

효과적인 지식창출을 위한 웹 상의 지식채굴과정 : 주식시장에의 응용*

Knowledge Discovery Process from the Web for Effective Knowledge Creation: Application to the Stock Market

김 경 재 KAIST 테크노경영대학원 (kkj@kgsm.kaist.ac.kr)
홍 태 호 KAIST 테크노경영대학원 (hongth@kgsm.kaist.ac.kr)
한 인 구 KAIST 테크노경영대학원 (ighan@kgsm.kaist.ac.kr)

ABSTRACT

This study proposes the knowledge discovery process for the effective mining of knowledge on the web. The proposed knowledge discovery process uses the *Prior knowledge base* and the *Prior knowledge management system* to reflect tacit knowledge in addition to explicit knowledge. The prior knowledge management system constructs the prior knowledge base using a fuzzy cognitive map, and defines information to be extracted from the web. In addition, it transforms the extracted information into the form being handled in mining process. Experiments using case-based reasoning and neural networks are performed to verify the usefulness of the proposed model. The experimental results are encouraging and prove the usefulness of the proposed model.

Keywords: knowledge discovery process; web; fuzzy cognitive map; case-based reasoning; neural networks

1. 서 론

지식경영의 중요한 목적 중의 하나는 지식을 보다 잘 보이게 하는 것이다 (Allee, 1997). 즉 암묵지를 표현하고 이를 여러 사람이 공유하게 하는 것이 지식경영의 중요한 역할 중의 하나인 것이다. 그러나 주식시장이나 채권시장과 같이 비구조적이고 비선형적이며 복잡한 관계를 가진 분야에서의 암묵지를 포착하고 표현한다는 것은 매우 어려운 일이다. 이러한 분야에서의 지식을 포착하기 위하여 많은 연구들이 시계열분석방법론 등의 통계적인 방법론이나 인공지능

기법 등을 적용하여 왔다. 그러나 통계적 방법론이나 인공지능기법은 방법론의 특성상 그 자료가 계량자료일 것을 요구하는 경우가 많다.

그러나, 실질적으로 계량자료에서 얻을 수 있는 지식의 양은 한정되어 있고, 분석방법론의 한계로 인하여 적절한 지식을 포착해 내지 못하는 경우가 많다. 특히 이러한 분야에서는 계량자료에서의 지식 뿐만 아니라 전문가의 지식이나 기업 내부의 비재무적 정보 또는 정치, 국제경제상황 등의 환경적 정보에 의한 지식이 많이 포함되어 있어서 여러 지식의 통합을 통하지 않고서는 적절한 지식을 추출하는데 어려움이 있다.

* 이 논문은 1998년도 한국과학재단 연구지원(과제번호: 98-0102-08-01-3)에 의하여 연구되었음.

이러한 추세는 최근 데이터와 데이터베이스의 폭발적 증가, 인터넷의 사용에 따라 더욱 심화되고 있다. 비재무적이고 환경적인 정보에 의한 지식을 포착하기 위하여 AHP (analytic hierarchy process) 등의 방법론을 적용하려는 시도를 해 왔다. 그러나 이런 연구들은 단순히 비재무적, 환경적 정보에 의한 지식을 포착하는 데에만 관심을 두었을 뿐, 두 분야의 지식을 통합적으로 고려하는 방법에 대한 연구는 이루어지지 못하였다. 두 분야의 지식을 통합적으로 고려하기 위해서는 계량자료 뿐만 아니라 비재무적, 환경적 자료가 수집되어야 하고 이들을 통합된 지식으로 변환하기 위한 변환과정을 필요로 한다.

본 연구에서는 비재무적, 환경적 정보를 수집하기 위하여 웹 (web) 상에서 정보를 수집하고, 수집된 정보를 통합지식으로 변환하기 위하여 전문가의 지식을 이용하는 방법을 제안한다. 기존 연구에서는 이러한 정보의 변환을 임의적이거나 주관적인 방법에 의하였으나 본 연구에서는 퍼지인식도 (fuzzy cognitive map)에 의한 선행지식관리시스템 (prior knowledge management system)을 이용한다. 선행지식관리시스템은 추출해야 할 정보의 정의를 위한 선행지식베이스 (prior knowledge base)의 구축과 이를 통해서 추출된 정보를 지식으로 변환해 주는 지식변환시스템을 제어하는 역할을 담당하게 된다.

