

과학철학에 있어서 AI의 활용 가능성

이 영 의

고려대학교 철학과

서울 성북 안암 5가 1 (우편번호 136-075)

요약

본 논문은 과학철학을 논의하는데 있어서 AI를 활용하는 가능성을 다루고 있다. AI를 이용한 과학철학은 그것의 논의 절차가 분명하게 표현된다는 점 이외에도 그러한 절차를 테스트할 수 있다는 장점을 갖고 있다.

1. 머릿말

인공지능(artificial intelligence:AI)은 전산학적 모델(computational model)을 이용하여 정신 상태나 그 과정을 연구하는 학문으로 볼 수 있다. AI 연구는 지식 표현(knowledge representation)이나 학습(learning)과 같은 주제를 연구하는데 있어서 전산학적 모델을 이용함으로써 괄목할 만한 성과를 이룩하여 왔으며, 그 결과 철학과 인지 심리학, 신경과학 그리고 언어학 등의 분야에서는 인지과학(cognitive science)의 관점에서 AI의 연구 성과를 적극적으로 수용하여 기존의 문제들을 다루려는 시도를 하고 있다.

이러한 점에서 과학철학(philosophy of science) 역시 예외는 아니다. 과학철학 분야에 나타난 최근의 경향은 그 주요 문제인 이론의 논리적 구조나 추리의 문제, 문제 해결(problem solving) 등에 AI 연구에서 얻어진 기법들을 적용하려는 것인데, 그 결과로 나타난 과학철학은 그것이 제시하는 바를 기계 공학적으로 다룰 수 있으며 그 이론을 구성하고 있는 요소간의 관계도 명백히 밝힐 수 있다는 장점을 지니고 있다.

본 논문에서는 AI의 지식 표현을 이용한 과학적 지식의 표현 방법을 제시하고자 한다. 필자는 이러한 논의를 통하여 철학과 AI연구는 상호간에 도움을 줄 수 있는 측면이 드러날 수 있다고 생각한다.

2. 인공지능과 과학철학

과학철학은 20세기에 들어서 대체로 논리 실증주의의 과학 철학(logical positivist philosophy of science : LPPS)과 역사적 과학철학(historical philosophy of science : HPS)에 의해서 주도되어 왔다고 볼 수 있다. LPPS는 그 명칭이 시사하듯이 고전적 경험론과 현대 기호 논리학을 기반으로 하여 과학을 분석했다. 따라서 LPPS의 주요 관심사는 이론의 논리적 구조나 관찰을 기술하는 명제, 그리고 명제가 확증하거나 반박하는 법칙과 이론 사이에 존재하는 논리적 관계였다. 따라서 그것은 과학적 연구가 어떻게 감각 경험에 근거하여 의미를 지닐 수 있는가를 보여 주는 검증 이론(verifiability)과 특정 사건과 법칙간의 관계를 다룬 확증 이론(confirmation theory) 내지 설명 이론(explanation theory)들을 제시했다.

이러한 LPPS의 분석이 보다 더 엄밀해질수록 실제로 과학적 연구가 수행되어지는 것과 멀어진다는 비판이 제기되어 왔다. Hanson이나 Kuhn과 같은 사람들은 LPPS의 특징인 형식적 엄밀성을 피하고 그대신 역사적인 사례 연구에 기반을 둔 HPS를 제안하였다. HPS는 LPPS가 다루지 않았던 주제들, 즉 과학적 발견(scientific discovery)의 본질이나 과학혁명(scientific revolution)의 구조와 같은 문제들을 다룰 수 있었다. 그리하여 HPS는 과학의 본질에 대한 보다 풍부하고 생생한 설명을 제시한다는 점에서 실제적인 장점을 지니고 있지만, 반면에 엄밀성을 결여하고 있다는 단점도 지니고 있다. 이와 같은 HPS는 형식 논리를 중요시하지 않는다는 점에서 LPPS와 근본적으로 차이가 난다고 볼 수 있다.

