

확률적 퍼지 룰 기반 학습 시스템의 적응 방법

Adaptation Methods for a Probabilistic Fuzzy Rule-based Learning System

이형욱¹, 변증남²

¹ 한국과학기술원 인간친화 복지 로봇시스템 연구센터

E-mail: helee@ctrsys.kaist.ac.kr

² 한국과학기술원 전자전산학과

E-mail: zbien@ee.kaist.ac.kr

요 약

지식 발견(knowledge discovery)의 관점에서, 단기간 동안 취득된 데이터 패턴을 학습하고자 하는 경우 데이터에 비일관적인(inconsistent) 패턴이 포함되어 있다면 확률적 퍼지 룰(probabilistic fuzzy rule) 기반의 지식 표현 방법 및 적절한 학습 알고리즘을 이용하여 효과적으로 다룰 수 있다. 하지만 장기간 동안 지속적으로 얻어진 데이터 패턴을 다루고자 하는 경우, 데이터가 시변(time-varying) 특성을 가지고 있으면 기존에 추출된 지식을 변화된 데이터에 활용하기 어렵게 된다. 때문에 이러한 데이터를 다루는 학습 시스템에는 패턴의 변화에 맞추어 갈 수 있는 지속적인 적응력(adaptivity)이 요구된다. 본 논문에서는 이러한 적응성의 측면을 고려하여 평생 학습(life-long learning)의 관점에서 확률적 퍼지 룰 기반의 학습 시스템에 적용될 수 있는 두 가지 형태의 적응 방법에 대해서 설명하도록 한다.

Key Words : Adaptation, Life-long Learning, Probabilistic Fuzzy Logic

1. 서 론

데이터베이스(database)에서의 지식 발견은 데이터로부터 의미 있고 유용하며 궁극적으로 이해 가능한 패턴을 얻어내는 논트리비얼 과정(non-trivial process)로 간단히 정의된다[1]. 다른 측면에서 보면, 지식 발견은 입출력 데이터로부터 입출력 관계(I-O relation)를 모델링 하는 것으로 볼 수도 있다. 이 때, 지식의 형태는 응용 분야에 따라 여러 가지 형태로 정의가 될 수 있으며, 적절한 알고리즘을 사용하여 입출력 관계의 모델을 알아내는 것을 학습 과정(learning process)으로 볼 수 있다.

일반적으로 학습 시스템의 선정 및 지식 표현 방법, 그리고 그에 대한 학습 알고리즘은 학습 대상(learning target)에 대한 분석 및 요구 조건에 따라 달라진다. 예를 들어, 신경망(neural network), 베이시안(Bayesian) 네트워크와 같은 학습 시스템들은 각각 수치화된 가중치(weight), 상태(state)간의 천이(transition)

확률 등으로 학습된 지식을 표현한다. 신경망은 내부 구조를 분석하기 힘든 단점이 있지만, 일반적인 비선형 함수의 학습 및 패턴 인식 분야에 효과적으로 적용되어 왔고, 베이시안 네트워크의 경우 상태 분할이 올바르게 이루어졌을 경우, 주어진 데이터에 대한 확률적인 특성을 효과적으로 모델링 할 수 있다는 장점을 가지고 있다.

하지만, 다루고자 하는 데이터가 그림 1과 같이 비일관적인 데이터 패턴을 가지고 있을 경우, 확률적 퍼지 룰 기반의 지식 표현 방법은 매우 효과적으로 사용될 수 있다[2].

