복잡계의 위상특성을 이용한 MDP 학습의 효율 분석

이승준○, 장병탁
서울대학교 컴퓨터공학부 바이오지능 연구실
{sjlee○, btzhang}@bi.snu.ac.kr

Using Topological Properties of Complex Networks for analysis of the efficiency of MDP-based learning

Seung Joon Yi0 Byoung Tak Zhang
Biointelligence Lab, School of Computer Science, Seoul National University

요 약

본 논문에서는 마르코프 결정 문제 (Markov decision problem)의 풀이 효율을 잴 수 있는 척도를 알아보기 위해 복잡계 네트워크 (complex network) 의 관점에서 MDP를 하나의 그래프로 나타내고, 그 그래프의 위상학적 성질들을 여러 네트워크 척도 (network measurements)들을 이용하여 측정하고 그 MDP의 풀이 효율과의 관계를 분석하였다.

실세계의 여러 문제들이 MDP로 표현될 수 있고, 모델이 알려진 경우에는 평가치 반복(value iteration) 이나 모델이 알려지지 않은 경우에도 강화 학습(reinforcement learning) 알고리즘등을 사용하여 풀 수 있으나, 이들 알고리즘들은 시간 복잡도가 높아 크기가 큰 실세계 문제에 적용하기 쉽지 않다. 이 문제를 해결하기 위해 제안된 것이 MDP를 계층적으로 분할하거나, 여러 단계를 묶어서 수행하는 등의 시간적 추상화(temporal abstraction) 방법들이다.

시간적 추상화를 도입할 경우 MDP가 보다 효율적으로 풀리는 꼴로 바뀐다는 사실에 착안하여, MDP의 풀이 효율을 네트워크 척도를 이용하여 측정할 수 있는 여러 위상학적 성질들을 기반으로 분석하였다. 다양한 구조와 파라미터를 가진 MDP들을 사용해 네트워크 척도들과 MDP의 풀이 효율간의 관계를 분석해 본 결과, 네트워크 척도들 중 평균 측지 거리 (mean geodesic distance) 가 그 MDP의 풀이 효율을 결정하는 가장 중요한 기준이라는 사실을 알 수 있었다.

1. 서 론

실세계의 다양한 문제들은 상태(state) 와 상태에서 취할 수 있는 행동(action), 그리고 상태와 행동의 결과인 보상(reward)의 형태로 주어지는 마르코프 결정 문제 (Markov decision problem, MDP)의 형태로 표현될 수 있다. MDP를 푼다는 것은 최적의 보상을 얻는 정책 (Policy)를 구하는 것을 의미하며, MDP는 환경의 모델을 알 경우 동적 프로그래밍(dynamic programming)등을 이용하여 풀 수 있고, 환경이 알려져 있지 않은 경우에도 시행착오를 통해 학습하는 강화 학습 (reinforcement learning)을 사용하여 풀 수 있다.

하지만, 이러한 방법들은 계산 복잡도가 높고, 기본적으로 이산적인 시간과 공간을 가정하기 때문에 연속적인 시간과 공간을 갖는 실세계 문제에 바로 적용하기가 어렵다. 주로 쓰이는 대안으로는 신경망과 같은 함수 근사 장치를 사용하는 방법으로 많은 성공적인 적용 예가 있어 왔으나, 이러한 근사 장치를 사용하더라도 문제가 복잡해짐에 따라 학습해야 할 파라미터의 수가 증가하는 것은 막을 수가 없다. 강화 학습의 경우 학습 난이도가 문제 크기나 파라미터 수가 커짐에 따라 지수적으로 증가하는 차원성의 저주[1] (curse of dimensionality) 문제 때문에 크고 복잡한 실세계 문제에 적용하기가 매우 어렵다.

이러한 문제를 해결하기 위해 제안된 방법이 시간적 추상화(temporal abstraction) 방법이다[1]. MDP가 크기가 커짐에 따라 두 상태 간에 더 많은 판단 단계를 거치게 되어 학습이 어려워지는 것을 막기 위해, MDP를 계층적으로 분할하거나 한번에 실행되는 여러 행동들의 묶음을 사용한다. 이러한 방법으로 요구되는 판단 단계의 수를 줄여, 크기가 큰 MDP를 보다 효율적으로 풀 수 있게 해준다.