그런데 AI에서도 과학철학에서와 마찬가지로 형식 논리에 의한 구분이 적용될 수 있다. AI가 정확히 무엇을 의미하는가는 1950년대에 그 용어가 등장한 이후로 아직까지도 논의의 대상이며, 따라서 그것의 목표와 연구 방법에 있어서도 다양한 견해가 공존한다. 그러나 그 다양성에도 불구하고 지능을 이해하는데 있어 앞에서 언급된 기준에 따라 현재의 AI연구는 대체로 두 진영으로 구분할 수 있다.[12] 그중 하나는 논리를 AI 연구의 가장 중요한 요소로 간주하여 주로 논리적 연역을 처리하는 형식 체계를 구성하는데 주안점을 준다. McCarthy나 Nilsson 등이 중심 인물인 이러한 진영은 “불의 이상(Booleam dream)”을 추구한다는 점에서 LPPS의 후예들이라고 볼 수 있다. 이와는 대조적으로 Minsky와 Schank에 의해서 대표되는 또하나의

진영은 AI 연구가 보다 더 성공하기 위해서는 형식 논리의 엄밀성을 피하고 대신 우리의 사고에서 발견되는 다양한 구조와 과정을 탐구해야 한다고 주장한다. HPS에 대응한다고 볼 수 있는 이러한 진영은 전자보다 훨씬 더 심리학적인 접근 방식을 취하고 있다.

그런데 최근에 과학철학에서는 AI를 이용하여 과학적 지식을 연구하려는 시도들이 나타나고 있다. 이러한 새로운 시도를 하는 사람들은 과학철학 분야에서 앞으로의 연구를 위한 강력한 도구로써 AI의 추리 과정을 사용할 것을 적극적으로 주장한다.[2,12] 따라서 그들은 컴퓨터 모의 실험(simulation)을 사용하여 과학적 발견이 이루어지는 과정을 연구하거나 [2,7,11] 과학적 사례를 분석함으로써 발견을 지배하는 기본 원리를 이끌어내려고 한다. [8]

물론 이러한 연구들은 시도 단계이며, AI의 기법이 아직은 LPPS와 HPS의 문제를 만족스럽게 다룰 수 없기 때문에 그러한 접근 방식에 따르는 문제들이 있을 것이라고 생각된다. 그러나 과학적 추리를 연구하는데 있어서 컴퓨터 모의 실험을 도입함으로써 과학철학이 얻을 수 있는 방법론적 이점을 간과해서는 안될 것이다. 그런데 여기서 말하는 방법론적 이점이란 다음과 같은 것이다.[12] 즉 a) 전산과학은 지식의 구조를 기술하는데 있어 체계적인 어휘들을 제공한다. b) 프로그램에 의해 어떤 가설을 수행하여 그것의 내적 정합성을 테스트할 수 있다. c) 프로그램을 실행시킴으로써 그 가설이 참인지 거짓인지를 알 수 있다.

이러한 관점에서 Thagard는 새로운 과학철학을 Minsky 계열의 AI와 HPS의 교차점으로 보고, 그것을 전산학적 과학철학(computational philosophy of science:CPS)이라 명명했다.[12] 처음에는 LPPS가 McCarthy 계열의 AI에 영향을 주었지만 이제는 역으로 과학철학이 AI의 영향을 받아 그 결과 CPS가 나타나게 되었다.

3. 과학적 지식의 구조

AI에서 말하는 지식 표현이란 데이터 구조(data structure)와 알고리즘(algorithm)을 이용하여 지식을 구조화하는 것이다. AI가 전산학적 모델을 사용하여 정신 능력을 연구할 수 있는 것은 우리의 사고 과정이 컴퓨터의 계산 과정과 일치한다고 전제하기 때문이다. 컴퓨터 프로그램은 데이터 구조와 알고리즘으로 이루어져 있으므로 표1에 나타난 것처럼 우리의 사고와 대응된다고 볼 수 있다.

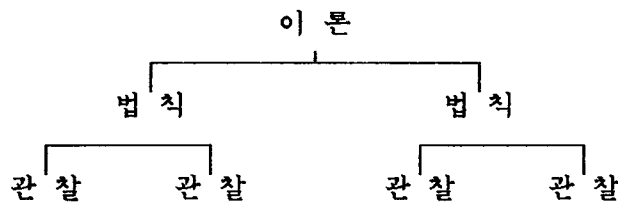
표1. 우리의 사고와 프로그램의 대응 관계

사고	프로그램
정신 구조	데이터 구조
과정	알고리즘

CPS도 과학적 지식에 대해서 AI와 비슷한 전제를 받아들인다. 즉, CPS 관점에 의하면 과학적 지식의 구조는 과학이 실제로 수행되는 과정과 분리될 수 없으며 과학적 지식의 표현은 그 구조와 더불어 그 과정을 반영해야 한다.

