 네 가지 요소들 중 하나로 조직의 지식을 모으고, 유지하며, 향후 접근할 수 있는 방법론의 마련을 들었다. 또한 O'Leary [26, 27, 28]와 Davenport [6]의 연구도 정보기술 활용을 통한 조직지식의 확보 방안 및 관련 주제를 소개하고 있다.

정보기술의 활용 관점에서 보는 지식경영 연구에서는 주로 지식생성, 지식저장, 그리고 지식추출 및 활용이라는 주제에 대하여 구체적인 정보기술의 적용가능성을 보여주는 것이 특징이다. 지식생성이란 기업내의 각 개인, 부서 등의 경영활동에서 확인된 여러 가지 형태의 명목지(explicit knowledge)와 암묵지(tacit knowledge)를 정리하여 이를 하나의 조직지식화 하는 것을 말한다. 이때 어떤 것을 명목지로 하고 어떤 것을 암묵지로 하여야 하는지에 대한 판단은 각 기업의 전략적 상황에 따라 달라진다. 정보기술의 활용관점에서 지식의 생성과정에서 함께 고려하여야 할 이슈는 생성된 지식의 표현이다. 지식의 표현 방법의 예로 규칙(rule), 의미망(semantic net), 프레임(frame), 인지지도(cognitive map) 등을 들 수 있다.

한편, 이렇게 하여 생성된 조직지식이 기업의 경쟁력 강화를 위하여 사용되기 위해서는 기업 내에 저장되어야 하는데 이와 관련된 이슈를 다루는 것이 지식저장 문제이다. 지식저장이란 지식생성을 통하여 생성된 조직지식을 지식베이스 형태로 저장하여 향후 경쟁력 강화를 위한 활용에 대비하는 것을 말한다 [1, 27]. 따라서, 조직지식을 어떠한 형태로 표현하여 어떠한 메모리 구조로 저장하는가와 같은 물리적인 스키마(physical schema)에 관한 면밀한 검토가 있어야 한다.

마지막 이슈인 축적된 지식의 추출 및 활용은 정보기술을 활용한 지식관리시스템

의 실효성에 매우 큰 영향을 미치는 요소라고 할 수 있다. 생성되고 저장된 조직지식의 목적이 그 전략적 활용에 있는 만큼 문제해결에 적절한 조직지식을 추출하여 의사결정에 도움을 줄 수 있어야 한다. 특히 Schatz [32]는 조직지식의 유용성은 환경이 변화되어도 조직의 의사결정이 효과적인 결정이 될 수 있도록 하는데 있다고 지적하였는데, 환경의 변화가 반영될 수 있는 모형의 구축을 통하여 문제해결에 적절한 지식을 추출하는 것이 중요한 과제라고 하였다.

본 연구는 정보기술의 활용을 통한 효과적인 지식경영 방법론을 제시하는데 그 목적이 있다. 본 연구에서는 위에서 서술한 정보기술의 활용 관점에서의 토론에 기초하여, 지식의 생성, 저장, 그리고 추출 및 활용이라는 지식경영의 주요 과제를 효과적으로 해결하는 방안으로써 인공지능기법(artificial intelligence)인 사례기반추론(case-based reasoning)과 유전자 알고리즘(genetic algorithms)을 이용한 통합방법론을 제시한다.

본 연구에서 제시하고 있는 방법론은 생성된 지식의 표현, 저장, 그리고 추출에 사례기반추론기법을 활용하였다는 점 이외에 다음과 같은 두 가지 특징을 가지고 있다. 첫째로는, 해결하고자 하는 문제에 가장 적절한 과거 지식이 추출되도록 함으로써 활용 효과를 높일 수 있도록 하였다는 점이다. 둘째로는, 환경의 변화를 반영할 수 있는 방안을 제시하고 있다는 점이다.

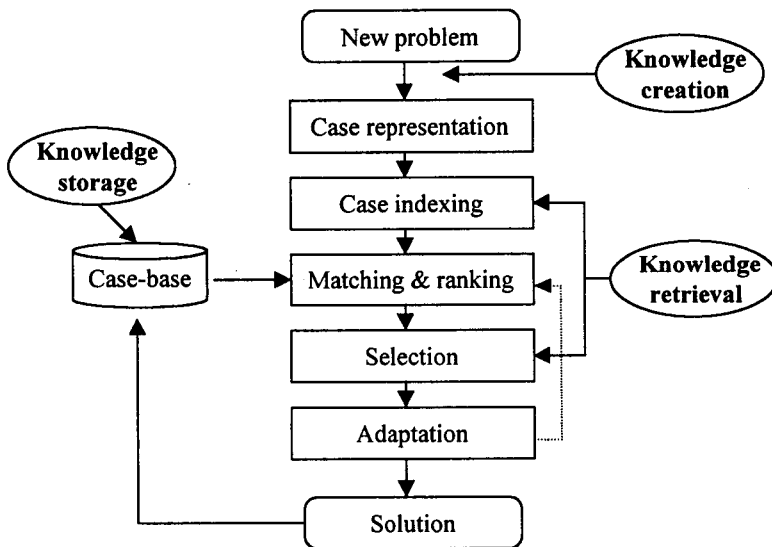
본 논문의 구성은 다음과 같다. 2장에서는 조직지식의 생성, 표현 그리고 저장을 위한 사례기반추론 기법의 적용에 관하여 설명한다. 3장에서는 저장된 지식의 효과적인 추출을 위한 사례기반추론 기법과 유전자 알고리즘의 통합 방법론을 제시한다. 4장에서는 본 연구에서 제시한 방법론을 신용평가부서의 지식관리모형을 통해 검증해 보며, 마지막 장은 결론과 연구의 한계점을 다룬다.