본 논문에서는 계층적으로 조직화된 MDP나 시간적 추상화를 도입한 MDP의 경우 일반적인 MDP보다 매우 빠른 속도로 풀 수 있다는 점에 주목하고, 어떠한 형태의 MDP가 보다 효율적으로 풀리는지에 대해 분석해 보고자 하였다. 이를 위하여 복잡계 네트워크 (complex network)적인 관점에서 MDP를 하나의 그래프로 보고, 그 그래프의 위상적 성질들을 여러 네트워크 척도 (network measurement) 들을 이용하여 측정한 뒤 그 MDP의 풀이 효율과의 상관관계를 분석해 보았다.

다양한 네트워크 모델들과 다양한 파라미터들의 조합으로 만든 MDP들을 사용한 실험 결과, MDP의 풀이 효율에 가장 영향을 끼치는 네트워크 척도는 평균 측지적 거리 (mean geodesic distance)임을 알 수 있었다.

2. 관련 연구

2.1. MDP와 Semi MDP

MDP는 $\langle S, A, T, R \rangle$로 정의할 수 있다. S는 유한한 상태들 s의 집합, A는 유한한 행동들 a의 집합, R은 (s,a,s') 의 함수로 주어지는 보상, T는 행동 a를 수행할 경우 상태 s에서 s'로 이동할 확률이다. 정책 π는 각 상태에서 취할 행동들로, MDP를 푸는 것은 장기적으로 최고의 보상을 얻을 수 있는 최적 정책 π^*를 구하는 것이다. MDP의 모델, 즉 R, T에 대해 알고 있을 경우 평가치 반복(value iteration) 등의 동적 프로그래밍 방법으로 π^*를 구할 수 있고, 모델에 대해 모를 경우에도 Q-learning 등의 강화 학습 알고리즘을 사용하여 π^*를 구할 수 있다.

MDP에서는 각 행동들에 단위 시간이 소요된다는 것을 가정하나, 각 행동들에 다양한 시간이 소요될 수 있도록 MDP 모델을 확장한 것이 Semi MDP이다. 시간적 추상화 (temporal abstraction) 을 MDP에 적용할 경우, 그 확장된 MDP를 Semi MDP로 나타낼 수 있다. Semi MDP의 경우에도, 평가치 반복이나 Q-learning 등의 방법을 약간의 수정을 거쳐 그대로 적용 가능하다[2].

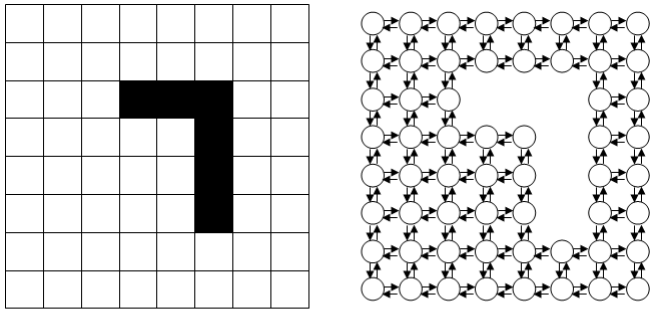
2.2 복잡계 네트워크와 네트워크 척도

실세계의 네트워크들을 위한 모델로는 예전부터 랜덤 네트워크 (random network)가 사용되어온 바 있으나, 최근의 연구 결과 실세계의 많은 네트워크들은 랜덤 네트워크 모델에서는 볼 수 없는 작은 세상 성질 (small world property) 나 높은 클러스터링 계수[3] (clustering coefficient), 척도 없는 도수 분포[4](scale-free degree distribution) 등의 여러 특징적인 성질을 공유하는 것이 밝혀졌다. 이러한 네트워크의 위상적 성질들은 네트워크의 연결 관계를 특징짓고 그 네트워크상에서 일어나는 여러 과정들에 영향을 준다. 이러한 네트워크들의 여러 위상적 성질들을 측정하는 데 사용되는 것이 네트워크 척도 (network measurement)로써 네트워크들을 분석하고 서로 다른 종류들로 구분하며, 원하는 성질을 갖는 네트워크를 디자인하는 등 다양한 분야에 사용된다.

