

산업분야에서의 지식 정보 추출에 대한 비교연구

Comparative Study of Knowledge Extraction on the Industrial Applications

우영광, 김성신, 배현, 우광방*

Young-Kwang Woo, Sungshin Kim, Hyeon Bae, and Kwang Bang Woo*

부산대학교 전기공학과

*연세대학교 자동화기술연구소

School of Electrical and Computer Eng., Pusan National University, Busan, Korea

*Automation Technology Research Institute, Yonsei University, Seoul, Korea

Tel: 051-510-2374, E-mail: sskim@pusan.ac.kr

요약

데이터는 어떤 특성을 나타내는 언어적 또는 수치적 값들의 표현이다. 이러한 데이터들을 목적에 따라 구성한 것이 정보이며, 문제 해결이나 패턴 분류, 또는 의사 결정을 위해 정보들간의 관계를 규칙으로 체계화하는 것이 지식이다. 현재 대부분의 산업 분야에서 시스템에 대한 이해를 높이고 시스템의 성능을 향상시키기 위해 지식을 추출하고, 적용시키는 작업들이 활발히 이루어지고 있다. 지식 정보의 추출은 지식의 획득, 표현, 구현의 단계로 구성되며 이렇게 추출된 지식 정보는 규칙으로 도출된다. 본 논문에서는 여러 산업 분야에 걸쳐 다양하게 적용되는 지식 정보 추출 방법들에 대해 그 영역별로 알아보고 여러 시험 데이터들과 실제 시스템에 클러스터링(CL), 입력공간 분할(ISP), 뉴로-퍼지(NF), 신경망(NN), 확장 행렬(EM) 등의 방법들을 적용시킨 결과들을 비교 분석하고자 한다.

Keywords: Knowledge & Rule Extraction, Clustering, Neuro-Fuzzy, Extension Matrix

1. 서론

인공지능은 특정 분야에서 전문가의 축적된 지식과 경험을 시스템화하여 필요할 때 사용하도록 하는 알고리즘이다. 여기서 지식이라는 용어는 정보와 혼용하여 사용되거나 구분되기도 하는데, 데이터 관점에서 엄밀하게 정의하면 정보란 데이터를 가공, 처리해서 얻어지는 산출물이고, 지식이란 정보의 체계적인 활용을 통해 축적된 노하우이다[1]. 정보는 그 자체로서도 경쟁력의 원천이 되지만, 지식으로 발전하게 되면 경쟁력은 더욱 강화된다[2]. 이에 본 논문에서는 지금까지 연구되어온 지식 추출의 방법들을 비교 검토하고자 하였다. 먼저 지식 추출을 위해 사용된 기존의 연구 동향을 전체적으로 파악하여 알고리즘에 따라 분류하였다. 크게 클러스터링을 이용한 방법[3], 입력 공간 분할을 이용한 방법[4], 뉴로-퍼지 모델링[5], 신경망을 이용한 방법[6] 등을 큰 항목으로 나누었다. 그리고 그 외 확장 행렬, 고차 분할 등에 대한 내용을 조사하여 수록하였다. 이와 같은 방법론적인 분류를 통해 지식도출을 위해 사용된 지능형 방법의 종류들을 신체적으로 검토하고자 하였다. 또한 방법론적인 분류와 함께 적용 분야에 대해서도 분류하였다. 실제

대상의 특성에 따라 적절한 방법이 선택되어야 좋은 결과를 얻을 수 있다. 대상에 대한 분류는 대상에 따른 효율적인 사용과 그 특성을 파악하는데 많은 정보를 제공할 것으로 판단된다. 본 논문은 지금까지 연구된 지능제어 기법들을 이용한 지식도출의 방법들을 방법과 대상에 따라 분류하여 적절한 방법의 선택을 위한 정보를 제공하고자 한다.