2. 조직지식 생성 및 저장을 위한 사례기반추론

2.1 사례기반추론기법 개요

사례기반추론(case-based reasoning)은 현재의 문제를 해결하기 위하여 저장된 과거의 사례(case)들 중 유사한 사례를 추출하여 활용하는 인공지능기법이다 [18, 19, 31, 35]. 이 기법은 사람의 추론 방식인 유사추론(analogical reasoning)과 그 추론절차가 유사하다는 특징을 가지고 있다. 최근 법률, 의학, 설계, 분석 등, 문제 해결을 위해서 폭 넓은 경험이 요구되는 분야에 활발히 적용되고 있다.

사례기반추론 기법의 추론과정을 살펴보면 다음과 같다. 우선 현재 해결하고자 하는 문제를 파악하고, 현재의 상황과 가장 유사한 과거의 사례를 추출한다. 추출된 사례는 과거 유사한 문제에 대해 해결한 결과를 포함하고 있게 되는데, 그 결과를 현재의 문제해결을 위해 사용하거나 참고함으로써 의사결정 등에 도움을 얻을 수 있게 된다. 지식경영의 중요한 요소가 지식의 생성과 저장, 그리고 효과적인 추출 및 활용이라는 점을 고려해 볼 때 지식경영시스템을 구성하는 모형에 적용가능성이 높음을 알 수 있다. [그림-1]은 사례기반추론 과정이 지식경영의 세 가지 단계에서 어떻게 적용되는지를 간략히 보여준다.



[그림-1] 지식경영과 사례기반추론 과정

2.2 지식의 생성과 표현

개인 혹은 조직이 지식을 형상화하고, 가진 지식을 컴퓨터가 이해할 수 있는 형태로 표현하는 것은 매우 어려운 과업이다. 그 동안 지식을 표현하는 다양한 방법론들이 제시되어 왔는데, 규칙(rule), 의미망(semantic net), 객체(object), 프레임(frame) 등이 이에 해당한다. 이렇게 표현된 지식은 사례의 형태로 저장하게 되는데, 사례는 과거의 경험에서 얻어진 지식을 표현하는 가장 쉬운 방법론이다 [4].

신용평가회사의 경우를 예로 들어보자. 신용평가회사의 주요업무 중 하나는 채권

발행 기업들의 신용상태에 따라 신용등급을 평가하는 것이다. 일반적으로 기업의 재무, 비재무적 요인들을 종합적으로 평가하여 그룹의사결정과정을 거쳐 최종 등급을 부여하게 된다. 한 기업의 신용도를 평가하기 위해서는 여러 경제 여건과 해당 기업의 특성 등을 고려하여야 하는 만큼 고도의 경험과 전문성이 필요하다고 할 수 있다. 이러한 노하우(know-how)는 신용평가회사가 가진 자산 중 가장 중요한 것에 해당할 것이다.

이러한 지식은 잘 관리되어야 한다. 평가가 부여된 하나 하나의 사례들은 신용평가회사의 중요한 지식이 생성되고 표현된 것으로 볼 수 있으며, 효과적으로 저장되고 추후 추출되어 활용되는 것이 필요하다. 신용평가과정에서 생성된 지식은 다음 [표-1]과 같은 형태로 표현될 수 있을 것이다. [표-1]은 어떤 기업의 주요 평가요인별 값과 평가된 최종 신용등급으로 구성된 사례의 예를 보여준다.

[표-1] 생성된 사례의 예

A 주식회사		일자: 09/01/97
요 인	현 황	
기업군	20대 그룹 계열사 소속	
기업형태	상장	
총자산	2500억원	
소유주지분	1100억원	
매출액	4000억원	
업력	25 년	
매출총이익	7.4 %	
총자산 대 현금흐름	9.4 %	
매출액 대 금융비용	14.1 %	
총자산 대 총부채	235.3 %	
감가상각 대 총비용	14.2 %	
운전자본회전율	2.4 %	
등 급	A2	

2.3 지식의 저장

위에서 언급한 바와 같이 지식은 사례의 형태로 생성되고 표현될 수 있음을 보았다. 이렇게 생성된 지식은 사례베이스(case-base)의 형태로 저장될 수 있다. 신용평가과정에서 발생한 모든 평가 사례들은 사례베이스에 저장하여 향후 의사결정에 사용하게 된다.

사례베이스에 지식을 저장하는데 있어서 고려하여야 할 점은 사례를 어떻게 저장하는 것인가 하는 것이다. 사례는 향후 추출되어 재사용을 전제로 하는 만큼 효율적이고 효과적인 추출이 가능하여야 한다. 이러한 체계적인 사례의 저장을 위해 사례의 인덱싱(case indexing) 방법론들이 다양하게 제시되고 있다. 사례의 인덱싱은 추후 효과적인 추출과도 연계되어 있다. 주로 사용되는 인덱싱 방법론으로는 수평적방법과 위계적(hierarchical) 방법이 있다 [2].

3. 지식추출을 위한 사례기반추론과 유전자 알고리즘의 통합적 접근방법

본 연구에서 제시하는 지식경영 시스템의 마지막 과제는 어떻게 개인 혹은 조직의 현재 문제 해결에 도움을 주는 축적된 지식, 즉 사례를 추출하는가 하는 것이다. 사례기반추론기법을 적용하는데 있어서 가장 어렵고 중요한 문제는 단순히 유사한 사례가 아니라 유사한 동시에 유용한 사례를 추출하는 것이다. 유용한 사례의 추출을 위해서는 적절한 유사성 평가함수를 사용하는 것이 필수적이다. 특히 적용분야의 목적에 적합한 평가함수를 사용하는데, 잘못 구축된 평가함수는 유사하기는 하나 유용하지 못한 사례를 추출하게 함으로써 의사결정에 도움을 주지 못할 가능성이 높다 [18].

