불확실성 기반 상태 공간 학습 알고리즘을 이용한 Exploration-Exploitation 딜레마에 관한 연구

A Study on the Exploration-Exploitation Dilemma using an uncertaintydriven state space learning algorithm

저자
(Authors)

안수진, 이상완
Su Jin An, Sang Wan Lee

출처
(Source)

한국지능시스템학회 논문지 27(4), 2017.8, 316-320 (5 pages)
Journal of Korean Institute of Intelligent Systems 27(4), 2017.8, 316-320 (5 pages)

발행처
(Publisher)

한국지능시스템학회
Korean Institute of Intelligent Systems

URL

http://www.dbpia.co.kr/Article/NODE07227292

APA Style

안수진, 이상완 (2017). 불확실성 기반 상태 공간 학습 알고리즘을 이용한 Exploration-Exploitation 딜레마에 관한 연구. 한국지능시스템학회 논문지, 27(4), 316-320.

Journal of Korean Institute of Intelligent Systems
Vol. 27, No. 4, August 2017, pp. 316–320
http://dx.doi.org/10.5391/JKIIS.2017.27.4.316

ISSN(Print) 1976–9172
ISSN(Online) 2288–2324

JKIIS

불확실성 기반 상태 공간 학습 알고리즘을 이용한 Exploration–Exploitation 딜레마에 관한 연구

A Study on the Exploration-Exploitation Dilemma using an uncertainty-driven state space learning algorithm

안수진* · 이상완*†
Su Jin An, and Sang Wan Lee†

*한국과학기술원 바이오및뇌공학과
*Department of Bio & Brain Engineering, KAIST

요 약

본 논문은 인간의 학습 과정에서 발생하는 자신의 학습 정도에 대한 불확실성을 평가하는 메타 인지 능력 기반 학습 과정을 형식화하고, 메타 인지 기반 상태 공간 학습 알고리즘을 제안하였다. 상태 공간 (state-space) 학습과정을 구현한 2단계 마르코프 의사결정 게임 데이터를 사용, 상태 공간 정보에 대한 근접도(Proximity)와 최소오류제곱을 이용하여 불확실성을 도출하고, 이것이 인간의 상태 공간 학습에 미치는 영향을 확인하였다. 또한 현재 가지고 있는 상태 공간에 대한 정보를 바탕으로 기계학습의 본질적 문제인 Exploration-Exploitation 딜레마의 trade-off를 예측하고, 개선 가능성을 보였다.

키워드 : 학습, 불확실성, 메타 인지, Exploration-Exploitation 딜레마, 의사결정

Abstract

This paper proposes a formal model and the algorithm for human's state space learning process based on metacognition which is seen as the human's capability to introspect their thought process and report their level of uncertainty. Given the 2-stage Markov decision process game data which augmented humans' state space learning process, the algorithm performs online updates of low dimensional embedding of state space on the basis of the uncertainty assessment. We found out that the uncertainty does use on the human learning process. Furthermore, by predicting the exploration-exploitation tradeoff using acquired knowledge about the state space, it is expected to improve the exploration and exploitation dilemma in machine learning.

Key Words : Learning, Uncertainty, Metacognition, Exploration-Exploitation dilemma, Decision making

Received: May. 18, 2017
Revised: Jul. 30, 2017
Accepted: Aug. 4, 2017
†Corresponding authors
sangwan@kaist.ac.kr

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 정보통신 · 방송 연구개발사업의 일환으로 수행하였음. (2016-0-00563, 자율지능 동반자를 위한 적응형 기계학습기술 연구개발)
이 논문은 KAIST의 지원을 받아 수행하였음. (과제번호 G04150045)
본 연구는 2017년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행하였음.(NRF-2017R1C1B 2008972)

1. 서 론

메타 인지 능력(metacognitive ability)[1]은 인간의 지식과 인지 영역에 대한 통제와 조절을 일컫는 것으로써, 학습 과정 중 자신의 학습 정도에 대한 불확실성을 평가하는 인간의 고유 능력을 포함한다. 메타 인지 능력은 인간의 학습 과정에서 학습 성취를 위한 행동을 계획하고 실행하는 것에 중요한 역할을 한다. 예를 들어 (i) 어떠한 문제를 해결하기 위해서 이미 알고 있는 방법을 고수할 것인지 (exploitation), (ii) 가능한 다른 방법을 탐색할 것인지(exploration)를 선택해야 하는 상황, 또는 (iii) 자신이 내린 의사 결정에 대한 확신 정도를 평가해야 하는 상황에, 이러한 메타 인지 능력을 사용하게 된다.