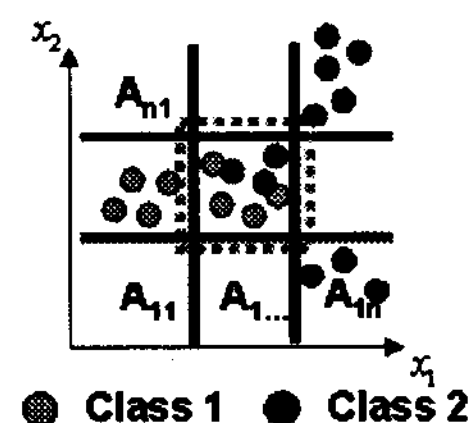


그림 1. 비일관적인 데이터 패턴의 예

특히, 그림 2와 같이 다루고자 하는 데이터에 분리 가능한 데이터와 분리 불가능한 데이터가 섞여 있을 경우, 클래스 분리도(class separability)에 근거하여 군집화 해석(cluster analysis)[3]을 수행하면 보다 의미 있는 지식을 추출해 낼 수 있다. 다시 말하면, 하나의 학습 시스템에서 분리 가능한 데이터들은 독립적인 클로로 만들고, 분리 불가능한 데이터 군은 확률 분포로 해석하는 것이 가능해진다. 이러한 관점에서 IFCS(Iterative Fuzzy Clustering with Supervision) 학습 알고리즘이 제안되었으며, 이것은 주어진 데이터 집합으로부터 반복 학습 과정을 통해 의미 있는 확률적 퍼지 룰 베이스(PFRB; Probabilistic Fuzzy Rule Base)를 추출해 낸다[4].

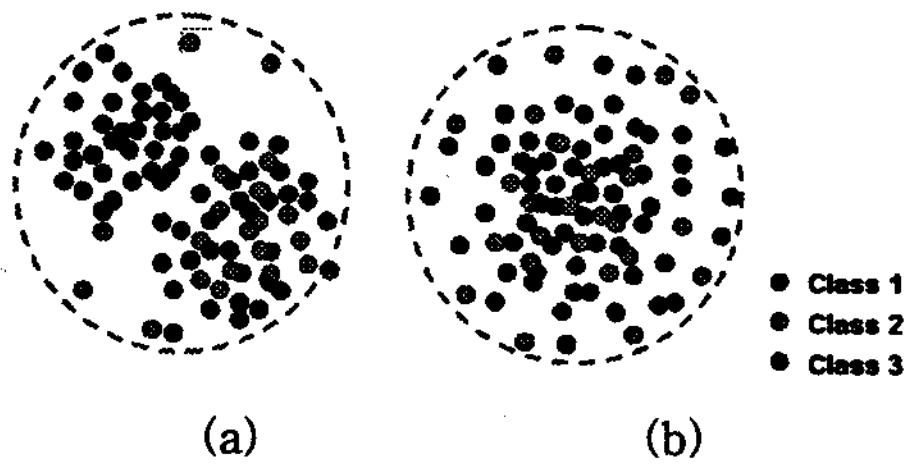


그림 2. 분리 가능한 패턴과 분리 불가능한 데이터 패턴의 예

하지만, 주어진 데이터 패턴이 시변 특성을 가지고 있을 경우 데이터의 변화에 맞추어 지속적으로 얻어진 지식을 수정/보완할 수 있는 적응성이 요구된다. 이러한 관점에서 평생 학습의 개념이 제안되었으며 평생 학습은 지속적인 적응과는 달리 기존의 지식과 충돌하지 않는 한 기존의 지식을 보존하는 개념도 포함한다[5]. 본 논문에서는 특히 평생 학습의 개념을 HiLS(Human-in-the-loop System)의 관점에서 귀납적 학습 과정과 연역적 학습 과정이 반복되는 지식의 축적과정으로 정의한다[2].

제 2 절에서는 이러한 평생 학습 시스템의 관점에서 확률적 퍼지 룰 기반의 학습 시스템 구조에 대해서 설명하도록 하고, 제 3 절에서는 이러한 학습 시스템에 사용될 수 있는 두 가지 형태의 적응 방법에 대해서 설명하도록 한다. 끝으로, 제 4절에서의 실험결과와 결론을 끝으로 논문을 마치도록 한다.

2. 확률적 퍼지 룰 기반 평생 학습 시스템

HiLS의 관점에서 퍼지 룰 기반의 평생 학습 시스템의 구조는 그림 3과 같이 표현될 수 있으며, 확률적 퍼지 룰 기반의 지식 표현 방법을 이용할 경우, 퍼지 룰 베이스는 확률적 퍼지 룰 베이스로 치환된다. 또한, IFCS 학습 알고리즘과 연동하면, 그림 4와 같은 구조를 가지는 학습 시스템으로 만들 수 있다.

