

확장된 논리적 진화방식을 이용한 개념학습

박명수, 최진영
서울대학교 전기컴퓨터공학부

Learning of concept with real-valued input with modified Logical Evolution method

Myoung Soo Park and Jin Young Choi
School of Electrical Engineering & Computer Science,
Seoul National University
E-mail : mspark@neuro.snu.ac.kr

Abstract

본 논문에서는 다수의 개념을 효과적으로 학습할 수 있는 논리적인 진화방식을, 보다 일반적인 문제에 적용하기 위하여 확장하였다. 기존의 알고리즘이 0 차 논리(zero-order logic)에 의해 표현되는 개념만을 학습할 수 있었던 점을 개선하여, 특징논리로 표현되는 개념까지 학습할 수 있도록 확장하였다. 또한 확장된 알고리즘을 이용하여, 실수입력을 가지는 개념 또한 학습할 수 있도록 하였다. 제안된 알고리즘은 이전의 알고리즘과 마찬가지로, 다수의 개념을 학습할 때, 이전의 학습 결과를 이용하여 새로운 개념을 쉽게 학습할 수 있는 특징을 가지고 있다. 제안된 알고리즘에 의해 특징논리로 표현되는 개념을 학습할 수 있으며, 다수의 개념을 학습할 때 성능향상이 이루어지는 것을 실험을 통해 확인하였다.

1. 서론

논리적인 진화방식(Logical Evolution method)[1]은 다수의 이진개념(binary concepts)들을 효율적으로 학습할 수 있도록 설계된 개념학습 알고리즘이다. 주어진 자료로부터 자료를 잘 기술하는 개념을 학습할 수 있을 뿐만 아니라, 다른 자료로부터 새로운 개념을 학습하게 될 경우에, 이전에 학습된 개념들을 이용하여 새로운 개념을 보다 효율적으로 학습할 수 있다. 그럼으로써, 이전의 학습 결과를 유지하면서 새로운 학습결과를 더하는 좁은 의미의 누적학습(Incremental learning) 뿐만

아니라, 이전의 학습결과를 새로운 학습에 능동적으로 이용하는 넓은 의미의 누적학습까지도 가능해졌다.

이러한 성능은, 기계학습분야에서 이전까지 연구된 두 가지 방식(divide and conquer, separate and conquer)의 개념학습 알고리즘들[3]과의 차이점에서 비롯한다. 학습을 위해 자료집합에 포함된 다수의 예(example)들을 동시에 이용하는 대신에, 논리적인 진화방식에서는 하나하나의 개별적인 예(individual example)들을 이용하여 학습이 이루어진다. 그럼으로써, 얻어지는 정보가 다른 문제의 해결, 즉 새로운 개념의 학습을 위해서 바로 이용될 수 있다. 자료가 시간차를 두고 조금씩 차례로 주어지는 경우에도 이용될 수 있다.

이러한 성질을 가지도록 하기 위해서, 논리적인 진화방식은 기본적인 단계에서부터 새로이 개발되었다. 그 결과 추가적인 연구가 필요한 부분을 많이 가지고 있다[2]. 첫째, 논리적인 진화방식은 일관성(consistency), 완결성(completeness)을 가지는 개념을 학습하는 것을 목표로 개발되었다. 학습된 개념이 자료를 잘 기술할 뿐만 아니라, 학습되지 않은 자료에 대한 일반화 성능을 가질 수 있도록, 정확성(accuracy), 불확실성에 대한 강인성(robustness) 등에 대한 추가적인 개선이 필요하다. 둘째, 논리적인 진화방식은 이진입력에 대해 정의되는 개념만을 학습할 수 있다. 개념의 표현(representation)을 0 차 논리(zero-order logic)에 한정하고 있으며, 보다 폭넓은 실제적인 응용을 위해서는 특징 논리(attributional logic), 1 차 논리(first/second-order logic) 표현의 개념까지 학습할 수 있도록 할 필요가 있다. 이 논문에서는, 두 번째 주제에 초점을 맞추어, 논리적인

진화방식이 특징논리에 의해 표현되는 개념까지 학습할 수 있도록 확장하였다. 확장된 알고리즘을 이용하면 실수입력을 가지는 개념까지 학습할 수 있다.

이 논문에서는 우선, 논리적인 진화방식에 대해 소개하였다. 특징논리로 표현된 개념을 학습할 수 있도록 확장된 논리적인 진화방식을 설명하고, 기술된 알고리즘을 이용하여 실수값을 입력으로 하는 개념에 을 학습한 결과를 간단히 제시하였다.