여기에서 과학적 지식은 이론(theory)과 법칙(law), 그리고 관찰(observation)로 구성된다. Campbell은 이러한 세 요소간의 관계를 밝혀주는 하나의 이론을 제안하고 있는데, 그에 의하면 이론은 법칙을 설명하는데 이용되고, 법칙은 자연 현상에서 발생하는 개개의 사건을 설명하고 예측하는데 이용된다.[1] 따라서 과학적 지식의 구조는 표2에서 나타난 바와 같은 구조를 지니게 될 것이다.

표2. 과학적 지식의 구조



위의 구조에서 관찰은 법칙에 의해서 일반화된다. 다시 말하면 법칙은 관찰을 조직한 것이다. 그런데 관찰과 이론간의 관계가 LPPS의 주장처럼 독립적인지, 아니면 HPS가 주장하듯이 관찰이 이론 의존적(theory-laden)인지는 위의 구조에는 나타나 있지 않다. 그러나 어떤 주장을 채택하더라도 법칙은 관찰을 일반화한 것으로 볼 수 있다. 과학적 지식의 구조에 대한 이러한 설명은 가장 일반적으로 받아들여지고 있으며 또한 전형적인 과학적 사례와도 일치한다. 이에 대한 예를 들면 빛에 대한 입자 이론과 파동 이론이 있다. 우선 빛의 굴절과 반사 현상을 관찰한 경우에 그것을 일반화시켜 “빛은 굴절한다” 또는 “빛은 반사한다”와 같은 일반 법칙(general law)을 구성하게 된다. 그런데 이러한 일반 법칙을 설명하기 위해서는 입자 이론과 파동 이론이 필요하게 될 것이다.

따라서 법칙은 여러가지 기능을 갖게 된다. 즉, 그것은 관찰을 조직하고, 새로운 관찰을 예측하기도 하며, 문제 해결(problem solving)과 설명에 중요한 역할을 하기도 한다. 논리적 관점에서 보면 이론은 일반 법칙처럼 보일 수 있는데, 그 이유는 법칙과 이론이 모두 if-then 형식을 갖고 있기 때문이다. 그러나 이론은 그 기원과 설명적 기능에서 법칙과 분명히 구별된다. 그것은 법칙이 관찰을 일반화한 것임에 비해 이론은 법칙을 설명하기 위해 제안된 것이기 때문이다. 그러므로 이론은 법칙을 얼마나 잘 설명하는가에 따라 평가된다. 더욱이 이론은 입자나 유전자와 같은 관찰되지 않은 이론적 실재(theoretical entities)를 가정하므로 때로는 법칙과는 달리 경험적 방식으로 발견될 수 없다는 특징을 갖고 있다.

4. 과학적 지식의 표현

이러한 구조를 지닌 과학적 지식의 표현에 AI는 과학철학에 많은 도움을 줄 수 있다. 현재까지 AI에서 지식 표현의 방법으로 제시된 것은 대략 Semantic Networks, Scripts and Frames, Production System, 그리고 Predicate Calculus 네 가지가 있다. [10] 여기에서는 Thagard가 개발한 Production System의 한 종류인 PI (process of induction)를 살펴보겠다. PI는 귀납(induction) 뿐만 아니라 개념 형성(concept formation)과 가설 발상법(abduction)을 수행하는 알고리즘을 갖고 있으며, 그 주기능은 문제 해결과 귀납의 일반 모델을 LISP 언어 체계에서 수행하는 것이다. [6] 과학적 지식의 요소인 관찰과 법칙 그리고 이론은 PI에서 다음과 같이 표현될 수 있다. [12]

a) PI는 관찰이나 추론 결과를 메시지(message)를 이용하여 표현한다. 하나의 메시지는 술어(predicate)와 논항(argument), 진리치(truth-value), 신뢰도(confidence), 메시지 명칭(message-name)과 같은 정보를 포함하고 있는 리스트(list)이다. 예를 들어 “화성은 붉다”라는 관찰 문장은 (red(Mars)true 1)로 표현된다. 이러한 구조에 의해 간단한 가설(simple hypothesis)도 표현할 수 있다. 즉, “화성에는 생명체가 없다”라는 가설은 (has-life(Mars) projected-to-be-false .7 hypothesis-26)로 표현된다. 이러한 리스트에서 .7은 어떤 메시지에 대한 그 체계의 신뢰도이고, hypothesis - 26은 메시지 명칭으로서 부가적인 정보의 저장에 이용된다.

b) 법칙은 if-then 형식을 지닌 규칙(rule)에 의해 표현된다. 우선 그 법칙의 성공과 실패를 추적하기 위해 규칙에 명칭이 부여된다. PI에서는 법칙의 성공과 실패가 ‘강도(strength)’라는 0과 1사이의 양으로 집약된다.