3. 네트워크 척도들과 MDP의 풀이 효율간의 관계 분석

3.1. MDP의 네트워크 표현

네트워크 척도를 사용하여 MDP를 분석하기 위해서는 일단 MDP를 네트워크 형태로 나타내는 게 필요하다. MDP의 상태 s_i는 네트워크의 노드 i에 일대일 대응되게 되고, 상태 변화 (s_i, a, s_j)는 네트워크의 에지 (i,j)에 대응되게 된다. 본 연구에서 다룬 결정론적인 MDP의 경우, 각 행동의 결과로 일어나는 상태 변화는 한 가지 뿐이므로, 네트워크의 각 에지는 MDP의 상태와 행동 (s,a)에 대응하게 된다. 그림 1은 중앙에 장애물이 있고 각 상태에서의 행동이 상하좌우로 이동하는 것으로 주어지는 2차원 격자 형태의 MDP와 거기 대응되는 네트워크를 보여주고 있다.



(a) (b)

그림 1. 장애물이 있는 2차원 결정론적 MDP와 그에 대응되는 네트워크 모델.

3.2 . 실험 설계

다양한 특성을 갖는 네트워크에서 네트워크 파라미터와 MDP의 풀이 효율간의 관계를 분석하기 위해, 다음의 네트워크 모델들을 사용하였다. 상태의 수는 4개부터 2500개까지, 모델들의 에지 생성 파라미터는 0부터 1.0까지의 값을 사용하였다.

- Regular lattice 모델
- Erdos-renyl 모델
- Watts-Strogatz 모델[3]
- Barabasi-Albert 모델[4]

파라미터별, 크기별로 생성된 네트워크들로부터 다음의 네트워크 척도를 측정하였다.

- 평균 측지 거리(mean geodesic distance)
- 전역 효율[5](global efficiency)
- 평균 차수(average degree)
- 최대 차수(maximum degree)
- 클러스터링 계수[6](clustering coefficient)
- 부분그래프 중앙도[7](subgraph centrality)
- 프랙탈 차원[8](fractal dimension)

또한 생성된 네트워크를 사용하여 대응되는 MDP를 푸는 데 걸리는 시간을 측정하였다. 보상으로는 목적 상태 1개에만 10, 나머지는 0을 주었고 파라미터로는 γ=0.9, ε=0.1을 사용하였다. 다음의 알고리즘을 사용하여, 출발점에서 목적 상태까지의 거리가 최적치의 110%안으로 수렴하는 시간을 측정하였다.

- 평가치 반복(value iteration)
- Q-학습(q-learning)

3.3 실험 결과

다양한 모델과 파라미터를 사용하여 만든 네트워크들로부터 그림 2와 같은 네트워크 척도 값들을 얻을 수 있었다. 평균, 최대 차수(c와 d), 클러스터링 계수 (e), 부분