2. 지식 추출 기법들의 추세

지식 추출의 방법은 현재 가장 발전된 지능형 방법들 중 하나라고 할 수 있다. 최근 10년간의 여러 가지 지식 추출 방법들의 현황에 대해 IEEE에 게재된 지식 도출에 관한 77편의 논문들 중 26편을 선별하여 동향을 분석하였다. 표 1에서 알 수 있듯이 1990년대 초반 뉴로-퍼지(NF)가 도

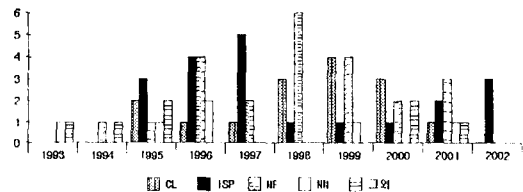


그림 1. 규칙추출 방법들의 추세.

표 1. 연도별 규칙 추출 관련 논문.

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
CL	0	0	2	1	1	3	4	3	1	0
ISP	0	0	3	4	5	1	1	1	2	3
NF	0	1	1	4	2	6	4	2	3	0
NN	1	0	1	2	0	0	1	0	1	0
기타	1	1	2	0	0	0	0	2	1	0

입된 이래로 꾸준히 지식 추출의 중요한 방법으로 사용되고 있고 클러스터링(CL)[7]과 입력공간 분할 방법(ISP)도 많이 적용되고 있다. 또한 이러한 주요 방법들 이외에도 확장 행렬[8], 고차 분할[9], 결정 트리(DT)[10] 등에 의한 방법들도 꾸준히 연구되고 있다. 주목할 만한 사실은 입력공간을 단순하게 등분하는 개념이었던 입력 공간 분할 방법이, 1990년대 후반 들어 다양한 인공지능 기법들의 적용으로 인해 잠시 주춤했다가 전문가의 지식을 반영하는 조건부적 입력 공간 분할로 바뀌면서 다시 활발히 적용되고 있다는 점이다.

3. 각 분야들에 대한 비교

본 절에서는 제어, 인식, 진단, 환경, 주식 시스템 등에서 지식 추출 방법들이 어떻게 적용되었는지 검토할 것이다. 각 영역별로 어떠한 방법들이 주로 사용되었는지 알아보고 그 이유에 대해서도 살펴본다.

3.1 제어 시스템에 대한 적용

제어 시스템에선 클러스터링이나 입력 공간의 분할에 의해 대략적인 규칙들을 정하고 그 규칙들을 최적화 하는 방법이 사용되었다. N. T. Gurman은 RuleNet이라는 클러스터링을 이용한 지식 정보 추출부분을 더한 모델과 거기에 퍼지 논리를 접목시킨 모델을 비교한 내용을 발표하였다[3]. Z. M. Yeh는 비 선형적이고 시간에 따라 변하는 2축 자동제어 장치를 개발하였다[11]. D. Kim *et al.*은 트럭의 Backer-upper 제어를 위해 Lamackin co-adapted 퍼지로지식을 사용하였다[12]. H. J. Cho *et al.*은 PID 제어에 있어 Genetic fine tuning 방법, Ziegler-Nichols 방법, Manual tuning을 비교한 결과를 발표하였다[13]. A. Lekova *et al.*은 교차로에서 발생하는 입력 상황에 따라 개개의 입력 공간들이 중첩되는 지점들을 유전자 알고리즘으로 찾은 뒤 분류된 규칙의 수를 최적화 하는 방법을 제안하였다[14]. B. J. Choi *et al.*은 진자의 슬라이딩 모드 제어에서, 입력과 제어선과의 거리로 대략적인 규칙을 만든 뒤 유전자 알고리즘으로 규칙의 개수를 줄이는 방법을 발표하였다[15]. S. F. Smith *et al.*은 진자를 제어하는데 있어 최적제

표 2. 제어 시스템 적용.