기계학습의 경우, 위의 (i), (ii) 방법을 선택할 때, 더 나은 방법을 추적하기 위해 많은 양의 데이터에 의존한 최적화 방식을 쓰게 된다. 이 때문에 초기학습에 많은 시간이 소요되며 현재 처한 환경에 대한 학습이 미진한 경우 에이전트는 Exploration - Exploitation 딜레마에 빠지기 쉽다. 이러한 문제는 마르코프 의사결정 과정에서 요구되는 온라인 및 순차적 데이터 학습 과정 시나리오에서 더욱 심각해진다.

본 논문에서는 인간의 메타 인지 능력을 기계 학습과 결합하여, 메타 인지 기반 상태 공간 학습과정을 형식화하고 메타 인지 기반 상태 공간 학습 알고리즘을 제안한다. 또한 현재 가지고

있는 상태 공간에 대한 정보를 바탕으로 Exploration-Exploitation 딜레마의 trade-off를 예측하고, 개선 가능성을 알아본다.

2. 마르코프 의사결정 게임

새로운 환경에 처한 에이전트의 학습 과정에서 변화하는 불확실성을 관찰하기 위해 본 논문에서는 인간의 순차적 의사결정을 형식화한 2단계 마르코프 의사결정 게임 데이터 [2]를 사용하였다. 이 논문에서는 보상 기반 전략과 관계없이 상태 공간 학습 과정만을 관찰하기 위해 본 실험 이전의 연습 실험 데이터만을 사용하였다.

연습 실험은 약 15분간 총 90회의 시도로 이루어졌다. 각각의 시도마다 피험자가 왼쪽 또는 오른쪽을 순차적으로 두 번 선택한 뒤 선택에 의한 보상을 얻게 된다. 보상으로는 각기 다른 값을 가지고 있는 4가지의 동전 중 하나가 주어지게 된다. 피험자는 어떠한 선택 조합이 가장 큰 값을 가지고 있는 동전을 획득할 수 있는지 학습하여야 한다. 그 과정에서 피험자는 동일한 선택을 하여 그 전략에 대해 확신을 쌓게 되거나 새로운 선택을 하여 새로운 전략을 탐색하게 된다.

만약 피험자가 4초 안에 아무런 선택을 하지 않을 경우 임의로 왼쪽 또는 오른쪽이 자동으로 선택되어 실험이 계속 진행된다. 왼쪽 또는 오른쪽을 선택함에 따라 달리 주어지는 보상은 0.5의 확률로 결정된다. 다시 말해 선택의 결과가 항상 보상을 보장하지 않는다는 구조를 가지고 있는 실험이다.

피험자는 먼저 연습 실험을 통해 다양한 선택을 시도해볼 수 있다. 이처럼 연습 실험을 통해 처음 접하는 실험의 상태 공간을 배우는 것이 위의 의사 결정 게임의 목표이다. 이 실험에는 총 22명(여성 6명, 남성 16명, 19세-40세)의 피험자가 참가하였다.

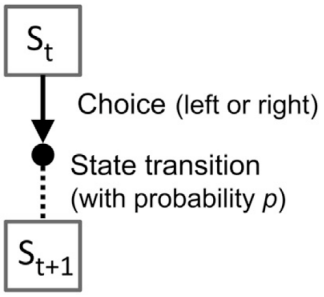

그림 1. 순차적 마르코프 의사결정 게임 [2]
Fig. 1. Sequential Markov Decision Task [2]

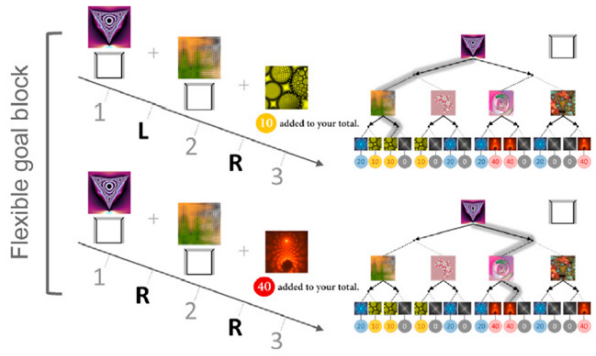

그림 2. 2단계 마르코프 의사결정 게임 [2]
Fig. 2. 2 Stage Markov Decision Task [2]

3. 불확실성 기반 상태 구조 학습 알고리즘

앞에서 기술한 실험을 통해 획득한 데이터를 이용하여 의사결정의 불확실성 예측 모델을 추적하였다. 본 논문에서는 Li *et al.* [3, 4]에서 제안한 모델을 바탕으로, 아래와 같은 구조를 가진 예측 모델 및 알고리즘을 사용하였다.