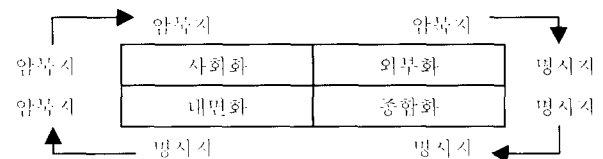
본 연구는 다음과 같은 순서에 의해 진행된다. 다음 장에서는 지식경영과 지식베이스시스템과의 관련성에 대해 토의해 보고, 3장에서는 선행지식관리시스템의 중심 기구인 퍼지인식도에 대한 기존 연구를 검토한다. 4장에서는 본 연구와 관련된 선행연구를 검토하고 5장에서는 본 연구에서 제안하는 연구모형에 대한 설명이 이어진다. 6장에서는 제안된 모형에 대한 검증작업과 결과가 제시된다. 마지막으로 7장에서는 본 연구의 한계점과 향후연구과제에 대한 논의가 이어진다.

II. 지식경영과 지식베이스시스템

Allee(1997)의 주장처럼 지식경영의 중요한 목적

중의 하나는 지식을 보다 잘 보이게 하는 것이라 할 수 있는데, 이러한 관점에서 Polanyi(1966)는 지식을 암묵지와 명시지로 구분할 수 있다고 주장하였다. 여기서 암묵지란 개인적이고 특정한 내용의 지식을 의미하며 이는 경험이나 예감, 직관이나 전문성, 실무에 종사함으로써 인한 실제적인 경험 등에 의해 형성되는 지식을 말한다. 반면에 명시지란 사람들 간에 공유되고 있는 지식으로 문서나 그림 등을 통하여 형성되는 지식을 의미한다 (Allee, 1997).

명시지는 일반적으로 표현하거나 이해하기가 용이하지만 암묵지는 이러한 것들이 용이하게 이루어지기 어렵다. 이런 측면에서 Nonaka and Takeuchi(1995)는 조직 내의 지식 창출을 암묵지와 명시지의 상호교차관계로 보고 지식변환 프로세스를 [그림 1]과 같이 제안하였다. 여기서 암묵지를 명시지의 형태로 변환하여 공유할 수 있도록 하는 작업이 매우 중요하다고 할 수 있다.



(그림 1) 지식변환 프로세스 (Nonaka and Takeuchi, 1995)

많은 연구들이 전문가시스템이나 인공지능방법론을 이용함으로써 지식경영에서의 지식창출과 지식공유를 지원할 수 있다고 주장하였다. Wiig(1997)는 지식경영에서의 지식베이스시스템의 역할에 대한 연구에서 지식경영의 기능을 네 가지 요소로 구분하고 각 기능에서의 지식베이스시스템의 역할을 제시하였다. 그는 지식경영의 역할을 지식 창조와 조달, 지식축적과 변환, 지식의 흡수, 지식의 적용 및 가치실현으로 나누고, 각 단계에서 지식베이스시스템이 유용하게 이용될 수 있음을 주장하였다. Allee(1997)는 전문가시스템이 전문가의 사고과정을 포착하며 그 지식을 이용하여 사람들을 인도할 수 있다고 주장하였다.

Davenport and Prusak(1998)은 지식경영의 도구인

전문가시스템, 사례기반추론, 인공지능망 등이 차지하는 역할을 사용자의 지식 요구정도와 해답을 찾는 데에 걸리는 시간을 기준으로 도시하였다. 그들은 특히 전문가의 지식을 요하지 않고 해답을 찾는 데에 걸리는 시간이 짧은 사례기반추론이 사례로부터 지식을 추출하는데 유용하다는 점을 강조하였다. 또한, McNeill and Freiburger(1997)는 퍼지시스템이 지식경영의 도구로서 사용될 수 있음을 제안하였다. 그들은 퍼지시스템, 퍼지칼만필터, 뉴로퍼지, 퍼지인식도 등을 그 예로 들었다. 특히 그들은 퍼지인식도가 여러 전문가의 의견을 결합하여 반영할 수 있다는 점에 주목하였다.