2. 논리적인 진화방식

논리적인 진화방식은 이진값을 입력으로 가지는 개념을 학습하기 위해 개발되었다. 모든 학습가능한 개념은 DNF(Disjunctive Normal Form) 형태로 표현될 수 있다. 입력변수들과 입력변수들의 부정(negation)들을 literal 이라고 정의할 때, 이진함수들은 literal 들을 AND 결합한 항들의 OR 결합으로 표현될 수 있다. 따라서, 그러한 개념을 구하는 것은 포함되는 항들을 모두 구하는 것과 같다. 입력이 주어지면 항들의 값을 구하고, 그 값들에 OR 를 취함으로써 개념에의 소속 여부를 결정할 수 있는 것이다. 논리적인 진화방식을 학습을 위해 그러한 항들을 찾는다.

항들은 항을 구성하는, OR 로 결합된 입력변수들의 상태에 의해 결정된다. 입력변수는 단순히 포함되는 경우와, NOT 연산을 통하여 결합되는 경우와, 항에 포함되지 않는 세 가지 상태를 가진다. 예를 들어

입력변수가 x_1, x_2, x_3 인 경우, 항 $x_1 \wedge \neg x_2$ 에서 x_1 은 첫번째 상태에, x_2 은 두번째 상태에, x_3 은 세번째 상태에 있다. 따라서, 모든 항은 변수의 개수가 n 개일 때 n 차원의 벡터로 표현이 가능하다. 논리적인 진화방식은 벡터의 형태로 항들을 저장하고, 그러한 벡터들을 결합하거나 추가하면서 주어진 자료를 기술할 수 있는 벡터들의 집합을 찾는다. 찾아낸 집합은 주어진 자료집합을 일관성과 완결성을 가지고 기술하는 개념에 해당한다. 그 과정이 아래 그림에 주어져 있다.

목표로 하는 개념에 대한 가설(hypothesis)는 두 개의 집합 중에서 아래 원에 포함된 항들의 집합에 해당한다. 그것들은 입력변수 각각에 해당하는 10 개의 칸으로 되어있고, 그 안에는 상태가 표시되어있다. 위쪽 원에 포함된 항들의 집합은 이전의 학습결과로, 그 중에서 적절한 것이 선택되어, 아래쪽 원에 포함된 항들을 보다 개선된 항으로 만드는 데에 이용된다. 그 기준은 적합도로 정의되며, 1) 원하는 출력을 가지고 2) 많은 literal 을 포함할수록 적합도가 크다. 그러한 기준에 의해 선택된 항이, 아래 쪽의 항 들 중에서 1) 가설의 출력 (포함된 항들의 값에 OR 를 취하여 계산되는)과 같고, 2) 많은 literal 을 포함하는 제일 책임있는(responsible) 항을 개선하는 데에 이용된다. 또는, 아래 집합에 새로운 항을 추가해야 하는 경우에 이용된다. 아래쪽 원에 포함된 항을, 위쪽 원에 포함된 항과 결합되어 개선된 항을 만들고 나면, 개선된 항이

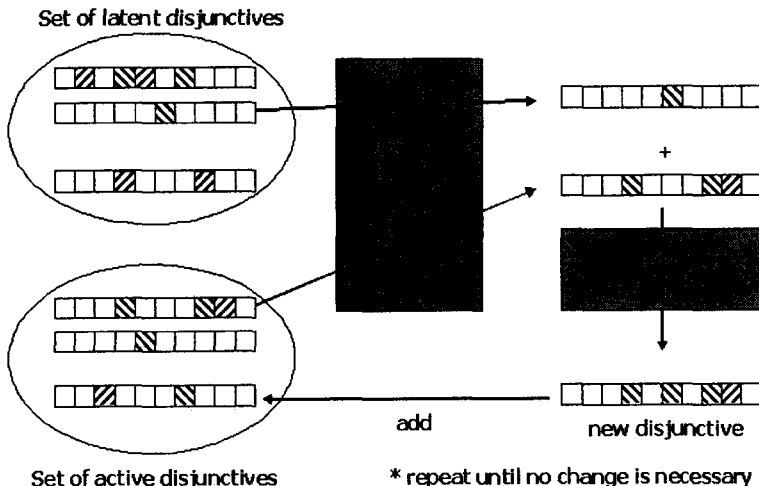


그림 1. 논리적인 진화방식에 의한 개념 학습

아래쪽 원에 포함되어, 이전의 항은 위쪽 원에 포함되어 다른 문제의 해결을 위해 이용될 수 있게 된다. 이러한 요소들을 이용하여, 알고리즘은 오른쪽 순서대로 학습을 진행한다. 학습은 자료에 포함된 모든 입력에 대해 올바른 출력을 낼 때 종료한다.