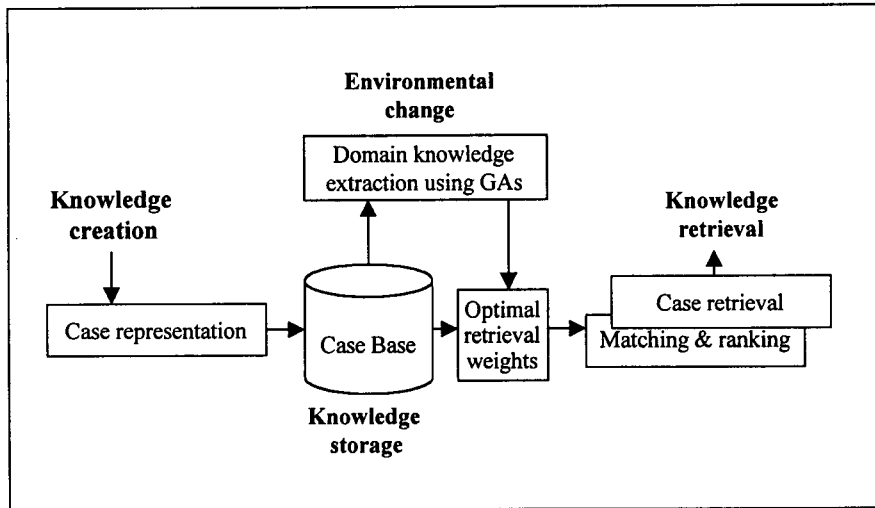
사례기반추론기법에서 가장 많이 사용되고 있는 유사사례추출 방법론은 nearest-neighbor 추출 기법으로, 이 방법론은 현재의 사례가 가지고 있는 입력 변인들과 사례베이스에 저장된 사례들의 변인들과의 유사성을 비교하여 가장 근사도가 높은 사례를 추출하는 것이다 [2, 19]. 예를 들어 위의 [표-1]에서 보여지고 있는 사례를 현재 평가가 필요한 기업이라고 가정하고, 위 기업의 최종 평가등급을 부여하고자 할 경우, 평가에 고려한 요인들을 분석하고 위의 요인들과 가장 유사한 속성을 가지는 사례를 저장된 사례에서 추출하게 된다. 사용자는 추출된 과거의 사례의 평가등급을 현재 사례에 대한 의사결정에 반영하게 된다.

그러나 nearest-neighbor 추출방법에 있어서 가장 큰 어려움 중의 하나는 비교하는 요인간의 중요도를 어떻게 반영하는가에 있다. 유사도의 측정 대상이 되는 변인들의 경우 적용분야의 특성에 따라 어떤 변인들은 더욱 중요히 고려하여야 하고, 어떤 변인들은 상대적으로 그 중요도가 낮을 수 있을 것이다. 예를 들어 신용평가를 하는데 있어서 기업의 형태보다 특정 규모의 대기업군에 소속여부가 상대적으로 더욱 중요한 요소라면, 유사 사례의 추출에 있어서 대기업군의 소속여부라는 변인의 일치, 혹은 불일치의 정도가 기업의 형태에 있어서보다 더욱 중요하게 고려되어야 할 것이다.

이러한 변인간의 가중치는 적용 분야의 특성에 고려한다고 볼 수 있다. 사례간의 유사도를 평가하고 추출하는데 있어서 적용분야의 특성을 고려하는 일은 매우 중요한 일이다. Barletta [2]와 Brown 과 Gupta [4]는 이러한 적용분야의 특성을 반영하는 일이 매우 중요한 일이나, 그 추출이 용이하지 않다고 주장하였다. Kolodner [18]도 적용 분야의 특성이 고려되지 않을 경우 사례베이스에 문제해결에 적합한 사례가 존재함에도 불구하고 찾지 못할 가능성이 있음을 경고하고 있다.

사례를 통한 추론과정에서 고려하여야 할 다른 한가지 중요한 점은 이러한 변인별 가중치가 정적(static)인 개념이 아니라 동적(dynamic)인 개념으로 이해하여야 한다는 점이다. 환경의 변화는 문제 해결 시 변인간 중요도의 변화를 가져오게 된다. 예를 들어 과거 신용평가에 있어서 10대 그룹 계열사라는 사실은 매우 중요한 요소였지만, IMF 경제체제하에서 이러한 사실은 상대적으로 그 중요성이 감소하였다는 점은 명백한 사실이다. 이는 유사성을 평가하는 가중치가 시간이 지남에 따라 변동할 수 있으며, 이러한 변화를 반영해 주어야 함을 의미한다. 환경에 따라 효과적인 지식경영시스템은 이러한 변화하는 환경을 반영할 수 있어야 할 것이다.

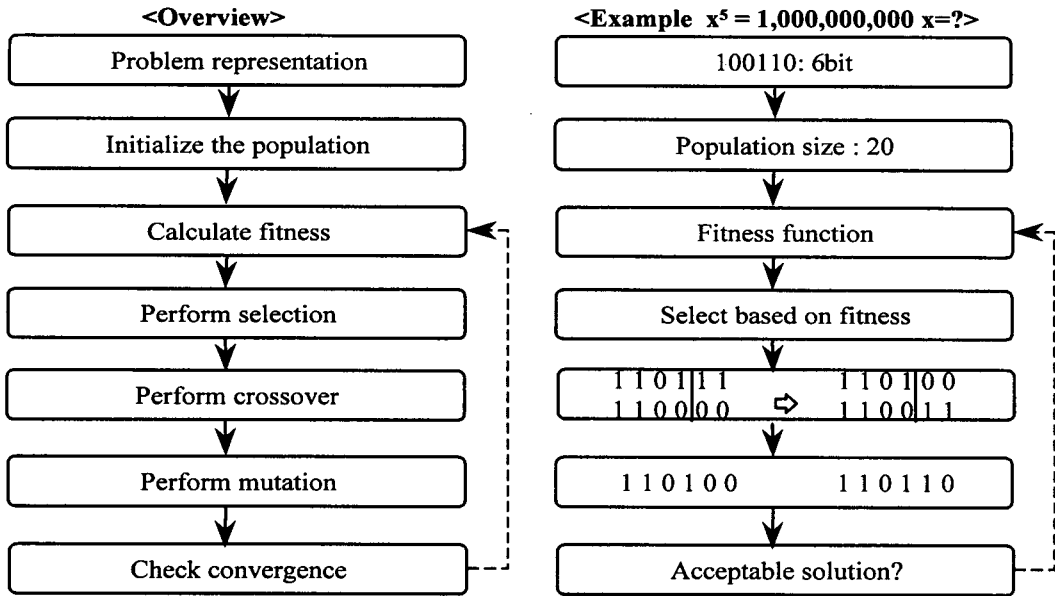
본 연구에서는 사례기반추론을 통한 적절한 과거지식을 추출하는데 있어서 가장 중요한 두 가지 문제, 즉 유사도의 평가 및 추출에 있어서 변인별 가중치로 표현되는 적용 분야의 일반지식을 반영하는 문제와 환경에 변화를 반영하는 방안으로 유전자 알고리즘의 학습기능을 활용하는 방법론을 제시하고자 한다. 제시된 모형은 [그림-2]에 간략히 나타나 있다. 이를 위해 다음에서는 유전자 알고리즘에 관한 방법론의 간략한 소개와 구체적인 모형 구축 방법론을 설명한다.