III. 퍼지인식도

퍼지인식도 (fuzzy cognitive map; FCM)는 인과관계의 추론을 표현하는 퍼지그래프 구조를 의미한다 (Kosko, 1986). 즉 퍼지인식도는 전문가의 지식을 추출하기 위하여 여러 개념들 간에 존재하는 인과관계를 분석하는 것이다. 따라서 퍼지인식도는 쉽게 계량화하기 힘들거나 불가능한 분야의 개념들의 인과관계를 전문가들이 가지고 있는 지식을 이용하여 계량화하고 표현하는 방법이라고 할 수 있다.

퍼지인식도는 인과관계지식을 기존의 If-Then 형태의 규칙과는 달리 행렬의 형태로 표현하기 때문에 이에 기초한 추론과정이 수렴되는 과정을 수학적 연산으로 계량화할 수 있다는 장점이 있다 (이건창 등, 1995). 또 퍼지인식도는 전문가가 여러 명 존재하는 경우에 여러 전문가의 지식을 통합하는 수단으로도 이용된다.

Kosko(1986)는 퍼지인식도의 일반적인 방법론과 인식도 (cognitive map)와의 관계에 대하여 설명하였고, Taber(1991)는 여러 개의 퍼지인식도를 결합하는 방법론을 정리하여 제시하였다. 또한 Kosko(1991)는 전술한 연구에 이어 여러 개의 퍼지인식도를 결합하는 방법과 퍼지인식도에 의한 추론과정을 소개하였다. 이진창 등(1995)은 퍼지인식도를 사용하여 주식시장에서 여러 개념들 간의 인과관계를 파악하고

이를 이용한 양방향추론을 하였다. 특히 이 연구에서는 여러 개의 퍼지인식도를 결합하는 방법으로 SOBA (Set operation-based augmentation algorithm)를 사용함으로써 기존의 단순합 또는 가중합에 의한 결합 알고리즘보다 자연스러운 결합이 가능하다고 주장하였다. Lee and Han(1998)은 EDI의 통제에 퍼지인식도를 이용하였는데 그들은 전문가의 지식을 얻기 힘든 분야에서도 구조방정식모형으로 퍼지인식도를 용이하게 구축할 수 있다고 주장하였다.

IV. 주식시장관련 선행연구

주가지수의 예측에 관련된 기존 연구로는 주로 재무변수를 이용하여 주가의 방향성이나 지수를 예측한 연구가 대부분이었다 (Shin et al., 1999). 그러나 이러한 연구들은 대부분 환경적 정보나 비재무적인 정보를 적절히 모형에 반영하지 못했다는 한계가 있었다. 환경적 정보나 비재무적인 정보를 주식시장의 예측에 적용한 연구로는 전술한 이진창 등(1995)의 논문과 Kohara et al.(1997)의 연구가 있다.

이진창 등(1995)은 퍼지인식도를 이용하여 주식시장에서의 전문가의 지식을 찾으려 하였다. 그러나 그들은 계량자료와 비계량자료를 통합적으로 이용하지 못하고 후자에 주로 의존하였다는 한계를 가지고 있다. 한편, Kohara et al.(1997)은 주식시장의 예측을 위해 신문기사의 헤드라인기사를 인공지능망의 입력변수로 추가하고 재무자료와 함께 학습시킴으로써 이를 추가하지 않았을 때의 성과보다 나은 예측성고가 나왔다고 주장하였다. 그들은 신문기사의 헤드라인에서 추출하여야 할 정보를 선정하기 위해 선행지식 (prior knowledge)을 설정하여 이에 해당하는 정보를 추출한 후 이를 계량화하여 인공지능망의 입력변수로 추가하였다. 그들의 연구에서는 환경적 자료와 재무자료, 그리고 계량자료를 모두 모형에 반영하였다는 데에 의의가 있다. 그러나 그들의 연구에서는 계량화의 기준이 주관적이고 필요한 정보의 선정과정도 임의적이었던 한계점을 가지고 있다.