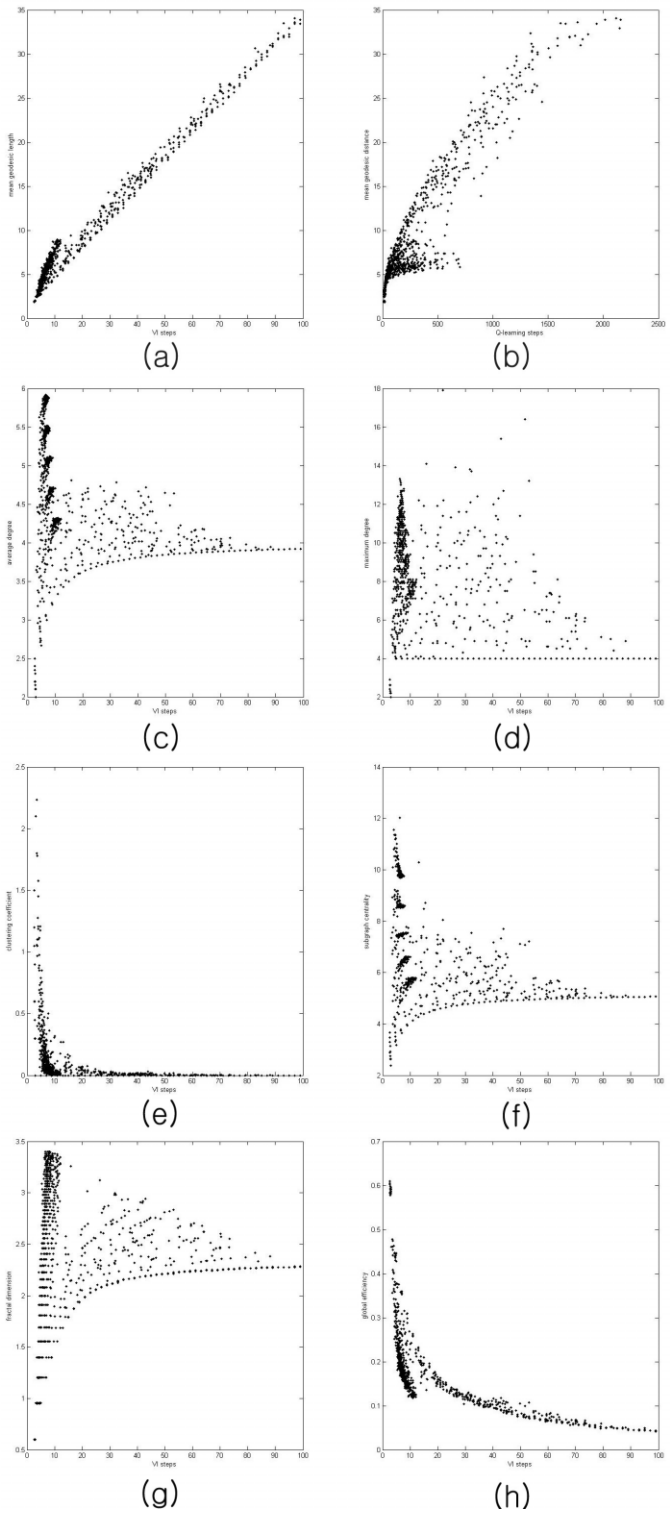


그림 2. 각 네트워크 척도와 MDP 풀이 효율간의 관계

그래프 중앙도 (f), 프랙탈 차원 (g)등과 달리, 평균 측지 거리는 평가치 반복(a) 와 Q학습 (b) 모두에 대해 다양한 네트워크 모델과 파라미터를 사용하더라도 공통적으로 관측되는 뚜렷한 상관관계가 있음을 볼 수 있었다. 또한 평균 측지 거리와 관계가 있는 전역 효율(h)의 경우에도 MDP의 풀이 효율과의 상관관계를 관측할 수 있었다.

4. 결론

MDP에 시간적 추상화를 적용할 경우 MDP의 풀이 효율이 개선된다는 점에 착안하여, 복잡계 네트워크의 관점에서 네트워크 척도들을 이용하여 MDP의 위상적 특징과 MDP의 풀이 효율과의 관계를 분석해 보았다. 다양한 네트워크 모델들과 다양한 파라미터들의 조합으로 만든 여러 MDP들을 사용한 실험 결과, MDP의 풀이 효율에 가장 영향을 끼치는 네트워크 척도는 평균 측지적 거리(mean geodesic distance)임을 알 수 있었다. 이 사실로부터, 작은 세상 네트워크를 사용한 MDP 모델[9]의 성능 향상을 설명할 수 있었다.

향후 과제로는 결정론적 MDP가 아닌 일반적인 MDP의 경우로 이러한 분석을 확장하고, 확률적인 MDP에서도 성능 향상을 보일 수 있는 네트워크 척도와 네트워크 모델을 찾는 일이 있을 것이다.

감사의 글

이 논문은 과학기술부 국가지정연구실사업(NRL)에 의하여 지원되었음.

참고 문헌

[1]Barto, A.G., Mahadevan, S. Recent advances in hierarchical reinforcement learning. Discrete Event Systems Journal, 13, 41-77, 2003.

[2]Sutton, R.S., Precup, D., Singh, S.P. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. Artificial Intelligence, 112, 181-211, 1999.

[3]Watts, D.J., Strogatz, S.H. Collective dynamics of 'small-world' networks. Nature, 393, 404-407, 1998.

[4]Barabasi, A.-L., Albert, R. Emergence of scaling in random networks. Science, 286, pp. 509-512,. 1999.

[5]Latora, V., Marchiori, M. Efficient behavior of small-world networks. Physics Review Letters, 87(19):198701, 2001.

[6]Barrat A., Weigt M. On the properties of small-world networks. European Physical Journal B, 13(3):547-560, 2000.

[7]Estrada, E., Rodriguez-Velaquez, J.A. Subgraphs centrality in complex networks. Physical Review E, 71:056103, 2005.

[8]Song, C. Havlin, S. Makse,H.A. Self-similarity of comlex networks. Nature, 433(27):392-395, 2005.

[9]이승준, 장병탁. 복잡계 네트워크를 이용한 강화 학습 구현, 한국정보과학회 가을학술발표 논문집, pp.232-234, 2004.