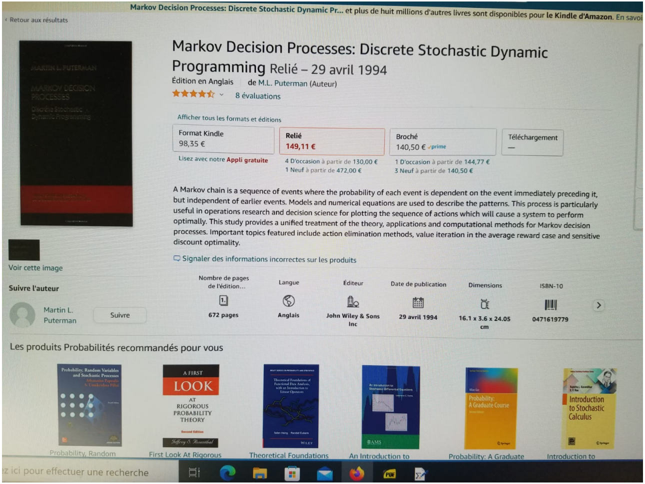
Dynamic Social Learning Under Graph Constraints

K. Avrachenkov (Inria)
V.S. Borkar, S. Moharir, S.M. Shah (IITB)

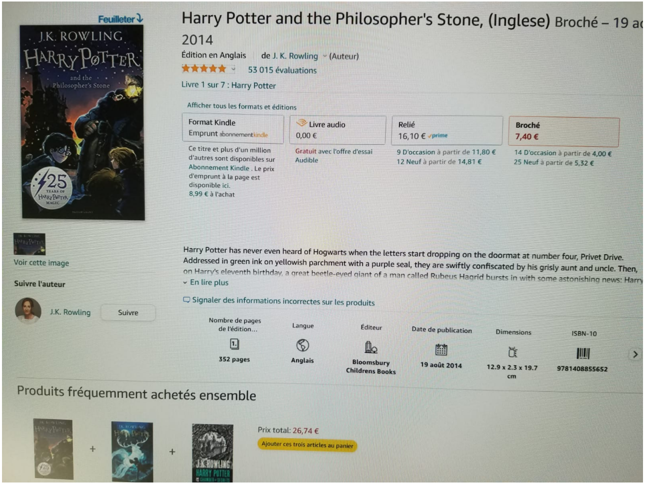
IEEE Trans. on Control of Network Systems, 9(3), 2021

GDR COSMOS, Avignon, 21/10/2022

Motivation


Markov Decision Processes: Discrete Stochastic Dynamic Pr... et plus de huit millions d'autres livres sont disponibles pour le Kindle d'Amazon.
‹ Retour aux résultats
Markov Decision Processes: Discrete Stochastic Dynamic Programming Relié – 29 avril 1994
Édition en Anglais de M.L. Puterman (Auteur)
8 évaluations
Afficher tous les formats et éditions
Format Kindle 98,35 €
Relié 149,11 €
Broché 140,50 € prime
Téléchargement —
Lisez avec notre Appli gratuite
4 D'occasion à partir de 130,00 €
1 Neuf à partir de 472,00 €
1 D'occasion à partir de 144,77 €
3 Neuf à partir de 140,50 €
A Markov chain is a sequence of events where the probability of each event is dependent on the event immediately preceding it, but independent of earlier events. Models and numerical equations are used to describe the patterns. This process is particularly useful in operations research and decision science for plotting the sequence of actions which will cause a system to perform optimally. This study provides a unified treatment of the theory, applications and computational methods for Markov decision processes. Important topics featured include action elimination methods, value iteration in the average reward case and sensitive discount optimality.
Signaler des informations incorrectes sur les produits
Voir cette image
Suivre l'auteur
Martin L. Puterman
Suivre
Nombre de pages de l'édition... 672 pages
Langue Anglais
Éditeur John Wiley & Sons Inc
Date de publication 29 avril 1994
Dimensions 16.1 x 3.6 x 24.05 cm
ISBN-10 0471619779
Les produits Probabilités recommandés pour vous
A FIRST LOOK AT RIGOROUS PROBABILITY THEORY
Probability: A Graduate Course
Introduction to Stochastic Calculus
Probability, Random
First Look At Rigorous
Theoretical Foundations
An Introduction to
Probability: A Graduate
Introduction to
ici pour effectuer une recherche