[그림-2] 지식추출을 위한 사례기반추론과 유전자 알고리즘의 통합방법론 개요

3.1 유전자 알고리즘 개요

유전자 알고리즘은 생물의 진화를 모방한 탐색기법으로 [7, 14, 10], 최적화 문제, 확률적 탐색 등에 많이 적용되고 있다 [5, 12, 20, 34]. 유전자 알고리즘은 특히 다른 인공지능 기법들과 통합하여 많이 적용되어 왔는데, 이는 인공지능 모형구축에 있어서 최적화 필요를 만족시키는데 효과적이라는 점에 서이다. 유전자 알고리즘은 초기화(initialization), 선택(selection), 교차(crossover), 그리고 돌연변이(mutation)과 같은 절차를 통해 탐색을 수행하게 된다 [7, 40]. 유전자 알고리즘의 탐색과정은 다음 [그림-3]에 나타나 있다.



[그림-3] 유전자 알고리즘의 탐색과정

초기화 단계에서는 염색체(chromosome)라고 불리는 개체를 미리 결정된 개체 수만큼 임의 생성하게 된다. 이렇게 생성된 개체들은 그 적응도를 평가하게 되는데, 평가에 사용된 함수를 적응함수(fitness function)이라고 한다. 적응함수의 선정은 적용 분야의 특성과 목적을 반영하는 만큼 중요한 절차라고 하겠다.

각각의 개체의 적응도가 평가되면 교차를 수행할 개체들을 선정하게 된다. 선정되는 개체들은 적응도가 높은 개체들로, 교차 조작에 의해 자손들을 생성한다. 적응도로 표현되는 우수한 유전적 특성을 가진 개체들을 통하여 생성된 자손들의 적응도가 높을 것이라는 시각이 내포되어 있다.

교차 조작은 선정된 개체들로부터 새로운 개체를 생성하게 하는데, 이를 통해 효과적인 탐색을 수행하게 된다. 교차는 확률에 의해 이루어지는데 다양한 조작 방법론이 제시되고 있다 [38].

돌연변이는 개체를 임의로 선정하여 염색체의 일부 값을 변경시킴으로써 새로운 개체를 생성한다. 돌연변이 역시 확률에 근거하여 이루어지는데, 이는 결국 탐색영역을 확장하고자 하는 이유에서이다. 이러한 조작이 이루어진 후 적응도를 다시 평가하고 위와 같은 유전적 조작이 수행되며, 반복적인 적용을 통하여 최종 해를 얻게 된다.

3.2 지식의 추출을 위한 Nearest-neighbor Matching

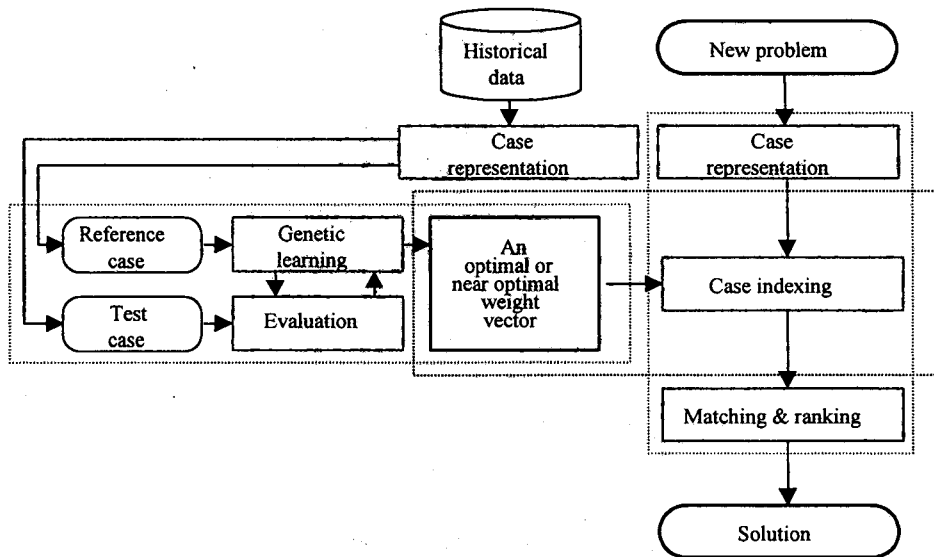
위에서 서술한 바와 같이 한 조직이 과거의 사례들을 체계적으로 축적하고 이를 효과적으로 활용하기 위해서는 단순히 유사한 사례가 아니라 현재 문제 해결에 유용한 사례가 추출되는 것이 필수적이다. nearest-neighbor 추출에 가장 많이 사용하는 유사도 측정방법은 다음과 같은 유클리디언 (Euclidean) 거리를 측정하는 것이다.