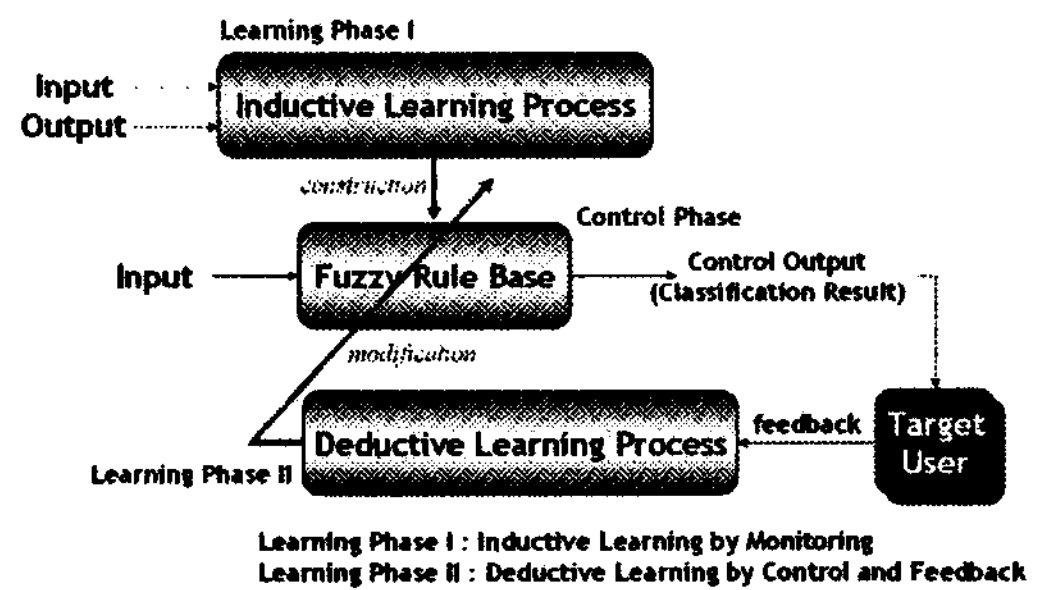


그림 3. HiLS의 관점에서 평생 학습 시스템의 구조

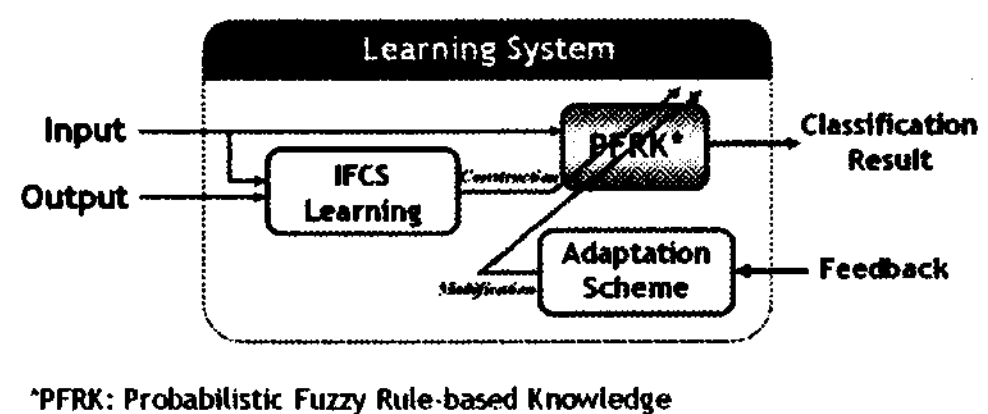


그림 4. 확률적 퍼지 룰 기반 평생 학습 시스템

이 때, 확률적 퍼지 룰 기반 지식(PFRK; Probabilistic Fuzzy Rule-based Knowledge)은 PFRB들의 집합으로 볼 수 있으며, 하나의 PFRB는 하나의 확률적 퍼지 모델(model)로 설명되기도 한다. IFCS 알고리즘이 주어진 데이터 집합으로부터 하나의 PFRB를 만들어 내는 것을 고려하면, 지속적으로 들어오는 데이터 패턴에 대해서 학습시스템은 일정 시간 또는 일정 데이터 수 간격으로 PFRB를 만들어 내게 된다. 또한, 만들어지는 PFRB는 기존에 보유하고 있던 PFRB와의 비교를 통하여 새로운 PFRB를 반영하는 방법을 결정하게 된다. 본 논문에서는 이러한 적응 방법을 모델 적응(model adaptation)이라고 부르기로 한다. 또한 보다 빠른 적응을 위해서는 학습 데이터 패턴 하나하나에도 지속적으로 PFRK를 수정할 수 있는 적응 방법 또한 요구되며 이것을 온라인 적응(on-line adaptation)이라고 부르기로 한다. 제 3절에서는 이러한 두 가지 적응 방법에 대해 보다 자세히 설명을 하기로 한다.

3. 확률적 퍼지 룰 기반 학습 시스템을 위한 적응 방법

모델 적용은 모델과 모델간의 적용으로 생각할 수 있고, 온라인 적용은 학습 데이터 패턴과 모델간의 적용으로 생각할 수 있다. 적용 과정은 새로운 정보를 기존의 정보에 병합(merge)할 것인지 아니면 추가(addition)을 하게 될 것인지를 결정하는데, 이러한 과정은 그림 5에서 보는 것과 같이 비교 대상간의 유사도(similarity)를 기준으로 결정하게 된다.