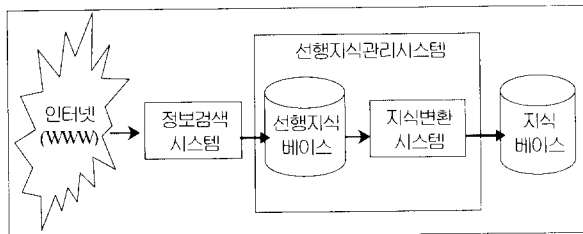
V. 연구모형

본 연구에서는 막대한 양의 정보원인 웹 상의 정보를 효과적으로 채굴하기 위한 지식채굴과정을 제안한다. 일반적으로 지식채굴과정은 Fayyad(1996)가 제시한 과정이 많이 이용된다. 그러나 이는 주로 오프라인 환경에서의 지식채굴에 관한 것이며 일반적인 입장에서 주로 계량자료의 분석 등에 적합한 과정이라고 할 수 있다. 따라서 웹 등에서 제공되는 환경적인 자료나 비재무자료 등의 비계량자료의 경우에는 그러한 과정으로 설명하기에는 어려움이 있다.

본 연구에서는 웹 상의 정보로부터 지식을 추출하기 위한 과정으로 선행지식베이스 (prior knowledge base)와 선행지식관리시스템 (prior knowledge management system)을 이용한다. 선행지식관리시스템은 퍼지인식도를 이용하여 웹에서 찾거나 하는 정보나 지식을 정의하며 선행지식베이스는 선행지식관리시스템을 통해 얻어진 선행지식을 저장하고 있다가 새로운 정보가 얻어지면 이를 선행지식을 이용하여 변환하는 역할을 담당한다. [그림 2]는 본 연구에서 제안하는 연구모형이다.

연구모형에서 선행지식베이스라는 개념을 이용하는 이유는 주식시장자료와 같이 암묵지가 많이 존재하는 분야에서 유용한 지식을 찾아내기 위하여 지식경영의 여러 연구에서 제안한 것과 같은 지식베이스시스템의 개념을 이용하는 것이 효과적이기 때문이다.

미래의 가치는 과거의 계량적 요소와 비계량적 요소에 의존하는 경우가 종종 있다. 예를 들어 주식의 미래가치는 과거의 주식이거나 경제지표는 물론, 정



[그림 2] 연구모형

치적이거나 국제적인 사건에 의해서도 영향을 받는다. 그러나 대부분의 연구에서는 계량자료의 분석에만 주력할 뿐, 정치적이거나 국제적인 사건, 투자자의 심리 상황 등의 비계량 정보는 고려하지 않았다 (Kohara et al., 1997).

주식시장의 자료는 앞에서 본 바와 같이 계량자료로서의 재무자료와 비계량자료로서의 비재무자료, 그리고 주식시장의 외부환경에 관련된 환경적 자료로 구분할 수 있다. 여기서 환경적 자료란 주식시장에 직접적인 영향을 미치는 요소는 아니지만 투자자의 심리 요인 등을 통하여 간접적으로 주식시장에 반영되는 요인을 의미한다. 재무자료에도 암묵지가 포함되어 있지만 비재무자료와 환경적 자료에는 재무자료보다 더 많은 양의 암묵지가 존재한다고 할 수 있다. 따라서 이런 자료들로부터 유용한 정보를 추출하고 이를 지식으로 변환하기 위해서는 주식시장의 전문가들의 지식을 이용할 수 있다. 이를 위해서는 전문가의 지식을 추출하여야 하는데, 전문가들의 지식을 용이하게 반영할 수 있는 방법으로 본 연구에서는 선행지식관리시스템이라는 개념을 사용한다.