$$DIS_{ab} = \sqrt{\sum_{i=1}^n w_i \times (f_{ai} - f_{bi})^2} \quad (1)$$

위의 식에서 DIS 는 사례간의 유클리디언 거리를 나타내고, n 은 변인의 수, w_i 는 변인 i 의 가중치를 나타낸다. 유사도의 평가에 있어서 유클리디언 거리가 짧을수록 그 유사도가 높으며, 현재 문제의 해답에 가까울 가능성이 높다고 볼 수 있다. 위에서도 언급한 바와 같이 지식경영시스템의 성공은 사례의 형태로 저장된 지식이 운데에서 유용하고 적절한 사례를 추출할 수 있는가에 달려있다고 할 때, 적용 분야의 특성을 포함하고 있는 정확한 변인간의 가중치 (w_i)를 반영하는 것이 매우 중요하다. 비록 Kolodner [19]가 속성별 중요도로 표현되는 적용분야의 일반화된 지식을 반영하는 방안으로 통계적 분석에 의한 방법, 전문가의 의견을 반영하는 방법 등을 제시하였으나, 어떠한 방법이 현재의 문제를 해결하는데 최적인지를 확인하는 것은 매우 어려운 일이다.

3.2.1 유전자 알고리즘을 통한 최적 가중치의 추출과 적용

이와 같이 적용분야의 일반화된 지식을 획득하기 어렵고 다수의 사례가 존재할 경우 귀납적 학습을 통하여 지식을 추출하는 것이 보다 효과적일 것이다. 본 연구에서는 적용 분야와 환경의 특성을 나타내는 최적 변인별 가중치를 찾아내는 방안으로서 유전자 알고리즘을 통한 기계학습(machine learning)을 적용하였다. 다음 그림-4는 통합 알고리즘의 구조를 간략히 보여주고 있다.



[그림-4] 사례기반추론과 유전자 알고리즘의 통합 방법론

유전자 알고리즘을 이용한 최적화 문제를 해결하는데 중요한 몇 가지 요소들은 탐색변수의 선정과 개체들의 성과를 측정할 적합도 함수의 결정이다. 본 연구의 경우 탐색하고자 하는 변수는 nearest-neighbor matching 알고리즘에 적용할 최적의 가중치의 조합이다. 즉 적용 분야의 일반적 특성이 표현된 가중치를 찾아냄으로써 이를 유용한 사례의 추출에 사용하게 된다.

적합도 함수의 경우 구축하는 시스템의 목적에 부합하는 것이 매우 중요하다. 신용평가회사의 지식경영의 경우 새로운 기업의 평가를 수행함에 있어서 과거 평가를 수행하였던 기업들 중 유용한 사례를 추출하고 활용하는 필요를 가지고 있다. 구체적으로는 추출된 과거 사례의 평가등급이 실제 등급과 일치할수록 그 유용성이 증가한다고 볼 수 있을 것이다. 본 연구에서는 평가용 기업들을 모형에 적용하였을 경우의 적중률을 적용도값으로 설정하였다. 신용등급평가와 같은 분류(Classification) 문제의 경우 적중률은 이와 같은 정확성을 측정할 수 있는 좋은 지표가 된다.

이러한 모형의 실제적인 구현과 검증을 위하여 저자들은 과거 자료들을 참조용 사례, 평가용 사례 그리고 검증용 사례로 구분하였는데, 각각의 기능은 다음과 같다.

- (1) 참조 사례군: 일반적인 사례기반추론모형의 사례베이스에 해당하는 것으로, 현재의 문제해결을 위해 추출대상이 되는 사례군이다.

- (2) 평가 사례군: 최적의 유전자 알고리즘을 통해 최적의 가중치 조합을 찾는 데 사용되는 사례군이다. 평가사례군을 이용해 확인된 분류 적중률은 유전자 알고리즘의 적응도를 평가하는데 사용된다.
- (3) 검증 사례군: 모형의 구축에 사용되지 않는 사례군으로 모형구축 후 검증에 사용되는 사례군이다.

본 연구에서 제시하는 지식관리모형에 사용한 적응도 함수를 수리적으로 표현하면 다음과 같다.

$$\begin{aligned} \text{Maximize} \quad & CR = \frac{1}{n} \sum_{i=1}^n CA_i \\ \text{s.t.} \quad & CA_i = \begin{cases} 1 & \text{if } O(T_i) = O(S_{j^*(i)}) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

$$S_{j^*(i)} = \text{Min}_{j \in R} \left(\sqrt{\sum_{k=1}^l w_k (T_{ik} - R_{jk})^2} \right)$$

for given $i (i=1,2,\dots,n)$

- CR = 평가사례군(T)의 분류 정확도
- CA_i = 평가사례군(T)의 i 번째 사례의 분류 정확 여부 (정확히 분류 = 1, 오분류 = 0)
- $O(T_i)$ = 평가사례군(T)의 i 번째 사례의 목표값(신용등급)
- $O(S_{j^*(i)})$ = 평가사례군(T)의 i 번째 사례와의 유사도가 가장 높은 참조사례군(R)의 j 번째 사례의 평가등급
- $S_{j^*(i)}$ = 평가사례군(T)의 i 번째 사례와 참조사례군(R)의 j 번째 사례와의 거리
- T_{ik} = 평가사례군(T)의 i 번째 사례의 k 번째 변인
- R_{jk} = 참조사례군(R)의 j 번째 사례의 k 번째 변인
- w_k = 사례의 k 번째 변인의 가중치
- l = 변인의 수
- n = 평가사례군(T)의 사례 수

4. 실험 및 결과

본 연구에 사용된 실험자료는 1993년부터 1995년까지 한국신용정보주식회사의 채권등급평가자료로 2,156개 회사의 재무비율과 해당 신용등급으로 구성되어있다. 신용등급은 A1, A2, A3, B, C 등 5단계로 나뉘는데, 기간별, 등급별 구성은 다음 [표-2]와 같다.