Motivation


Feuilleter
J.K. ROWLING
HARRY POTTER
and the Philosopher's Stone
25 YEARS OF HARRY POTTER MAGIC
Voir cette image
Suivre l'auteur
J.K. Rowling
Suivre
Harry Potter and the Philosopher's Stone, (Inglese) Broché – 19 a
2014
Édition en Anglais de J. K. Rowling (Auteur)
53 015 évaluations
Livre 1 sur 7 : Harry Potter
Afficher tous les formats et éditions
Format Kindle
Emprunt abonnementkindle
Livre audio
0,00 €
Relié
16,10 € prime
Broché
7,40 €
Ce titre et plus d'un million d'autres sont disponibles sur Abonnement Kindle . Le prix d'emprunt à la page est disponible ici.
8,99 € à l'achat
Gratuit avec l'offre d'essai Audible
9 D'occasion à partir de 11,80 €
12 Neuf à partir de 14,81 €
14 D'occasion à partir de 4,00 €
25 Neuf à partir de 5,32 €
Harry Potter has never even heard of Hogwarts when the letters start dropping on the doormat at number four, Privet Drive. Addressed in green ink on yellowish parchment with a purple seal, they are swiftly confiscated by his grisly aunt and uncle. Then, on Harry's eleventh birthday, a great beetle-eyed giant of a man called Rubeus Hagrid bursts in with some astonishing news: Harry
En lire plus
Signaler des informations incorrectes sur les produits
Nombre de pages de l'édition...
352 pages
Langue
Anglais
Éditeur
Bloomsbury Childrens Books
Date de publication
19 août 2014
Dimensions
12.9 x 2.3 x 19.7 cm
ISBN-10
9781408855652
Produits fréquemment achetés ensemble
Prix total: 26,74 €
Ajouter ces trois articles au panier

Model

There are m items to choose.

The reward in our model:

$$\left(\frac{\tilde{\mu}_i \times \#\{\text{item } i \text{ purchases}\}}{\#\{\text{total purchases}\}}\right)^{\alpha}, \quad \tilde{\mu}_i = \mu_i + \zeta_i, \quad i \in [m],$$

was inspired by

Shah, V., Blanchet, J., & Johari, R.
Bandit learning with positive externalities. NeurIPS 2018.

This positive α-homogeneous reward captures 'positive reinforcement' (aka 'herding behaviour' or 'increasing returns') can be either concave or convex utility function.

Model

Agents arrive one at a time. And let

$\xi(n) = i$ if the n-th agent picks object i,

$$x_i(n) := \frac{\#\{\text{item } i \text{ purchases}\}}{n}, \;\; n \geq 1$$

is the fraction of agents who picked i till time n.

$$x_i(n+1) = x_i(n) + \frac{1}{n+1}\left(\mathbb{I}\{\xi(n+1) = i\} - x_i(n)\right), \qquad (1)$$

with $x_i(0) = \frac{1}{m}$ $\forall i$.

This can be viewed as a stochastic approximation iteration.

Model

Graphical constraints: We assume that the choice in the $(n+1)$-st time slot is constrained by the choice made in the n-th slot.

E.g., given the present choice, only some selected 'nearby' or 'related' choices are recommended. (Of course, the complete graph is a particular case.)

We consider an undirected graph $G = (V, E)$ where V, E are resp., its node and edge sets, with $|V| = m$; and assume that G is irreducible with self-loops.

Model

Selection Policy: The agents choose items according to

$$\mathbb{P}(\xi(n+1) = j|\mathcal{F}_n) = (1-\varepsilon(n))\tilde{p}^{\alpha}_{\xi(n)j}(x(n)) + \varepsilon(n)\chi_j(\xi(n)). \tag{2}$$

Here:

$$\tilde{p}^{\alpha}_{ij}(x) := \mathbb{I}\{j \in \mathcal{N}(i)\}\frac{\hat{f}^{\alpha,n}_j(x)}{\sum_{l\in\mathcal{N}(i)} \hat{f}^{\alpha,n}_l(x)}\ , \tag{3}$$

for $\hat{f}^{\alpha,n}_i(x) := (\hat{\mu}_i(n)x_i(n))^{\alpha}$, where

$$\hat{\mu}_i(n) := \frac{\sum_{k=0}^{n} \mathbb{I}\{\xi(k) = i\}\tilde{\mu}_i(k)}{\sum_{k=0}^{n} \mathbb{I}\{\xi(k) = i\}}$$

is the empirical estimate of μ_i at time n recursively computed by

$$\hat{\mu}_i(n+1) = \left(1 - \frac{1}{S_i(n+1)}\right)\hat{\mu}_i(n) + \frac{\tilde{\mu}_i(n+1)}{S_i(n+1)},$$

if $\xi(n+1) = i$ and remains the same otherwise.