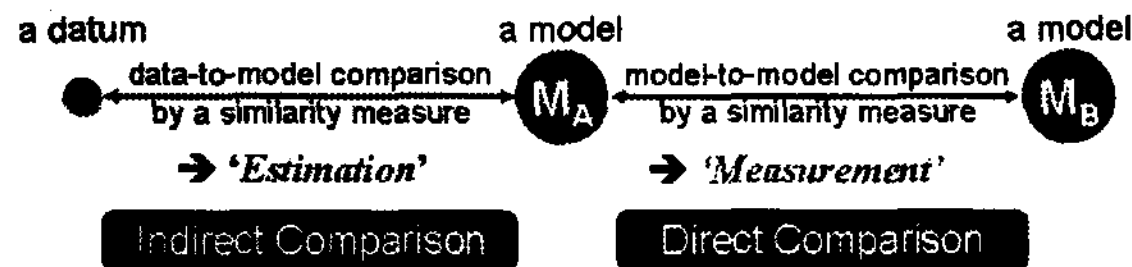


그림 5. 유사도 비교 방법

그림 5에서 보는 것과 같이 모델과 모델간의 유사도는 적절한 유사 측정도(similarity measure)를 정의함으로써 대상을 직접 비교할 수 있다. 하지만, 데이터 패턴과 모델의 유사도는 직접적인 비교 방법이 없기 때문에 추정(estimation)을 통하여 간접 비교를 수행해야 한다는 차이점을 가지고 있다.

3.1 모델 적용 방법

모델과 모델간의 유사도 비교는 Zadeh의 sup-min compatibility[6]를 기본으로 하여 두 개의 PFRB간의 유사도를 정의하여 사용한다. 먼저 두 개의 확률적 퍼지 룰 R_i, R_j 에 대해서 퍼지 유사도 $SimFS$ 는 다음과 같이 정의된다.

$$SimFS(R_i, R_j) = \sup_{u \in U} \min(\mu(R_i), \mu(R_j)) \quad (1)$$

여기서 확률적 퍼지 룰의 표현식은 일반적인 퍼지 룰의 표현식에서 결론절이 구분하고자 하는 클래스의 확률값을 갖도록 한다[2][4]. 이로부터 확률적 유사도 $SimPB$ 는 다음과 같이 정의된다.

$$SimPB(R_i, R_j) = \sum_{k=1}^K \min(P_i^k, P_j^k) = \sum_{k=1}^K \min(\Pr(y = c^k | x = \tilde{A}_i), \Pr(y = c^k | x = \tilde{A}_j)) \quad (2)$$

식 (1)과 식 (2)로부터, 두 개의 확률적 퍼지 룰 사이의 확률적 퍼지 유사도 $SimPBFS$ 는 다음과 같이 정의할 수 있다.