인터넷 상에서 지식을 채굴하는 과정은 다음과 같다. 무한한 정보원으로부터 필요한 정보를 찾기 위해서는 먼저 어떤 정보가 필요한 지를 파악하여야 한다. 본 연구에서는 Kohara et al.(1997)의 연구에서의 임의성과 주관성을 배제하기 위하여 선행지식관리시스템을 이용하는데 이는 무한한 정보원로부터 추출하여야 할 정보를 정의해 주는 선행지식베이스의 구축과 추출된 정보를 지식으로 변환해주는 지식변환시스템의 역할을 하게 된다. 선행지식관리시스템의 구축과 관리, 유지는 기본적으로 퍼지인식도에 의해서 이루어진다. [그림 3]은 본 연구에서 사용된 퍼지인식도이다. 이 퍼지인식도는 신문기사의 헤드라인으로 자주 사용되는 키워드를 주식시장 분석 전문가들에게 제공하고 그 결과를 설문조사에 의해 얻은 후 여러 개의 퍼지인식도를 통합하여 얻은 것이다.

	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀	C ₁₁	C ₁₂	C ₁₃	C ₁₄	C ₁₅	C ₁₆	C ₁₇	C ₁₈	C ₁₉	C ₂₀	C ₂₁	C ₂₂	C ₂₃	C ₂₄	C ₂₅	C ₂₆
C ₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₂	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₃	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₄	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0
C ₅	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0
C ₆	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₇	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0
C ₈	0	0	0	0	0	0	0	0	0	+1	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₉	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₀	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₂	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₃	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₄	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₅	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₆	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0
C ₁₇	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₈	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₁₉	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0
C ₂₀	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₂₁	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0
C ₂₂	0	0	-1	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0
C ₂₃	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₂₄	0	0	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0
C ₂₅	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C ₂₆	0	0	0	0	0	0	0	0	0	0	0	+1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[그림 4] 퍼지인식도 행렬의 예

서 균형상태에 이르렀음을 의미한다.

따라서 “주가폭등”이라는 기사는 C2에서 “조정국면”과 “경계심리”라는 요소에 긍정적인 영향을 미치는 것임을 알 수 있다. 그러므로 퍼지인식도를 이용하여 비계량자료로 추출된 정보를 계량화할 수 있으며 재무자료와 함께 통합적인 추론을 수행하는 모형을 만들 수 있다.

V. 모형의 검증

제안된 연구모형이 유용한 지를 검증하기 위하여 다음과 같은 실험을 하였다. 검증에 사용된 자료는

중앙일간지 인터넷 홈페이지의 1998년 1월 3일부터 1999년 4월 23일까지 기사의 헤드라인뉴스 부분을 사용하였고, 동 기간의 한국종합주가지수와 거래량을 바탕으로 한 기술적 지표를 자료로 사용하였다. 실험에 사용된 지표는 증권시장이나 관련연구에서 주로 사용되는 기술적 지표로 구체적인 산식은 부록에서 제시되고 있다.

검증과정을 위해 먼저 퍼지인식도에서 전문가들에 의해 구축된 선행지식을 이용하여 키워드를 정의한다. 예를 들어 “주가폭등”이라는 선행지식을 표현하기 위해 [표 1]과 같이 키워드 셋(keyword set)을 표현한다.

[표 1] 선행지식을 이용한 키워드 셋

선행지식	키워드 셋	키워드1	키워드2	키워드3
주가폭등	1	주가	큰폭	상승
	2	주가	큰폭	하락
	3	주가	폭등	
	4	주가	폭락	

정의된 선행지식을 이용하여 키워드 매칭 (keyword matching)을 하기 위해서는 다음과 같은 논리적 규칙을 따른다.

$If (KW_{1k} \subseteq A_{jk}) \text{ and } (KW_{2k} \subseteq A_{jk})$

$\text{and } (KW_{3k} \subseteq A_{jk}),$

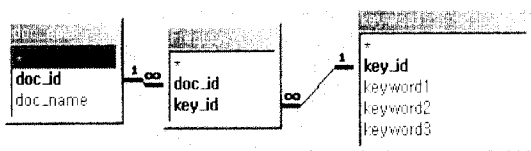
$\text{then } PK_k$

($KW_{ik} = k$ 번째 선행지식의 키워드 i ,

$A_{jk} = k$ 번째 선행지식의 키워드 셋 j ,

$PK_k = k$ 번째 선행지식)

또한 키워드 셋과 헤드라인뉴스 문서로부터 관계형 데이터베이스를 이용하여 질의를 통해 해당 정보를 추출해 내기 위해, [그림 5]와 같은 관계를 갖는 데이터베이스를 구축하였다.