Convergence analysis

We first note that $\hat{\mu}_i(n) \to \mu_i$ a.s. $\forall i$.

Thus a.s., $\lim_{n\uparrow\infty} \hat{f}_i^{\alpha,n}(x) = f_i^\alpha(x) := (\mu_i x_i)^\alpha$ and

$$\lim_{n\uparrow\infty} \tilde{p}_{ij}^\alpha(x) = p_{ij}^\alpha(x) := \mathbb{I}\{j \in \mathcal{N}(i)\} \frac{f_j^\alpha(x)}{\sum_{l\in\mathcal{N}(i)} f_l^\alpha(x)}.$$

As $\alpha \downarrow 0$, the process approaches a standard random walk on the graph that picks a neighbor with equal probability.

As $\alpha \uparrow \infty$, the process at i will (asymptotically) pick the $j \in \mathcal{N}(i)$ for which $\mu_j x_j = \max_{k\in\mathcal{N}(i)} \mu_k x_k$, uniformly.
I.e., the process will tend to select a neighbour item greedily.

Convergence analysis

We first analyze the convergence for fixed α by the standard stochastic approximation techniques based on limiting ODE.

Let $\varphi_i^\alpha(x) := f_i^\alpha(x) \sum_{j \in \mathcal{N}(i)} f_j^\alpha(x)/x_i$ and consider the ODE

$$\dot{x}_i(t) = \frac{x_i(t)\varphi_i^\alpha(x(t))}{\sum_k x_k(t)\varphi_k^\alpha(x(t))} - x_i(t). \tag{4}$$

This is also the ODE limit for Vertex Reinforced Random Walk introduced by Pemantle and analyzed further by Benaim.

Convergence analysis

Note that every equilibrium of (4) satisfies the fixed point equation

$$\pi(i) = h_i(\pi) := \frac{f_i^\alpha(\pi) \sum_{j \in \mathcal{N}(i)} f_j^\alpha(\pi)}{\sum_k f_k^\alpha(\pi) \sum_{\ell \in \mathcal{N}(k)} f_\ell^\alpha(\pi)}, \quad \forall i \in [m]. \qquad (5)$$

There can be more than one equilibrium point.

Convergence analysis

Further note that ODE (4) has the same trajectories and the same asymptotic behavior as

$$\dot{z}_i(t) = z_i(t)\left(\varphi_i^\alpha(z(t)) - \sum_j z_j(t)\varphi_j^\alpha(z(t))\right), \qquad (6)$$

i.e., $z(t) = x(\tau(t))$ for some $t \in [0,\infty) \mapsto \tau(t) \in [0,\infty)$ which is strictly increasing and satisfies $t \uparrow \infty \Longleftrightarrow \tau(t) \uparrow \infty$.

Now we recognize that ODE (6) is the replicator dynamics.

Convergence analysis

Let $A := [[a_{ij}]]_{i,j \in \mathcal{V}}$ be the (symmetric) adjacency matrix of G. Then, for $x = [x_1, \cdots, x_m] \in \mathcal{S}_m$,

$$\varphi_i^\alpha(x) = \frac{\partial}{\partial x_i}\Psi^\alpha(x) \text{ for } \Psi^\alpha(x) := \frac{1}{2\alpha}\sum_{i,j} a_{ij} f_i^\alpha(x) f_j^\alpha(x).$$

Thus (6) corresponds to the replicator dynamics for a potential game with potential $-\Psi^\alpha$.

We make the following technical, generically true, assumption: The equilibrium points (5) are isolated and hyperbolic, i.e., the Jacobian matrix of h at these points does not have eigenvalues on the imaginary axis.

Convergence analysis

Theorem

For each $\alpha > 0$, the local maxima of $\Psi^\alpha : \mathcal{S}_m \mapsto \mathbb{R}$ are stable equilibria of (4) and the iterates of (1) converge to the set thereof, almost surely. In particular, the probability of convergence of $\{x(n)\}$ in (1) to any local maximum of Ψ^α in $\mathcal{S}_m$ is strictly positive.

Convergence analysis (annealed dynamics)

Taking inspiration from 'simulated annealing', let us now slowly increase α or equivalently slowly decrease the 'temperature' $T(n) = 1/\alpha(n)$ as follows:

$$T(n+1) = (1 - b(n))T(n), \;\; n \geq 0, \tag{7}$$

where $1 > b(n) \downarrow 0$ are stepsizes satisfying

$$\sum_n b(n) = \infty, \;\; nb(n) \stackrel{n\uparrow\infty}{\rightarrow} 0, \;\; b(n) = o(c(n)). \tag{8}$$

The second condition implies $\sum_n b(n)^2 < \infty$.