(i) 이전에 경험한 상태 공간을 낮은 차원으로 사영시키는 부분
(ii) (i)의 결과를 바탕으로 현 상태 공간에 대한 불확실성을 계산하는 부분
(iii) 도출한 불확실성에 기반을 두어 현 상태 공간을 예측하는 부분

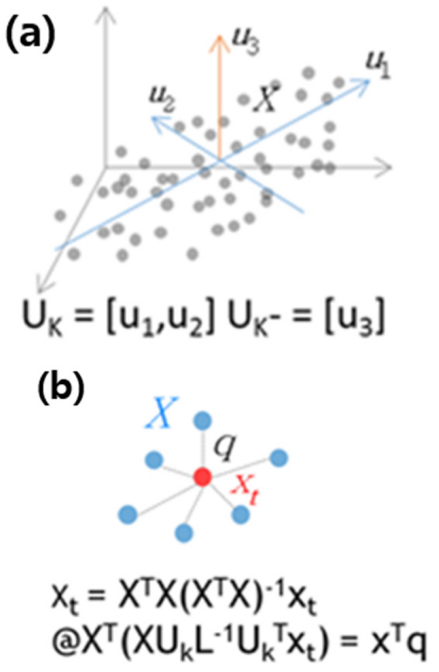

그림 3. (a) 경험한 상태 공간을 낮은 차원으로 사영시키고, (b) 이를 바탕으로 불확실성을 도출
Fig. 3. (a) Low dimensional embedding of the state space (b) Computing the uncertainty q with the state space previously learned

모델은 그림 3.(a)에서 보는 바와 같이, 이전에 경험한 높은 차원의 상태 공간을 낮은 차원으로 사영시킨다. 근접도를 기반으로 불확실성을 도출하는 알고리즘의 특성상 높은 차원에서의 불분명한 거리 계산을 배제, 더 정확한 불확실성을 도출하고자 하였다.

그림 3.(b)에서 X는 이전에 경험했던 상태 공간 정보를 나타내며, x_t는 현재 시간 t에서의 상태 공간을 나타낸다. 불확실성 q는 이전에 경험했던 상태 공간 정보에 대한 근접도와 최소오류제곱을 이용하여 도출한다.

현 상태 공간 x_t는 불확실성 q와 경험했던 상태 공간 X에 대한 선형 결합으로 표현될 수 있다. 이미 경험한 바 있는 상태 공간이 현 상태 공간과 근접하다면 불확실성 q는 작을 것이다. 이와 반대로 현 상태 공간 x_t를 이전에 경험하지 못했다면, x_t와 X 사이의 근접도가 작아질 것이므로 불확실성 q는 상대적으로 클 것이다.

본 논문에서 제안한 모델은, 도출한 불확실성 q가 기준 값보다 작다면 경험한 상태 공간 정보를 이용해 현 상태 공간을 예측하며, 이를 바탕으로 Exploration-Exploitation trade-off를 예측할 수 있다. 반면 도출한 불확실성이 기준 값보다 크다면 모델은 상태 공간을 유추할 수 없다고 보고, 현 상태 공간을 관찰한 후, 경험했던 상태 공간 정보에 현 상태 공간 정보를 추가하게 된다.

실험에서는 매번의 선택마다 위의 모든 단계를 반복하였다. 그러므로 실험 과정에서의 단계적으로 변화하는 불확실성을 관측할 수 있게 된다.

4. 시뮬레이션 및 결과

모델이 도출한 상태 공간에 대한 불확실성을 (i) 아주 높음, (ii) 높음, (iii) 낮음으로 나눠 실제 실험에서 얻은 행동 데이터와 비교 분석하였다. 그 결과 모델이 도출한 불확실성이 낮을수록 피험자들이 더 높은 수행 점수를 획득했음을 알 수 있다. 또한 수행