순서에 포함된 두 가지 연산, 즉 생성(Creating)과 특수화(Specialization)은 학습과정의 핵심적인 요소이다. 생성은, 원 안에 포함된 항들만을 결합하여서는 원하는 출력을 만들 수 없을 때, 망에 새로운 항을 추가하는 것이다. 특수화는, 원 안에 포함된 항들을 선택하고 결합하여, 망에 보다 개선된 항을 추가하는 것이다. 이 과정에서 일관성과 완결성을 보장하기 위하여 아래쪽 원에 포함된 항들에는 추가적인 정보가 포함된다.

포함되는 추가적인 정보는 바로 항을 생성할 때 주어지는 입력에 대해, 값이 T 가 되는 literal 들의 집합을 말한다. 이것을 저장하여 특수화를 할 때, 결합되는 항이 그러한 literal 들로만 구성되었는지를 판단한다. 만약 포함된 literal 로만 구성된 경우는 결합하여 특수화된 항을 만들 수 있다. 그러나, 만약 포함되지 않은 literal 로 구성되었을 경우에는, 특수화를 위해 이용하지 않는다. 만약 모든 항이 특수화에 이용될 수 없다면, 이것은 자료집합 내부에 모순되는 자료가 포함되어있음을 의미한다. 그러면 그러한 자료는 학습될 수 없는 것으로 취급하여, 학습자료에서 제외시킨다.

이상에서 논리적인 진화방식에 대해 간략히 기술하였다. 이제 특징논리(attributional logic)에의 적용을 위해, 특수화와 생성을 자세히 살펴보고 그것을 확장하도록 하겠다. 항에 포함된 추가정보를 이용하여

확장된 알고리즘이 실수입력을 처리할 수 있도록 하는 방법을 기술하겠다.

3. 특징논리로 표현되는 개념을 학습할 수 있도록 확장된 논리적인 진화방식

특징논리와 0 차 논리의 차이점은, 포함된 변수가 가질 수 있는 특징값이 2 개 이상이 될 수 있다는 점이다. 그렇기 때문에, 1) 생성된 항이 저장해야 하는 추가정보와, 2) 항의 선택기준과, 3)항을 생성 및 결합시키는 과정에 차이가 생긴다. 이러한 세 가지 변경과정의 예가 아래의 Table 에 주어져있다

0 차 논리개념을 학습하기 위해 생성된 항은 추가정보로서, 학습에 이용될 수 있는 literal set 을 포함한다. 이것을 부정하는 항이 결합되어, 생성시에 이용된 자료에 대해 잘못된 출력을 내는 것을, 즉 생성시에 학습된 자료를 망각하지 않도록 하기 위해서이다. 이러한 목적을 달성하기 위해, 특징논리개념을 학습하기 위해 생성되는 항은, 자료에 포함되었던 특징값을 포함한다. 두 가지 방식의 차이는 0 차 논리의 경우는 사용될 수 있는 literal 이 추가정보에 포함된 것으로 제한되는데 반하여, 여기서는 추가정보를 배제하지 않는 한도 내에서 literal 이 포함될 수 있다는 점이다. 그럼으로써, 생성시에 이용된 자료를 망각하지 않도록 할 수 있다.

항의 생성과정에서는 이전의 알고리즘이 현재의 입력에 대해 T 가 되는 literal 하나만을 포함하는데 반하여, 새로운 알고리즘에서는 그것과 함께 포함되는 같은 입력변수에 대한 모든 literal 들이 동시에 포함된다.

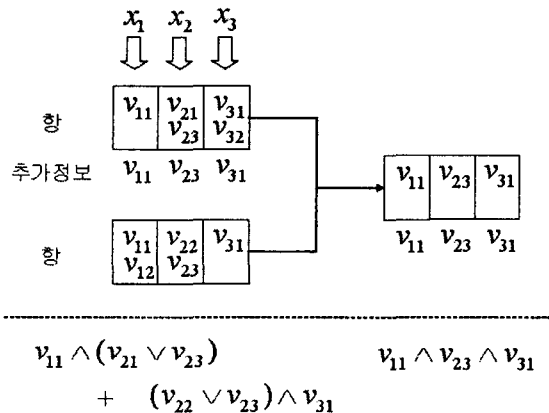


그림 2. 확장된 논리적 진화방식에 의한 항 특수화의 예

그럼으로써, 특수화를 위한 일반적인 항을 포함시킬 수 있다. 그러한 항의 특수화는 포함된 literal 들을 하나하나 제거해나가는 과정이 된다. 그 과정에서 추가정보에 포함된 입력값을 T 로 하는 literal 은 제거되지 않는다.