$$SimPBFS(R_i, R_j) = SimFS(R_i, R_j) \cdot SimPB(R_i, R_j) \quad (3)$$

또한 식(1)과 식(3)로부터, 두 개의 PFRB M_A 와 M_B 사이의 유사도는 확률적 퍼지 유사도 $SimPBFRB$ 로 다음과 같이 정의할 수 있다.

$$SimPBFRB(M_A, M_B) = \frac{\sum_{j=1}^{N_A} \{ (\sum_{i=1}^{N_B} SimPBFS(R_i^A, R_j^B)) / \sum_{i=1}^{N_B} SimFS(R_i^A, R_j^B) \}}{N_A} \quad (4)$$

여기서 N_A 와 N_B 는 각 PFRB의 룰의 수를 의미한다. 식 (4)에서 정의된 유사도를 이용하면 두 개의 모델 사이에 직접 비교가 가능하게 되며, 이를 이용하여 유사도가 높을 경우는 기존의 모델에 병합과정을 거치게 되고, 유사도가 낮을 경우는 새로운 모델로 간주하여 PFRK에 추가로 저장을 하게 된다. 단, 실제 시스템을 구현하는 경우에는 메모리의 제약이 있을 수 있으므로, 제한된 수의 모델에 대해서 모델간의 유사도가 최소화되도록 적용 과정을 수행하면 효율적인 PFRK를 구성할 수 있게 된다.

3.2 온라인 적용 방법

온라인 적용 방법은 PFRK 안에 다수의 PFRB가 있을 경우 현재의 데이터 패턴에 가장 적합한 PFRB를 선택하고 그 중에 가장 유사한 확률적 퍼지 룰을 선택하여 병합하는 과정을 말한다. 이 때, 주어진 데이터 패턴은 퍼지 싱글톤 방법[7]을 이용하여 퍼지 룰로 변환하게 되고, 식 (1)의 $SimFS$ 를 사용하면 PFRB 안에 가장 유사한 퍼지 룰을 찾을 수 있다.

하지만, 하나의 학습 데이터 패턴으로부터 만들어진 퍼지 룰은 확률적 정보가 없으므로 다수의 데이터 패턴으로부터 추출된 확률적 정보를 가지고 있는 PFRB와 직접적인 비교를 할 수가 없다. 그렇기 때문에, 지속적으로 들어오는 데이터 패턴에 대해서 어떤 모델이 가장 현재의 패턴을 잘 설명하고 있는지에 대한 추정치를 사용해야 한다. 이를 위해서 본 논문에서는 PFRK 안에 n 개의 PFRB가 있다고 할 경우, 다음과 같이 모델 추정기를 회귀 형태로 만들어 사용하게 된다. 시간 t 에 들어오는 데이터 패턴을 $(x(t), c^k(t))$ 라고 하고, $x(t)$ 를 입력 벡터, $c^k(t)$ 를 클래스 레이블이라고 하자. 또한 PFRK $M_B = \{M_{B1}, M_{B2}, \dots, M_{Bn}\}$ 와 같이 확률적 퍼지 모델 집합으로 구성되었다고 하자 (N_{B_i} 는 M_{B_i} 의 룰 수). 이 경우, M_{B_i} 의 시간 t 에서의 모델 추정치 $M_i(t)$ 는 다음과 같은 수식을 갖는다.

$$M_i(t) = 1 - e^{-\gamma(\sum_{i=0}^t m_i(t) - \frac{1}{\gamma} \ln(1 - M_i(t-1)))}$$

$$m_i(t) = \max_{i=1, \dots, N_{B_i}} \mu_i(x(t)) P_i(c^k(t)|x(t)), \quad (5)$$

어떤 시점에서 모델 추정치의 값이 가장 높다는 것은 현재 학습 시스템으로 들어오고 있는 패턴이 해당 모델로 가장 잘 설명이 된다는 것을 의미하기 때문에, 온라인 적응을 수행할 때는 식 (6)을 따라 모델을 선택한다.