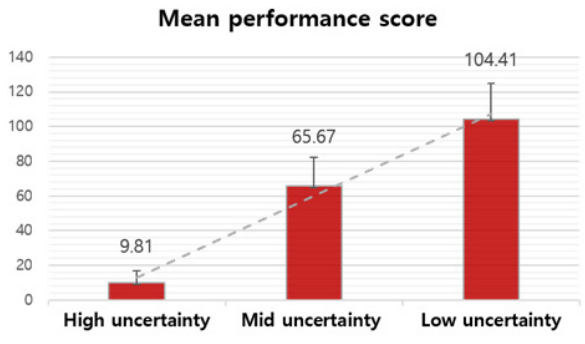

그림 4. 불확실성에 따른 인간의 실제 수행 점수
Fig. 4. Analysis of human task performance based on the uncertainty estimated by the proposed model

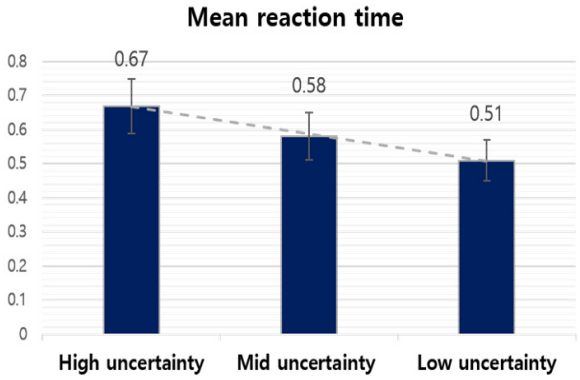

그림 5. 불확실성에 따른 인간의 실제 수행 시간
Fig. 5. Analysis of human reaction time based on the uncertainty estimated by the proposed model

시간 역시 더 빨라진 것을 관찰할 수 있었다. 이는 모델이 도출한 불확실성이 인간의 행동 패턴을 잘 반영했다는 사실을 보여준다.

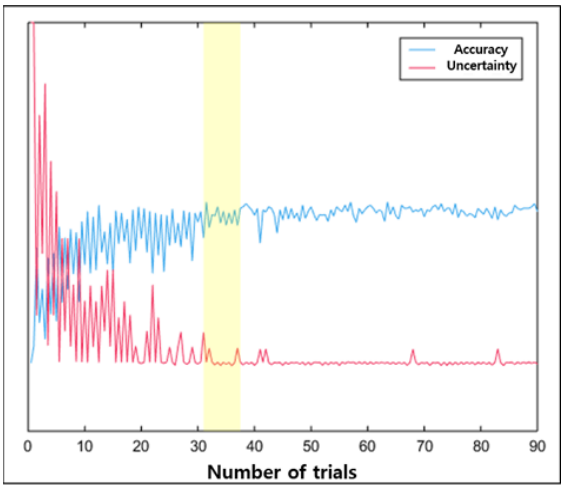

그림 6. 불확실성과 상태 공간 예측 정확도 변화
Fig. 6. Uncertainty and the model prediction accuracy

모델에서 예측한 불확실성은 평균 40회 시도 이후에 낮은 상태로 유지되었다. 이는 평균 40회 시도 내에 피험자가 실험의 상태 공간을 모두 학습했다는 것으로 볼 수 있다. 그림 7.에서 보여주는 바와 같이 40회 시도 이후의 시도들 (학습후기, 그림 7. 오른쪽)에서 모델의 상태 공간 예측 정확도가 높아진 것이 위의 결과를 뒷받침한다.

이처럼 불확실성을 바탕으로 예측한 Exploration-Exploitation trade-off 패턴을 실제 Exploration-Exploitation 행동 패턴과 비교해 보았을 때, 불확실성이 비슷한 상태 공간에 대해 학습 초기에는 exploration 비율이 exploitation에 비해 상대적으로 높은 것을 볼 수 있었다. 이는 학습 초기에 전체적인 상태 공간을 배우기 위해 불확실성이 높은 상태 공간에 대해 상대적으로 exploration을 많이 한다는 모델의 예측 결과와 일치한다.

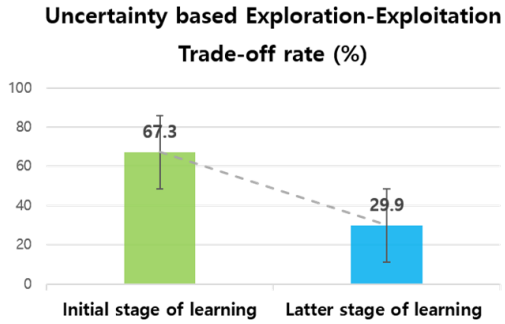

그림 7. 불확실성 기준 Exploration–Exploitation trade–off 비율 (%)
Fig 7. Uncertainty based Exploration-Exploitation trade-off rate (%)