적용분야	특징	방법 논문
로봇 항해	클러스터링으로 초기규칙을 정한 다음 GA등을 이용한 방법으로 최적화	CL [3]
컨투어 제어	원하는 Contour 곡선과 퍼지 추론을 이용하여 원하는 범위까지의 오차를 임계값으로 하여 규칙을 감소	ISP [11]
트럭제어	대략적인 규칙을 정하고 소속함수를 GA로 최적화	NF GA [12]
PID제어	초기 소속함수를 선택하고 GA로 최적화	NF GA [13]
교통교차로	대략적인 규칙을 정하고 GA에 의한 최적화로 규칙의 개수를 감소	NF GA [14]
진자제어	오차와 오차변화량으로 초기 규칙을 만든 후 거리공식으로 규칙 수를 감소	NF GA [15]
	최적제어 표를 구성해서 클러스터들로 규칙을 추출	CL [16]
압력제어	입력의 상태에 따른 규칙을 정하고 GA로 소속함수를 최적화	NF GA [17]
	클러스터들로 규칙을 구성	CL [18]

어표를 구성한 뒤 클러스터링으로 규칙을 도출하는 것을 제안하였다[16]. O. F. Nelles *et al.*은 디젤 엔진의 속도, 연료 유입량, 압력 등 세 가지 입력을 가지고 이후의 압력의 상태를 조절하는 시스템을 개발하였다[17]. M. Setnes는 발효 공정이 진행되는 동안 압력을 제어하기 위하여 가 변수에 대해 클러스터의 개수를 변화시키면서 최적의 규칙을 도출하였다[18].

3.2 인식 시스템

인식 시스템에서는 주로 전문가의 지식이 반영된 조건에 의한 입력 공간의 분할 방법이 사용되었다. K. Nozaki *et al.*은 경험적 방법에 의해 규칙을 도출하는 방법을 발표하였다. 입력공간을 임의의 k 개로 나누고 그 숫자만큼 소속함수를 구성한 다음 소속함수의 가중치를 a 승으로 정의하여 규칙을 찾았다[4]. S. H. Huang *et al.*은 입력을 퍼지값으로 분류하여 여러 개의 이진수로 표현하고, 지식에 따라 신경망으로 학습시켜 규칙을 구성하는 방법을 제안하였다[19]. S. Chiu는 입력공간을 클러스터링하여 초기규칙을 도출한 후, Gradient Descent 알고리즘을 사용하여 소속함수를 최적화 하는 방법을 적용하였다[20]. N. R. Pal *et al.*은 초기규칙을 정하고, Gradient Descent로 조정된 다음 결과를 확인함으로써 규칙을 줄이는 알고리즘을 선보였다[21].

3.3 진단 시스템

진단 시스템에서는 이미 구성된 모델의 규칙들 중에서 의사결정에 비교적 덜 영향을 미치는 규칙을 구분해 제거하는 방법이 주로 사용되었다. T. Pfeufer *et al.*은 자동차의 속도를 직접적으로 조절하는 actuator의 상태를 파악하는 방법으로 뉴로-퍼지 규칙 노트에서 역전파로 가중치를

표 3. 인식 시스템 적용.

적용분야	특징	방법 논문
패턴 인식	입력공간을 k 개로 분할하여 소속함수를 구성하고 가중치를 a 승으로 정의하여 최	ISP [4]
	적의 파라미터에 해당하는 규칙 생성	
	퍼지값으로 입력을 분류, 이진으로 표현	ISP [19]
	하고, 신경망으로 학습하여 규칙 구성	NF [19]
	입력공간을 클러스터링하여 규칙 결정	CL [20]
	결정트리로 규칙 확정	DT [10]

표 4. 진단 시스템 적용.

적용분야	특징	방법 논문
자동차 actuator	역전파로 가중치를 주어 의사결정에 영향을 적게 미치는 뉴런을 제거하는	ISP [21]
	방법으로 규칙 생성	NF [21]
자동차 타이어압력	역전파로 가중치를 줘서 의사결정에 영향을 적게 미치는 뉴런을 제거하는	ISP [22]
	방법으로 규칙 구성	NF [22]

주어 의사결정에 영향을 적게 미치는 뉴런을 제거하는 방법을 제안하였다[21]. M. Ayoubi *et al.*은 자동차의 타이어 치우침과 고유 회전 주파수를 이용하여 전달함수를 정의하고, 뉴로-퍼지 방법으로 학습시켜 의사결정에 영향을 적게 미치는 뉴런을 제거하는 방법을 개발하였다[22].