이러한 생성과 특수화를 위해 항을 선택하는 기준은 다음과 같이 변경된다. 원하는 출력을 낼 수 있는 것 중에서 많은 literal 들을 포함할수록 큰 적합성(fitness)를 가진다는 것이 이전의 기준이다. 새로운 기준은, 각 입력변수에 해당하는 literal 들의 곱을 기준으로 한다. 그럼으로써, 보다 효율적으로 이전의 항들을 개선할 수 있는 항을 찾을 수 있다. 개선해야 하는 항을 선택하는 경우에도 마찬가지로 기준이 이용된다.

이상에서 확장된 방법은, 실수 입력을 가지는 개념의 학습을 위해 이용될 수 있다. 항을 특수화하는 과정에서, 추가정보로 인하여 모든 특수화가 실패할 경우가 발생할 수 있다. 이 경우에는 자료집합 내에 현재의 입력변수들에 의하면, 같은 입력값을 가지고 출력이 다른 자료가 존재한다. 실수가 아닌 값(앞서의 이진값이나, 다수의 특징값)들을 입력으로 가지는 개념의 경우, 이러한 자료는 학습할 수 없다. 일반화를 고려하여 어느 쪽을 선호할 수는 있지만, 그것은 정확한 학습이라고 할 수 없다. 실수입력을 가지는 개념을 학습하는 경우에는, 이러한 경우에 입력공간을 세분화하여 표현할 필요성이 있다고 판단할 수 있다. 실수값을 가지는 입력변수들을 논리적인 변수로 변환하기 위한 양자화(quantization)가 부적절하게 이루어졌다고 보는 것이다. 그리고, 그러한 시점에서 주어진 자료를 기준으로 그것을 포함하는 양자화된 입력영역을 재분할하여 실수변수를 입력으로 하는 개념을 학습할 수 있다.

4. 실험

두 가지 실험을 통하여 알고리즘의 성능을 확인하였다. 첫째로, 확장된 알고리즘이 특징논리에 의해 표현되는 개념과 실수 값을 입력으로 하는 개념들을 학습할 수 있는지를 확인하기 위하여 UCI 에서 제공하는 benchmark 자료를 이용하였다. 전체 자료 중에서 20 개의 자료집합을 이용하여 학습을 하였고, 모든 경우에 학습을 위해 주어진 자료에 대해

일관성과 완결성을 가지는 개념을 학습할 수 있음을 확인하였다. 둘째로, 유사한 문제가 주어진 경우에 이전의 학습결과를 재활용하여 보다 효율적으로 학습하는 특징이 확장된 알고리즘에서도 유지되는지를 확인하기 위하여 인접합 실수영역들에 대해 정의되는 개념을 10 개 정의하고 이에 대한 자료를 임의로 추출하여 학습하였다. 그러한 학습과정을 1000 번 반복하였다. 그 결과, 각각을 학습하는 경우에 비해 이전에 학습된 경우가 증가할수록 개별적인 문제에 대한 학습속도가 빨라지는 것을 관찰하였고, 이를 통해 이전 정보가 재활용되는 것을 확인하였다.

5. 결론

이 논문에서는 다수의 개념을 효과적으로 학습할 수 있는 논리적인 진화방식을, 보다 일반적인 문제에 적용하기 위하여 확장하였다. 특징논리를 다룰 수 있도록 확장하는 방법을 제시하였고, 제시된 방법을 이용하여 실수입력을 가지는 개념 또한 학습할 수 있도록 하였다. 그러한 결과를 실험을 통해 확인하였다. 일반화 성능 등에 대한 추가적인 연구를 통하여, 보다 일반적인 다수의 개념들을 동시에 효과적으로 학습할 수 있도록 할 필요가 있다.

참고문헌

- [1] M. S. Park, J. Y. Choi, C.-H. Choi, "Logical Evolutionary Networks for Inductive Learning," International ICSC Congress on Intelligent Systems & Applications(ISA'2000), eember 12-15, 2000 at the University of Wollongong, Australia
- [2] Myoung Soo Park, Jin Young Choi, "Analysis on the Learning condition of Logical Evolution method," 2001 Korea Brain Society Annual Conference, Soongsil University, Seoul, Korea, July 29, 2001
- [3] Michalski, R.S., "A Review of Machine Learning Methods" in *Machine Learning and Data Mining : Methods and Applications*, John Wiley & Sons, 1996