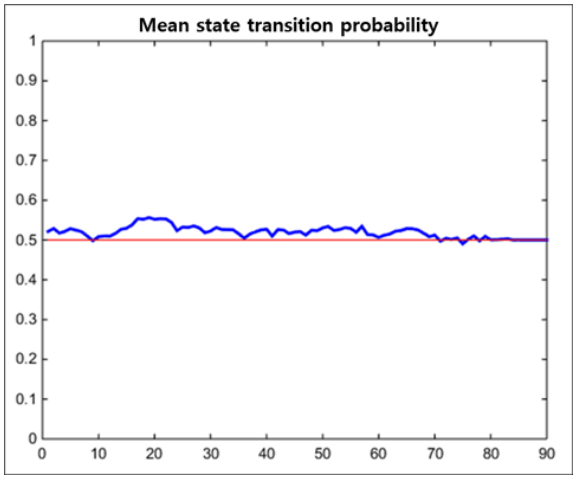

그림 8. 모델의 상태 공간 변화 확률 예측 변화 추이
Fig. 8. State transition probability prediction

나아가 모델은 피험자의 행동 데이터를 바탕으로 실험의 상태 공간 변화 확률을 예측한다. 예측한 상태 공간을 바탕으로, 모델은 임의의 확률 (0.5)로부터 시작해 상태 공간 변화 확률을 조정하게 된다.

5. 결론 및 향후 연구

불확실성 기반 상태 공간 학습 모델을 통해 예측한 불확실성이 인간의 상태 공간 학습에 영향을 미치는 것을 확인하였다.

모델이 도출한 불확실성이 낮은 경우, 실제 인간의 행동 결과에서 더 높은 수행 점수와 더 빠른 수행 시간을 보이는 것을 관찰할 수 있었다. 이는 본 논문에서 제안한 모델이 실제 인간의 행동 패턴을 잘 반영했다고 볼 수 있다.

모델이 예측한 불확실성은 평균 40회 시도 이후에 낮게 유지되었다. 모델의 상태 공간 예측 정확도 역시 40회 시도 이후에 높게 유지되었다.

또한 불확실성을 기준으로 예측한 Exploration - Exploitation trade-off와 실제 인간의 Exploration - Exploitation 패턴 모두 학습 초기에는 exploration 비율이 exploitation에 비해 상대적으로 높은 것을 볼 수 있었다.

위의 결과로써 본 모델은 기계학습의 본질적 문제인 Exploration-Exploitation 딜레마 개선에 활용될 수 있을 것으로 보인다.

향후 위의 결과를 바탕으로 실제 인간의 행동 패턴 예측이 가능한 모델 개발을 통해 초기 학습이 빠른 기계 학습 알고리즘 개발하고자 한다.

References

[1] Flavell, J. H., & H., J., "Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry," American Psychologist, vol. 34, no. 10, pp. 906-910, 1979.

[2] Lee, S. W., Shimojo, S., & O' Doherty, J. P, "Neural Computations Underlying Arbitration between Model-Based and Model-free Learning," Neuron vol. 81, no. 3, pp. 687–699, 2014.

[3] Li, L., Littman, M. L., & Walsh, T. J., "Knows what it' knows," In Proceedings of the 25th international conference on Machine learning - ICML 08, pp. 568–575, 2008.

[4] Li, L., Littman, M. L., Walsh, T. J., & Strehl, A. L., "Knows what it knows: a framework for self-aware learning," Mach Learn vol. 82, pp. 399–443, 2011.

저 자 소 개

안수진(Su Jin An)

2014년 : St.Xavier's College, 컴퓨터공학과 공학사

2017년 : KAIST 바이오및뇌공학과 석사

2017~현재 : KAIST 바이오및뇌공학과 박사과정

관심분야 : Brain-inspired AI, computational neuroscience

Phone : ※ 개인정보 표시제한

E-mail : sujinan@kaist.ac.kr

이상완 (Sang Wan Lee)

2003년 : 연세대학교 전기전자 학사

2005년 : KAIST 전자전산학과 석사

2009년 : KAIST 전자전산학과 박사

2010년~2011년 : MIT 박사후연수연구원

2011년~2015년 : CALTECH 박사후연수 연구원

2015년~현재 : KAIST 바이오및뇌공학과 조교수

관심분야 : Brain-inspired AI, computational neuroscience

Phone : ※ 개인정보 표시제한

E-mail : sangwan@kaist.ac.kr