[그림 5] 정보검색을 위한 데이터베이스 관계

추출된 정보는 퍼지인식도를 이용하여 앞에서의 예시와 같은 행렬연산을 통해 벡터형태로 변환된다. 이를 미리 구축되어 있는 기술적 지표들과 함께 사용하여 사례기반추론과 인공지능망 실험을 시행한다. 사례기반추론과 인공지능망은 비구조화된 문제로부터 지식을 추출하는데 널리 이용되는 방법으로, 사례기반추론은 과거의 사례로부터 특정지식을 추출하는데 유용하고 인공지능망은 과거 사례로부터 일반화된 지식을 추출하는데 유용한 것으로 알려져 있다. 본 연구에서

는 두 방법론의 비교분석보다는 상이한 특징을 갖는 두 가지 지식추출방법을 이용하여 본 연구에서 제안된 연구모형이 두 방법론에서 유용한 지를 살펴 보고자 한다. 실험의 산출물은 다음날 주가지수의 증가의 방향성이다.

제안된 연구모형의 검증은 위해 기술적 지표만으로 구성된 자료를 이용하여 같은 방식으로 사례기반추론과 인공지능망 실험을 실시한다. 사례기반추론과 인공지능망 실험에 사용된 기술적 지표는 7개이고, 사용된 자료 중 사례기반 구축과 인공지능망 모형에서의 학습용 자료는 98년 1월 3일부터 98년 12월 28일까지의 292개의 자료이며, 구축된 사례기반과 학습용 자료를 사용하여 추론하는 데 이용된 검증용 자료는 99년 1월 4일부터 99년 4월 23일까지의 75개 자료이다. 사례기반추론에서 유사도를 결정하는 방법으로는 최근접이웃법 (nearest-neighbor method)을 사용하였으며 한번의 추론에 사용되는 이웃의 수는 5개로 고정하였다. 인공지능망에서는 한 개의 은닉층을 갖는 역전파 인공지능망을 사용하였으며 은닉층 내의 은닉노드 수는 7개로 고정하였다. 이러한 과정을 통해 얻어진 추론결과를 [표 2]와 같이 적중율로 환산하여 제시하였다. 적중율은 전체 검증대상사례 중에서 기대값과 예측값이 일치한 사례가 차지하는 비율을 백분율로 표시한 것이다.

[표 2] 검증 결과 (단위: 적중율)

추론방법	사례기반추론	인공지능망
기술적지표로 추론	57.3%	54.7%
제안된 방법	60.0%	60.0%

추론 결과, 제안된 방법에 의한 결과가 기술적 지

표만을 이용하여 추론한 결과보다 우월한 성과를 보이는 것으로 나타났다. 이는 제안된 방법에서 사용된 웹 상의 정보가 추론과정에서 유용한 역할을 했다는 것을 의미한다. 그러나 두 방법의 성과가 큰 차이를 보이지 않고 있는데 이는 추론에 사용된 자료의 부족과 인터넷기사 원천의 부족 등에 의한 것으로 생각된다. 따라서 향후 연구에서는 이러한 정보원천의 추가와 이에 따른 검증이 필요하다고 생각된다.

VI. 결 론

본 연구에서는 주식시장에서의 지식채굴을 위해 기존의 계량자료 외에 웹 상의 환경적, 비계량적 자료를 지식채굴과정에 포함하는 모형을 제안하였다. 이 모형에는 환경적, 비계량적 자료를 포함함으로써 발생할 수 있는 임의성과 주관성을 감소시키고 해당 분야 전문가의 지식을 반영하기 위하여 퍼지인식도 개념을 응용한 선행지식관리시스템을 사용하였다.