[표-2] 데이터 구조

연 도	신 용 등 급					사 례 수
	A1	A2	A3	B	C	
1993	52	157	283	336	4	832
1994	53	186	261	295	23	818
1995	47	112	183	159	5	506
계	152	455	727	790	32	2,156

본 연구에서는 1차적으로 구축한 약 168개 정도의 재무변수 및 비재무변수들을 단일변량 t 검정, Kruskal-Wallis 검정 및 전문가 의견을 고려 약 27개정도로 압축하였다. 그리고 선택적 변수분석과정을 거쳐 9개의 재무변수와 5개의 비재무변수를 선정하였다. 통계분석을 통해 선정된 개의 비재무변수 중 감사의견과 감사수행 회계법인은 전문가의 의견을 고려 제외하고 최종적으로 12개의 변수를 선정하였다. 선정된 변수는 [표-3]에 나타나 있다.

[표-3] 선정변수

	변 수 명
X1	기업군
X2	기업형태
X3	총자산
X4	소유주지분
X5	매출액
X6	업력
X7	매출총이익
X8	총자산 대 현금흐름
X9	매출액 대 금융비용
X10	총자산 대 총부채
X11	감가상각 대 총비용
X12	운전자본회전을

선정 변수 중 기업군은 대기업집단 관계사 여부를 나타내는 것으로, 10대, 20대, 30대, 40대 그룹 계열사 및 기타 기업 등 5개로 분류하였다. 저자들은 2,156개의 자료를 연도별로 3개로 구분하고, 각각의 자료군들을 참조용 사례와 평가용 사례, 그리고 검증용 사례로 구분하였다. 위에서 간단히 설명한 바와 같이 평가용 자료군은 유전자 알고리즘을 통해 탐색되는 최적의 가중치의 적절성을 평가하는데 사용하였다. 즉 검증용 자료군의 분류 정확도를 유전자 알고리즘의 적합도 함수 (fitness function)로 설정하여, 설정된 검증용 자료들을 가장 정확히 분류하는 변수별 가중치를 찾고자 하였다. 한편 환경의 변화에 따른 가중치의 변화를 반영하기 위하여 각각의 사례군을 연도를 고려하여 구성하였다. 사례군의 구체적인 구성은 다음 [표-4]에 나타나 있다.

[표-4] 실험데이터 구조

(a) 모형의 구축 (최적 가중치 추출)

	평가용 사례군	사례수	참조용 사례군	사례수
연도 1 (1993)	1993년 사례 중 100개 기업*	100	1993년 사례 중 평가용 미포함 사례	732
연도 2 (1994)	1994년 사례 중 100개 기업*	100	1994년 사례 중 평가용 미포함 사례	718

* 추출방법 : 평가등급별 도수분포를 고려한 임의추출

(b) 모형의 검증

연도	검증용 사례군	사례수	참조용 사례군	사례수
연도 2 (1994)	1994년 사례 중 100개 기업	100	1994년 사례 중 검증용 미포함 사례	718
연도 3 (1995)	1995년 사례 중 100개 기업	100	1995년 사례 중 검증용 미포함 사례	718

** 해당 연도의 데이터를 통해 추출한 최적가중치 적용

[표-5] 유전자 알고리즘을 활용하여 추출한 변수별 가중치

연도	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
연도 1 (1993)	0.031	0.840	0.344	0.574	0.296	0.880	0.128	0.769	0.287	0.198	0.871	0.365
연도 2 (1994)	0.320	0.376	0.124	0.802	0.152	0.412	0.711	0.737	0.887	0.201	0.180	0.292

[표-5]는 유전자 알고리즘을 통해 추출된 요인별 중요도를 보여준다. 유전자 알고리즘을 통해 추출된 변수들의 중요도 가중치들은, 만일 가장 유사한 사례를 탐색하는데 사용되었을 경우 높은 적중률을 가능하게 하였을 것이다. 또 한가지 특기할 만한 사실은 1993년 자료로 구성된 사례들로부터 추출된 가중치와 1994년 자료들로부터 추출된 가중치가 다르게 나타나고 있다는 점이다. 이는 경제환경의 변화가 신

용등급 결정에 있어서 요인간의 가중치를 변화시켰다고 해석해 볼 수 있을 것이다. 예를 들면 어떤 법률 규정의 변화로 대기업의 그룹사로 편입되어 있다는 사실이 신용평가에 있어서 덜 중요한 요소로 취급될 수 있다는 것이다. 이러한 관점에서 가장 최근의 자료로부터 얻어진 변수간의 가중치를 적용하는 것은 환경의 변화를 반영하는 하나의 방안이 될 수 있을 것이다.

이러한 접근방법의 효과를 검증해 보기 위하여 일반적인 형태의 사례기반추론 모형의 적중률과 최근의 자료를 이용하여 유전자 알고리즘으로 추출한 가중치를 반영한 모형의 적중률을 비교해 보았다. [표-6]는 동일한 가중치를 적용하여 가장 유사한 사례를 추출하였을 경우의 적중률과 최근 연도의 최적 가중치를 적용하여 유사한 사례를 추출하였을 경우의 적중률을 비교하고 있다.

【표-6】 적중률 비교 (%)

연 도	기 존*	통 합 모 형	
		적용 가중치**	적 중 률
연도 2 (1994)	48.0	연도 1 (1993)	69.0
연도 3 (1995)	51.0	연도 2 (1994)	67.0

* 변인간의 가중치 동일

** 해당 연도 데이터를 통해 유전자 알고리즘으로 추출한 가중치 반영

저자들은 1994년 자료들 중 무작위로 100개의 검증용 사례를 선정하고 나머지 사례들을 사례 베이스로 사용하였다. 모형의 효과를 측정하기 위해 검증용 사례로 선정된 기업들과 가장 유사한 속성을 가지는 사례를 참조용 사례군에서 추출하고, 그 기업의 신용등급이 검증용 사례와 얼마나 일치하는지를 평가하였다. 검증에 사용된 자료군의 구성은 [표-4] (b)에 자세히 나타나 있다.