예를 들어 “구리는 전도체이다”라는 규칙은 다음과 같이 표현된다.

Name: Rule-22
Data-type: rule
Concepts-attached-to: copper
Condition: If x is copper
Action: Then x conducts electricity
Strength: .7

c) 이론, 법칙, 그리고 관찰은 그 자체내에 개념을 포함하므로 그것에 대한 표현을 필요로 한다. PI에서 개념은 메시지나 규칙보다 좀더 복잡한 형식을 지닌 Minsky의 프레임(frame)과 비슷한 구조로 표현된다. 따라서 개념은 그것의 위계망(hierarchical network)내에서의 자신의 위치에 관한 정보를 갖게 되며, 여러가지 성질들을 기술하는 법칙들을 갖는다. 예를 들어 음파 개념은 다음과 같이 표현된다.

Name: sound
Data-type: concept
Activation: 0
Superordinates: physical phenomenon, sensation
Subordinates: voice, music, whistle, animal sounds
Instances:
Activated-by:

Rules:

Rule 0: If x is heard, then x is a sound.
Rule 1: If x is a sound and y is a person and x is near y,
then y hears x.
.....

위의 표현에서 음파 개념에 부여된 규칙들은 각각에 대한 표현을 갖고 있다. 예를 들어 규칙 1은 다음과 같이 표현될 수 있다.

Name: Rule-3
Data-type: rule
Concepts-attached-to: sound
Conditions: (sound(\$x>true)
(person(\$y>true)
(near(\$x \$y>true)

Action:	(hears(\$y \$x>true)
Slot:	person-effect
Status:	default
Strength:	.7
Activation:	0
Old-matches:	nil
Current-match:	nil
Satisfies-goal?:	nil
Projection-status:	nil
Current-value:	0
Action-instances:	nil

d) 이제 마지막으로 이론이 어떻게 표현되는가를 살펴보겠다. 이론은 규칙이나 개념처럼 분명한 데이터 구조가 아니라 관련된 구조들의 집합이다. 따라서 음의 파동이론은 다음과 같이 표현될 것이다.

Wave theory of sound:
 Concepts: sound.wave.
 Theoretical concept: sound-wave.
 Rules: If x is sound, then x is a wave.
 If a is sound, then x is a sound-wave.
 Problem solution: Explanation of why sound propagates.
 Explanation of why sound reflects.

4. 맺는말

본 논문에서는 AI의 지식표현을 사용하여 과학적 지식을 표현하는 방법을 제시하였다. 이러한 방법에 의해 표현된 과학적 지식은 술어논리로 표현된 기존의 표현들보다도 더욱 엄밀하고 그것이 담고 있는 정보가 많다는 점에서 LPPS의 이상을 어느 정도 달성한 것이라고 생각된다. 또한 그것은 PI와 같은 프로그램내에서 귀납 추리나 문제 해결 과정을 통하여 과학의 과정을 구체적으로 모의실험함으로써 HPS가 주장하는 실제성을 얻을 수 있을 것이다. CPS는 AI에서의 지식 표현이 과학철학의 주제들을 그 표현 대상으로 삼을 경우 그 표현 형식이나 의미에 대한 분석을 제공함으로써 역으로 AI에 영향을 줄 수 있을 것으로 생각된다.

참고문헌

1. Campbell, N.R.(1957) Foundations of physics. Dover.
2. Darden, L.,(1987) Viewing the history of science as compiled hindsight. AI Magazine, 8(2) : 33-41.
3. Fetzer, J.H.(1990) Artificial intelligence : its scope and limits. Kluwer academic.
4. Graubard, S.R.(1988) The artificial intelligence debate. MIT Press.
5. Haugeland, J.(1985) Artificial intelligence. MIT Press.
6. Holland, J., Holyoak, K., Nisbett, R., and Thagard, P.(1986) Induction. MIT Press.
7. Langley, P., Simon, H.A., Bradshaw, G.L. and Zytkow, J.M.(1987) Scientific discovery : Computational explorations of the creative process. MIT Press.
8. Nickles, T.(1980) Scientific discovery : Logic and rationality. D.Reidel.
9. Searle, J.(1984) Minds, brains and science. Harvard Univ. Press.
10. Sharples, M., Hogg, D., Hutchison, C., Torrance, S., and Young, D. (1989) Computers and thought. MIT Press.
11. Simon, H.A.(1977) Models of discovery. D.Reiel.
12. Thagard, P.(1988). Computational philosophy of science. MIT Press.