본 연구는 비계량적 자료를 지식채굴과정에 보다 객관적이고 체계적인 방법에 의해 포함하는 모형을

제시하였다는 점에서 기존 연구의 지식채굴과정이 대부분 계량적 자료의 분석에만 치우쳤다는 한계점을 보완한 것이다.

모형의 검증결과에서 제시된 바와 같이 본 연구에서 제안된 지식채굴과정이 유용한 방법론이라고 할 수 있다. 그러나 본 연구의 한계는 정보원천의 부족으로 인해 성과차이의 유의성에 대한 충분한 검증이 부족하였다는 점이다. 또한 응용사례를 주식시장에 국한함으로써 다른 분야에서도 유용한 방법론이 될 수 있을지에 대한 검증을 하지 못하였다. 따라서 향후 연구에서는 인터넷 정보원천의 다양화를 위해 이동성을 갖는 검색 시스템이나 지능을 가진 검색시스템을 이용하여 충분한 정보원천을 확보하여야 할 것이다. 또한 주식시장 이외의 다른 분야에서의 유용성에 관한 검증을 시도함으로써 제안된 지식채굴과정이 일반적인 방법론인지에 대한 논의가 필요하다고 하겠다.

부 록

본 연구에서 사용된 기술적 지표의 산식은 <표-부록>과 같다.

[표-부록] 기술적 지표 (Kolb and Hamada, 1988; Achelis, 1995; Gifford, 1995; Edwards and Magee, 1997)

지표명	산식	지표명	산식
Stochastic %K	$\frac{C_t - LL_{t,5}}{HH_{t,5} - LL_{t,5}} \times 100$	ROC	$\frac{C_t}{C_{t-n}} \times 100$
A/D Oscillator	$\frac{H_t - C_{t-1}}{H_t - L_t}$	Disparity	$\frac{C_t}{MA_n} \times 100$
CCI	$\frac{(M_t - SM_t)}{(0.015 \times D_t)}$	OSCP	$\frac{MA_5 - MA_{10}}{MA_5}$
RSI	$100 - \frac{100}{1 + \frac{\sum_{i=0}^{n-1} Up_{t-i}}{\sum_{i=0}^{n-1} Dw_{t-i}}}$		

C: 종가, L: 저가, H: 고가, LL_n: n기간 중 최저가, HH_n: n기간 중 최고가, MA: 이동평균, M_t: $\frac{(H_t + L_t + C_t)}{3}$,
 SM_t: $\frac{\sum_{i=1}^n M_{t-i+1}}{n}$, D_t: $\frac{\sum_{i=1}^n |M_{t-i+1} - SM_t|}{n}$ Up: Upward price change, Dw: Downward price change.