Convergence analysis (annealed dynamics)

A behavioral interpretation is that the agents exhibit a herd behavior, weighing in public opinion more and more with time.

Let

$$D := \{i \in V : \mu_i = \max_j \mu_j\}.$$

Theorem

$\sum_{i \in D} x_i(n) \to 1$ *a.s.*

The proof is based on stochastic approximation technique with multiple time scales.

The case of complete graph

In the case of complete graph (no graphical constraints),

$$\Psi^{\alpha}(x) = \left(\sum_i f_i^{\alpha}(x) \right)^2,$$

is convex for $\alpha \geq 1$. Nice observation...

However, for $\alpha \in (0, 1)$, the problem is equivalent to max $\sum_i f_i^{\alpha}(x)$, which is strictly concave and we can say more.

The case of complete graph

For $\alpha \in (0, 1)$, using the Lagrange multiplier technique, we can even obtain an explicit expression for the unique stationary point:

$$x_i(\infty) = \frac{\mu_i^{\alpha/(1-\alpha)}}{\sum_{k=1}^{m} \mu_k^{\alpha/(1-\alpha)}} \,. \tag{9}$$

From (9), as $\alpha \to 1$, the frequencies $x_i(\infty)$ start to concentrate on D.

Numerical examples

Let us illustrate the importance of annealing for convergence of $x(n)$ to D.

Consider a graph composed of two cliques (2 and 8 nodes) connected through a single edge.

We set $\mu_i = 1$ for $i \in$ clique-1 and $\mu_i = 0.5$ for $i \in$ clique-2.

Cooling schedule:

$$\alpha(n+1) = \alpha(n) \left(1 - \frac{1}{n \log n}\right)^{-1}$$

Trembling hand parameter:

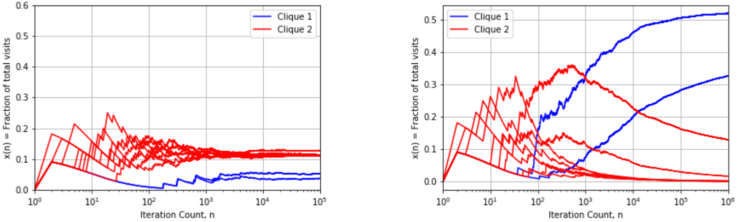
$$\varepsilon(n) = \frac{1}{\log(n+1)}$$

Numerical examples

Some points to note:

- If we initialize the walk in clique-2 and *do not* increase $\alpha \to \infty$, then the relative frequencies converge to non-zero values for nodes in clique-2. (We have set $\alpha = 10$.)
- If we initialize the walk in clique-2 and *do* increase $\alpha \to \infty$, then the chain moves to clique-1 and stays there.

Numerical examples



(a) Initialize in clique-2, α fixed. (b) Initialize in clique-2, $\alpha \to \infty$.

Figure: Fraction of Total Visits, $x(n)$ Vs. Iteration Count for the two clique experiment

Numerical examples

For the unconstrained case, we have tried $\mu = (2, \frac{1}{4}, \frac{1}{2}, 1)$ with the fixed $\alpha = 0.85 < 1$.

The dynamics always converges to the stationary point $(0.98, 0.000, 0.000, 0.019)$.

This demonstrates that in the unconstrained case for the values of $\alpha < 1$ even not so close to one, a very significant portion of the mass is concentrated on the optimal node!

Relations to MABs

With linear topology and the rewards $\mu = (2, \frac{1}{4}, \frac{1}{2}, 1)$, let us make an important comparison with the multiarmed bandit literature.

The ϵ-greedy policy has a stationary distribution that is seen to concentrate equally on items $1, 4$ as $\epsilon \downarrow 0$. In particular, it is a suboptimal distribution.

Of course, in the fully connected case the $\varepsilon(n)$-greedy policy with $\varepsilon(n) = \frac{1}{n}$ converges to the optimal, as shown in Theorem 3 of (P. Auer *et.al.*, 2002). Thus, a standard bandit algorithm can fail in the graph-constrained framework.

New book (2022)

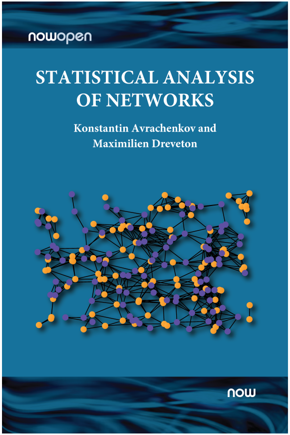


Table of Contents:

Thank you!

Any questions?