$$i^*(t) = \arg \max_{i=1, \dots, n} M_i(t) \quad (6)$$

4. 실험 결과 및 결론

IFCS 학습 알고리즘은 UCI repository of Machine Learning Database의 벤치마크 데이터 패턴(예. Iris, Wine, Breast-Cancer)을 이용했을 경우뿐만 아니라 비일관적인 데이터의 비율이 많은 경우에도 주어진 데이터로부터 효과적으로 확률적 퍼지 룰 기반을 추출해 낼 수 있음을 보인바 있다[4]. 하지만 보다 많은 정도의 비일관적인 데이터 패턴을 가지고, 동시에 시변 특성을 가지고 있는 경우를 분석하기 위해서 TV 시청 패턴을 이용하였다. 실험을 위해 25명의 1년 동안 누적된 TV 시청 패턴에서 선호도가 높은 장르를 선택하는 패턴 분류의 문제를 고려하였다. 여기서 학습 주기는 2주 간격 또는 한달 간격을 고려하였다.

표 1. 실험 결과

CASE	Avg. Test Success Rate(%) for Genre Selection	
	2 weeks	1 month
No Adaptation	58.14	59.37
Model Adaptation only	65.04	63.82
On-line Adaptation only	73.97	72.30
Model Adaptation + On-line Adaptation	76.73	74.09
Self-Test by IFCS	92.87	88.85

표 1의 실험 결과에서 보는 것과 같이, 학습한 데이터 패턴으로 자가 테스트를 하게 되면 90%에 가까운 성공률을 보이지만, 적응 과정을 거치지 않았을 경우 기존에 학습한 지식을 다음 기간에 적용하면 성공률이 매우 낮아지는 것을 알 수 있다. 하지만 모델 적응과 온라인 적응 방법을 병행함으로써 성공률을 증가시킬 수 있음을 알 수 있다.

실험 결과에서 보게 되면, 모델 적응 방법의 경우 온라인 적응에 비하여 높은 성공률의

증가를 보이지 않고 있지만, 데이터 패턴이 반복해서 들어올 경우는 온라인 적응보다 더 높은 성공률의 증가를 보일 수 있다.

일반적인 학습 시스템은 패턴 인식의 관점에서 벤치 마크 데이터 패턴과 같이 클래스간 분리도가 높은 데이터 패턴에 대해서 주로 연구가 이루어졌다. 하지만, 사람의 행동 패턴 또는 생체 신호와 같이 비일관적인 데이터 패턴이 많고 시변 특성을 가지고 있는 데이터 패턴을 다루는 경우는 정적인 패턴 분류뿐만 아니라 변화에 대처할 수 있는 학습 시스템의 적응력에도 중점을 두어야 한다.

이러한 학습 시스템을 서비스 로봇 분야에 적용할 경우 행동 패턴 인식 등으로 사람의 의도 파악 기능을 구현할 수 있고, 로봇의 지능 향상에도 기여를 할 수 있을 것이라 판단된다.

참 고 문 헌

- [1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview", in *Advances in Knowledge Discovery and Data Mining*, MIT Press, Cambridge, Mass., pp. 1-36, 1996.
- [2] Z. Zenn Bien and Hyong-Euk Lee, "Effective Learning System Techniques for Human-Robot Interaction in Service Environment", *Knowledge-Based Systems*, vol. 20, issue 5, pp. 439-456, 2007.
- [3] Nikhil R. Pal and J. C. Bezdek, "On Cluster Validity for the Fuzzy c-Means Model", *IEEE Trans. Fuzzy Systems*, vol. 3, No. 3, pp. 370-379, 1995.
- [4] Hyong-Euk Lee, Kwang-Hyun Park, and Z. Zenn Bien, "Iterative Fuzzy Clustering Algorithm with Supervision to Construct Probabilistic Fuzzy Rule Base from Numerical Data", *IEEE Trans. Fuzzy Systems*, accepted, 2007.
- [5] S. Grossberg, "Nonlinear neural networks: principles, mechanisms and architectures", *Neural Networks*, vol. 1, no. 1, pp. 17-61, 1988.
- [6] V. V. Cross and T. A. Sudkamp, *Similarity and Compatibility in Fuzzy Set Theory*, New York; Physica-Verlag, 2002
- [7] L. A. Jadeh, "Soft computing and fuzzy logic", *IEEE Software*, vol. 11, no. 6, pp. 48-56, 1994.