3.4 환경 시스템

환경 시스템에서는 환경 전문가들이 가지는 지식이 규칙에 반영된다. M. Cococcioni *et al.*은 바다의 오염도를 MERIS 데이터를 이용해 진단하기 위해, 클러스터링으로 초기 규칙을 정한 후 유전자 알고리즘으로 규칙을 최적화 하는 방법을 제안하였다[23]. E. F. Carrasco *et al.*은 축산 폐수 처리 공정의 산화 상태를 결정하기 위한 전문가 지식에 기반한 퍼지 추론 시스템을 개발하였다[24]. J. P. Steyer *et al.*은 산업폐수 처리장의 혐기공정 제어를 위하여, 센서를 통해 얻어지는 값들을 입력으로 전문가의 지식을 이용한 퍼지 제어를 설계하였다[25]. S. Dash *et al.*은 화학 처리 공정 진단에 대한 내용을 발표하였다. 센서로부터 얻을 수 있는 프로파일을 패턴 매칭 방법으로 분류하여 규칙을 구성하였다[26]. A. Muller S *et al.*은 하· 폐수 처리 공정의 외란을 퍼지 논리를 이용하여 위험한 입력조건에서의 패턴을 규칙화하여 공정을 제어하였다[27].

표 5. 환경 시스템 적용.

적용분야	특징	방법 논문
바다의 오염도 측정	클러스터링으로 초기 규칙을 정하고 GA로 최적화	CL [23]
		GA [23]
하폐수 처리장의 산성화 상태	플랜트로부터 얻을 수 있는 센서값을 전문가 지식에 의해 규칙화시켜 퍼지 제어기 설계	ISP [24]
산업폐수처리장의 혐기공정	전문가의 지식에 기반한 퍼지 제어기 개발	ISP [25]
화학공정	센서 프로파일의 강함을 전문가 지식에 의해 규칙화시켜 공정을 진단	ISP [26]
하폐수 처리공정	외란들에 대한 조건을 전문가의 지식에 의해 규칙화	ISP [27]

표 6. 주식 시스템 적용.

적용분야	특징	방법 논문
유전자 퍼지모델링	입력 공간에서 퍼지 클러스터링을 사용하여 규칙 생성	CL [28]
퍼지모델링	입력공간에서 퍼지 클러스터들을 이용하여 규칙을 구성	CL [29]
Tehrna지수	예측신경망의 '층을 규칙으로 표현	NN [30]
주식매매의사결정지원시스템	신경망의 각 층이 규칙을 표현	NN [31]
주식시장 timing 시스템	전문가 지식에 기반한 규칙 구성	ISP [32]
뉴로-퍼지 시스템	신경망의 2-3층의 링크가 퍼지 규칙을 표현.	NF [33]

3.5 주식 시스템

주식 시스템에서는 여러 방법들이 다양하게 사용되고 있다. D. D. Ettes는 주식의 평가를 위해 Singleton 퍼지 모델과 TS-퍼지 모델을 예측 모델로 하여 모델을 유전자 알고리즘으로 최적화하는 시스템을 개발하였다. 시뮬레이션을 위해 Dutch AEX의 20개 대표주의 매일의 가격 데이터를 사용하였다[28]. M. Setnes *et al.*은 TS-퍼지 모델을 이용한 주식 시장의 분석에 대한 연구내용을 발표하였다. Dutch AEX-주가지수의 모델링에 관하여 다른 두 모델 구조(Scenario model, Prediction model)가 사용되었고, 퍼지 클러스터링을 사용하여 언어적 규칙을 표현하였다[29]. H. Kgaloozadeh *et al.*은 이란(Iran) 주식시장의 장기적 예측을 목적으로 신경망을 사용한 결과를 선보였다. 성능은 ARIMA 모델과 비교하여 나타냈다[30]. S. C. T. Chou *at al.*은 규칙에 기반한 신경망을 사용한 장기, 단기 매매의 예측에 따른 매매 신호를 예측할 수 있는 지능적인 거래 지원 시스템을 제안하였다. 신경망의 초기 구조를 구성하는데 추론의 형태로 전문가 지식을 사용하였다[31]. S. S. Lam은 본 연구에서 퍼지 전문가 시스템에 유전자 알고리즘을 접목한 새로운 주식 시장 timing 시스템을 개발하였다. 유전자 알고리즘은 퍼지 매매 규칙의 최적화에 사용되었다. 적당한 퍼지 규칙이 유전자 알고리즘을 기반으로 한 정보로부터 선택된다[32]. Z. Pan *at al.*은 금융 모델링과 예측을 위하여 뉴로-퍼지 시스템을 사용하였다. 신경망의 2-3층의 링크가 규칙을 표현하며 역전파를 통해 학습되었다[33].