참 고 문 헌

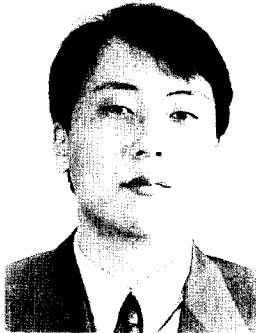
- 이건창, 주석진, 김현수 (1995), 퍼지인식도에 기초한 인과관계 지식베이스 구축과 양방향 추론방식에 관한 연구, 한국전문가시스템학회지, 제1권, 제1호, 1-22.
- Achelis, S. B. (1995), *Technical Analysis from A to Z*, Probus Publishing.
- Allee, V. (1997), *The Knowledge Evolution: Expanding Organizational Intelligence*, Butterworth-Heinemann.
- Davenport, T. H., and L. Prusak (1998), *Working Knowledge*, Boston, Massachusetts, Harvard Business School Press.
- Edwards, R. D., and J. Magee (1997), *Technical Analysis of Stock Trends*, Chicago, Illinois, John Magee.
- Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth (1996), "From Data Mining to Knowledge Discovery: An Overview", In U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, and R. Uthurusamy (Eds.), *Advances in Knowledge Discovery and Data Mining*, Cambridge, Mass., AAAI/MIT Press, 1-34.
- Gifford, E. (1995), *Investor's Guide to Technical Analysis: Predicting Price Action in the Markets*, London, Pitman Publishing.
- Kohara, K., T. Ishikawa, Y. Fukuhara, and Y. Nakamura (1997), "Stock Price Prediction using Prior Knowledge and Neural Networks", *International Journal of Intelligent Systems in Accounting, Finance and Management*, Vol. 6, 11-22.
- Kolb, R. W., and R. S. Hamada (1988), *Understanding Futures Markets*, Scott, Foresman and Company.
- Kosko, B. (1986), "Fuzzy Cognitive Maps", *International Journal of Man-Machine Studies*, vol. 24, 65-75.
- Kosko, B. (1991), *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*, Prentice-Hall.
- Lee, S., and I. Han (1998), "Fuzzy Cognitive Maps for EDI Controls Design: Integrated Approach with Structural Equation Modeling", In *Proceedings of '98 Joint Conference of the Korea Management Information Systems and Expert Systems Society*, Seoul, Korea, 103-105.
- McNeill, D., and P. Freiburger (1997), "Webs of Cognition", In R. L. Ruggles III (ed.), *Knowledge Management Tools*, Butterworth-Heinemann.
- Nonaka, I., and H. Takeuchi (1995), *The Knowledge-Creating Company*, Oxford: Oxford University Press.
- Polanyi, M. (1966), *The Tacit Dimension*, London, Routledge & Kegan Paul.
- Shin K., K. Kim, and I. Han (1999), "Financial Data Mining Using Genetic Algorithms Technique: Application to Korean Stock Market", In *Proceedings of Pacific Asia Conference on Expert Systems*, Los Angeles, California.
- Taber, R. (1991), "Knowledge Processing with Fuzzy Cognitive Maps", *Expert Systems With Applications*, Vol. 2, 83-87.
- Wiig, K. M. (1997), "Roles of Knowledge-Based Systems in Support of Knowledge Management", In J. Liebowitz, and L. C. Wilcox (eds.), *Knowledge Management and Its Integrative Elements*, Florida, CRC Press.

○ 저 자 소 개 ○



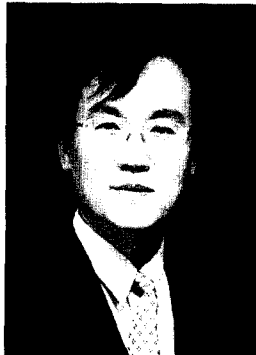
김 경 재 (Kyoung-Jae Kim)

중앙대학교 경영학과를 졸업하고 KAIST 테크노경영대학원에서 경영공학석사를 취득하였으며 동대학원에서 박사과정을 수료하였다. 현재 KAIST 테크노경영대학원 박사과정 재학중이다. Expert Systems with Applications 등에 논문을 발표하였으며, 주요 연구관심분야는 인공지능의 재무응용, 데이터마이닝, Intelligent Agent, 지식경영 등이다.



홍 태 호 (Tae-Ho Hong)

한국과학기술대학 산업공학과를 졸업하고 KAIST에서 경영정보공학으로 석사학위를 취득하였으며, 현재 KAIST 테크노경영대학원 박사과정에 재학중이다. 주요 연구관심분야는 데이터 마이닝, 지능형 에이전트, 인터넷상에서의 지식추출과정, 그리고 인공지능을 응용한 재무정보시스템 등이다.



한 인 구 (In-Goo Han)

서울대학교 국제경제학사, KAIST 경영과학석사를 취득하고 University of Illinois at Urbana-Champaign에서 회계정보시스템을 전공하여 경영학박사를 취득하였다. 현재 KAIST 테크노경영대학원 부교수로 재직중이다. Decision Support Systems, Expert Systems with Applications, Information & Management, International Journal of Intelligent Systems in Accounting, Finance and Management, International Journal of Electronic Commerce 등에 많은 논문을 발표하였으며, 주요 연구관심분야는 지능형 신용평가시스템, 인공지능을 이용한 주가예측, 지식자산 가치평가, 정보시스템 감사 및 보안 등이다.