유사사례의 추출은 두 가지 접근방법을 적용하였다. 첫째는 전통적인 방법론으로 모든 속성들의 가중치가 동일하다고 가정하여 동일한 가중치를 적용하였다. 둘째는 본 연구에서 제시하는 방법론, 즉 유전자 알고리즘을 활용하여 가장 최근의 자료로부터 추출한 가중치를 적용하는 방법을 사용하였다. 예를 들어 검증결과 [표-6]에서 보여지고 있듯이 최근 사례를 활용하여 최적의 가중치를 구하고 이를 모형에 반영하였을 경우 적중률이 크게 향상된다는 것을 알 수 있다. 가중치를 적용하지 않은 모형의 분류 정확도가 48-50% 정도인 반면, 최적가중치를 반영하였을 경우 정

확도가 67~69%에 이르고 있다. 이러한 결과는 지식관리시스템의 구축에 있어서 환경의 변화를 반영할 수 있는 방법론의 적용이 매우 중요함을 시사하고 있다.

[표-7]은 통합 방법론의 통계적 유의성을 검증하기 위해 비모수검정기법인 McNemar 검정(McNemar test)을 수행한 결과이다. McNemar 검정은 두 개의 연관된 이분화된 변수가 동일한 평균을 가지는가를 chi-square 분포를 활용하여 검정한다. 분석결과 기존 모형과 통합 방법론의 분류 정확성의 차이는 5% 수준에서 유의함을 보여주고 있다.

[표-7] McNemar 검정 결과

연 도	기존모형 대 통합모형	
	Chi-square	p-value
연도 2 (1994)	6.780	.009
연도 3 (1995)	4.500	.034

실험결과 사례기반추론과 유전자 알고리즘을 결합한 통합 방법론이 효과적인 지식의 생성 및 저장, 그리고 유용한 정보를 추출을 가능하게 함을 보여주고 있다. 또한 주기적인 학습을 통해 획득한 지식을 모형에 반영하는 것이 효과적인 방법론임을 제시하고 있다.

5. 결 론

사례기반추론기법은 인공지능망(neural networks)이나 귀납적 학습방법(inductive learning) 혹은 유전자 알고리즘 등의 인공지능 기법들과는 지식의 획득과 사용에 있어서 근본적으로 다른 접근방법이다. 위에 열거한 인공지능기법들이 데이터를 통해 변인간의 연관 관계를 찾아가는 과정을 통해 일반화된 지식을 획득하는 방법이라면, 사례기반추론은 사례 고유의 지식을 활용한다.

사례기반추론은 지식의 생성 및 저장, 향후 활용 등 지식의 지식경영이라는 측면에서 볼 때 그 적용이 간편하고, 적절히 구축될 경우 효과를 기대할 수 있다. 그러나 적용분야의 일반화된 특성이나 환경의 변화가 반영되지 못할 경우 유용하지 못한 사례들이 추출되며, 나아가 지식경영시스템으로서의 가치를 가지지 못할 위험이 있다.

본 연구에서는 사례기반추론과 유전자 알고리즘의 통합방법론을 활용한 지식경영 시스템을 제안하였다. 본 연구에서 제시하고 있는 방법론은 사례기반추론을 활용하여 지식의 생성과 축적을 용이하게 하는 한편, 유전자 알고리즘을 통한 학습을 통하여 적용 분야의 특성과 환경의 변화에 따른 적용이 가능하도록 설계하였는데 그 특징이 있다. 본 연구에서 제시된 방법론을 신용평가부서의 지식관리모형에 적용해 본 결과 효과성이 입증되었다. 추후 보다 일반적인 형태의 조직이나 기업의 지식경영시스템의 효과를 높일 수 있는 인공지능기법 적용방안에 관해 연구가 필요하리라고 생각된다.

참 고 문 헌

- [1] Abecker, A., Bernadi, A., Hinkelmann, K., Kuhn, O. and Sintek, M., "Toward a technology for organizational memories", IEEE Intelligent Systems, May/June 1998, pp.40-48.
- [2] Barletta, R., "An introduction to case-based reasoning", AI EXPERT, Vol.6, No.8, 1991, pp.42-49.
- [3] Bennis, W., Changing Organizations, New York, McGraw-Hill, 1966.
- [4] Brown, C.E. and Gupta, U.G., "Applying case-based reasoning to the accounting domain", Intelligent Systems in Accounting, Finance and Management, Vol.3, 1994, pp.205-221.
- [5] Colin, A.M., Genetic algorithms for financial modeling, In Deboeck, G.J. (Eds.), Trading On The Edge. New York, John Wiley, 1994, pp.148-173.
- [6] Davenport, T.H., Some principles of knowledge management, 1996, <http://www.bus.utexas.edu/kman>.
- [7] Davis, L. Handbook of genetic algorithms, Van Nostrand Reinhold, NY., 1991.
- [8] El Sawy, O.A., Gomes, G.M. and Gonzalez, M.V., "Preserving institutional memory: the management of history as an organizational resource", Academy of Management Best Paper Proceedings, Vol.37, 1986, pp.11-122.
- [9] Fiol, C.M. and Lyles, M.A., "Organizational learning", Academy of Management Review, Vol.10, 1985, pp.803-813.
- [10] Goldberg, D.E., Genetic Algorithms in Search, Optimization and Machine Learning, MA, Addison-Wesley, 1989.
- [11] Haire, M., Biological models and empirical histories of the growth of organizations, In M. Haire (Eds.), Modern Organization Theory. New York, Wiley, 1959, pp.272-306.
- [12] Han, I., Jo, H. and Shin, K.S., "The hybrid systems for credit