4. 결론

기업환경의 급속한 변화와 해결하고자 하는 시스템의 문제영역이 점차 확대되고 복잡해짐에 따라 데이터를 분석하여 자동으로 지식을 추출하는 방법에 대한 요구가 급속히 증가하고 있다. 학습능력을 지닌 프로그램의 능력은 컴퓨터의 가공할 만한 정보처리 능력뿐 아니라 정보통신 기술의 정보 유통 기능과 연계하여 대량의 데이터를 실

시간에 분석하고, 이를 통해 시스템의 효율적인 관리와 최적화를 위한 방법들이 지속적으로 연구되고 있다. 이러한 추세에 맞추어 현재 지식 추출 분야에 활용되는 방법들의 현황과 그 적용분야에 대해 알아보았다. 결론적으로 전문가의 지식기반 시스템에선 공간분할과 클러스터링 등의 방법이 많이 사용되었고 조건에 따른 시스템의 결과가 확연히 드러나지 않는 경우에는 신경망등을 이용하여 정해진 대략적인 규칙을 유전자 알고리즘으로 최적화하는 방법이 주로 사용되었으며 두 가지의 경우가 함께 사용되기도 하였다.

참고문헌

- [1] Jay Liebowitz, *Knowledge Management Handbook*. pp. 1-3. CRC Press, 1999.
- [2] Michael J. A. Berry, *Data Mining Techniques*. pp. 5-7. John Wiley & Sons, Inc. 1997.
- [3] Nadine Tschichold-Gurman, "The neural network model RuleNet and its application to mobile robot navigation," *Fuzzy Sets and Systems* 85, pp.287-303, 1997.
- [4] Ken Nozaki *et al.*, "A simple but powerful heuristic method for generating fuzzy rules from numerical data," *Fuzzy Sets and Systems* 86, pp.251-270, 1997.
- [5] C. T. Lin, *Neural Fuzzy Control Systems with Structure and Parameter Learning* pp.8-10, 16-20. World Scientific 1994.
- [6] Sigeru Omatu *et al.*, *Neuro-Control and its Applications* pp. 7-26, Springer 1995.
- [7] Yan Shi *et al.*, "An improvement of neuro-fuzzy learning algorithm for tuning fuzzy rules," *Fuzzy Sets and Systems* 118, pp. 339-350, 2001.
- [8] X. Z. Wang *et al.*, "A new approach to fuzzy rule generation-fuzzy extension matrix," *Fuzzy Sets and Systems* 123, pp. 291-306, 2001
- [9] Abe, S. *et al.*, "A Function Approximator Using Fuzzy Rules Extracted Directly From Numerical Data," *Proceedings of 1993 International Joint Conference on Neural Networks*, vol. 2, pp. 1887-1892, 1993.
- [10]Pal, N. R. *et al.*, "Fuzzy rule extraction from ID3-type decision trees for real data," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 31, no. 5, pp. 745-754, 2001.
- [11]Zong-Mu Yeh, "A cross-coupled bistage fuzzy controller for biaxis servomechanism control," *Fuzzy Sets and Systems* 97, pp. 265-275, 1998.
- [12]Daijin Kim *et al.*, "An optimal design of neuro-FLC by Lamarckian co-adaptation of learning and evolution," *Fuzzy Sets and Systems* 118, pp. 319-337, 2001.
- [13]Hyun-Jun Cho *et al.*, "Fuzzy-PID hybrid control-Automatic rule generation using genetic algorithms," *Fuzzy Sets and Systems* 92, pp. 305-316, 1997.
- [14]A. Lekova *et al.*, "Redundant fuzzy rules exclusion by genetic algorithms," *Fuzzy Sets and Systems* 100, pp. 235-243, 1998.
- [15]Byung-Jae Choi *et al.*, "Design of a single-input fuzzy logic controller and its properties," *Fuzzy Sets and Systems* 106, pp. 299-308, 1999.
- [16]Song, F. Smith *et al.*, "A general cell state space based TS type fuzzy logic controller automatic rule extraction and parameter optimization algorithm," *The 25th Annual Conference of the IEEE Industrial Electronics Society*, vol. 3, 1999.
- [17]Nelles, O, Fischer *et al.*, "Fuzzy rule extraction by a genetic algorithm and constrained nonlinear optimization of membership functions," *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*, vol. 1, pp. 213-219, 1996.
- [18]M. Setnes, "Supervised fuzzy clustering for rule extraction," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 4, pp. 416-424, 2000.
- [19]Samuel H. Huang *et al.*, "Extract intelligible and concise fuzzy rules from neural networks," *Fuzzy Sets and Systems* 132, pp. 233-243, 2002.
- [20]S. Chiu, "Method and software for extracting fuzzy classification rules by subtractive clustering," *1996 Biennial Conference of the North American Fuzzy Information Processing Society*, pp. 461-465, 1996.
- [21]Thomas Pfeufer *et al.*, "Application of a hybrid neuro-fuzzy system to the fault diagnosis of an automotive electromechanical actuator," *Fuzzy Sets and Systems* 89, pp. 351-360, 1997.
- [22]Mihiar Ayoubi *et al.*, "Neuro-fuzzy systems for diagnosis," *Fuzzy Sets and Systems* 89, pp. 289-307, 1997.
- [23]M. Cococcioni *et al.*, "Automatic extraction of fuzzy rules from MERIS data to identify sea water optically active constituent concentration," *2002 Annual Meeting of the North American Fuzzy Information Processing Society*, pp. 546-551, 2002.
- [24]E. F. Carrasco *et al.*, "Diagnosis of acidification states in an anaerobic wastewater treatment plant using a fuzzy-based expert system," *Control Engineering Practice*, Available online 28, 2003.
- [25]Steyer, J.-P *et al.*, "Fuzzy control of an anaerobic digestion process for the treatment of an industrial wastewater," *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, vol. 3, pp. 1245 -1249, 1997.
- [26]Sourabh Dash *et al.*, "Fuzzy-logic based trend classification for fault diagnosis of chemical processes," *Computers & Chemical Engineering*, vol. 27, issue. 3, pp. 347-362, 2003.
- [27]A. Muller S *et al.*, "Fuzzy control of disturbances in a wastewater treatment process," *Water Research*, vol. 31, issue 12, pp. 3157-3167, 1997.
- [28]Drs. Dennis Ettles, "Trading the stock markets using Genetic Fuzzy Modeling" *CIFER*, pp. 22-25, 2000.
- [29]M. Setnes *et al.*, "Fuzzy Modeling in Stock-Market Analysis," *Proceedings of the IEEE/IAFE 1999 Conference on Computational Intelligence for Financial Engineering*, pp. 250 -258, 1999.
- [30]Hamid Kgaloozadeh *et al.*, "Long Term Prediction of Tehran Price Index(TEPIX) Using Neural Networks" *2001 Joint 9th IFSA World Congress and 20th NAFIPS International conference*, vol. 1, pp. 25-28, 2001.
- [31]Seng-Cho Timothy Chou *et al.*, "A Rule-Based Neural Stock Trading Decision Support System," *Proceedings of the IEEE/IAFE 1996 Conference on Computational Intelligence for Financial Engineering*, pp. 148-154, 1996.
- [32]S. S. Lam, "A Genetic Fuzzy Expert System for Stock Market Timing," *Proceedings of the 2001 Congress on Evolutionary computation*, vol. 2, pp. 410-417, 2001
- [33]Zouhong Pan *et al.*, "A Neural - Fuzzy System for Forecasting," *Proceedings of the Thirtieth Hawaii International Conference on System Sciences*, vol. 5, pp. 549-558, 1997.