- rating", Journal of the Korean Operations Research and Management Science Society, Vol.22, No.3, 1997, pp.163-173.
- [13] Hedberg, B.L.T., How organizations learn and unlearn, In P.C. Nystrom and W.H. Starbuck (Eds.), Handbook of Organizational Design, 1, New York, Oxford University Press, 1981.
- [14] Holland, J.H., Adaptation in natural and artificial systems, Ann Arbor, The University of Michigan Press, 1975.
- [15] Huber, G.P., "The decision making paradigm of organizational design", Management Science, Vol.32, No.5, May 1984, pp.138-155.
- [16] _____, "Organizational learning: the contributing processes and the literatures", Organization Science, Vol.2, No.1, 1991, pp.88-115.
- [17] _____ and McDaniel, R.R., "The nature and design of post-industrial organizations", Management Science, Vol.30, No.8, August 1986, pp.928-951.
- [18] Kolodner, J., "Improving human decision making through case-based decision aiding", AI Magazine, Vol.12, No.2, 1991, pp.52-68.
- [19] _____, Case-Based Reasoning, Morgan Kaufmann, San Mateo, CA, 1993.
- [20] Koza, J., Genetic Programming, Cambridge, The MIT Press, 1993.
- [21] Levitt, B. and March, J.G., "Organizational learning", Annual Review of Sociology, Vol.14, 1988. pp.319-340.
- [22] Mardsen, J.R. and Pingry, D.E., The intelligent organizations: some observations and alternative views, In B.R. Konsynski (Eds.), Proceedings of the 21st Annual Hawaii International Conference on Systems Science, Vol.III, 1988, pp.19-24.
- [23] Mason, R.M., "The role of metaphors in strategic information systems planning", Journal of Management Information Systems, Vol.8, No.2, Fall 1991, pp.11-30.

- [24] _____, Effective intelligent organizations: knowledge is not enough, Proceedings of the 25th Hawaii International Conference on Systems Science, Vol.IV, 1992, pp.464-471.
- [25] Minch, R.P., Hypermedia KM for intelligent organizations, Proceedings of the 23rd Annual Hawaii International Conference on Systems Science, Vol.IV, 1990, pp.300-306.
- [26] OLeary, D.E., "Knowledge management systems: converting and connecting", IEEE Intelligent Systems, May/June 1998a, pp.30-33.
- [27] _____, "Using AI in knowledge management: knowledge bases and ontologies", IEEE Intelligent Systems, May/June 1998b, pp.34-39.
- [28] _____, "Enterprise knowledge management", Computer, Vol.31, No.3, March 1998c, pp.54-61.
- [29] Paradice, D.B., The role of memory in intelligent information systems, In B.R. Konsynski (Eds.), Proceedings of the 21st Hawaii International Conference on Systems Science, Vol. III, 1988, pp.2-9.
- [30] Porter, M.E., Competitive Advantage, New York, The Free Press (Macmillan), 1985.
- [31] Riesbeck, C.K. and Schank, R.C., Inside Case-Based Reasoning, Lawrence Erlbaum Associates, Hillsdale, NJ. 1989.
- [32] Schatz, B.R., "Building an electronic community system", Journal of Management Information Systems, Vol.8, No.3, 1991-1992, pp.87-107.
- [33] Senge, P.M., "The leaders new work: building learning organizations", Sloan Management Review, Vol.32, No.1, 1990, pp.7-23.
- [34] Shin K.S. and Han, I., Using genetic algorithms to support case-based reasoning: Application to corporate bond rating, Proceedings of 2nd Asia-Pacific Decision Sciences Institute Conference, Taipei, 1998, pp.341-344.
- [35] Slade, S., "Case-based reasoning: A research paradigm", AI

- Magazine, Vol.12, No.1, 1991, pp.42-55.
- [36] Stata, R., "Organizational learning: the key to management innovation", Sloan Management Review, Vol.30, No.3, 1989, pp.63-74.
- [37] Stein, E.W. and Zwass, V., "Actualizing organizational memory with information systems", Information Systems Research, Vol.6, No.2, 1995, pp.86-117.
- [38] Syswerda, G., Uniform crossover in genetic algorithms, In Schaffer, J.D. (Eds.), Proceedings of 3rd Intl Conference of Genetic Algorithms, San Maeto, CA, Morgan Kaufmann, 1989.
- [39] Wiseman, C., Strategic Information Systems, Homewood, IL, Irwin, 1988.
- [40] Wong, F. and Tan, C., Hybrid neural, genetic and fuzzy systems, In Deboeck, G.J. (Eds.), Trading On The Edge, New York, John Wiley, 1994, pp.245-247.

Hybrid AI Approach to Knowledge Management by Integrating Case-Based Reasoning and Genetic Algorithms

Kun Chang Lee, Kyung-shik Shin

Abstract

Recently, the importance of knowledge management (KM) is drawing a great attention from researchers and practitioners because it is expected to provide a real knowledge-based intelligence to decision-makers in an organization, yielding a satisfactory level of competitive advantage in a highly turbulent and competitive market. There are three major issues in KM - how to create, how to store, and how to retrieve organizational knowledge (or memory). However, since KM studies are still in their infancy, there are few rigorous approaches to these three issues. Contrarily, this paper is aimed at proposing a hybrid AI approach to tackling those three issues of KM by integrating Case-Based Reasoning (CBR) and Genetic Algorithms (GAs). We tackled three issues of KM in the following fashion:

- (1) Knowledge creation- We assume that organizational knowledge is created in a set of cases. For illustration, we assumed KM in a Credit Evaluation Department of credit rating agency. Therefore, organizational knowledge is created in a form of cases which imply a contextualized piece of knowledge representing an experience of individuals in the Credit Evaluation Department.
- (2) Knowledge storage- Organizational knowledge is stored in a form of case base in which all the cases acquired in the process of credit evaluation are systematically kept and organized.
- (3) Knowledge retrieval- Appropriate type of organizational knowledge

should be retrieved considering the change of environment so that decision makers can take advantage of the KM in which organizational knowledge is created, stored, and retrieved to provide a knowledge-based intelligence essential for ensuring a competitive advantage. We developed a new knowledge retrieval mechanism in which GAs play an essential role of adjusting the KM mechanism to the environment change and accordingly retrieving appropriate organizational knowledge.

We applied our approach to three years data for credit evaluation and proved that our hybrid AI approach to KM is very robust in creating, storing, and retrieving organizational knowledge.