

За да намерим модел, трябва да се откажем от съвършенството



СЕМИНАР
„АЛГЕБРА И ЛОГИКА“

Димитър Добрев

d@dobrev.com

Институт по математика и информатика
Българска академия на науките

Каква е целта на този доклад?

Целта е да ви запознае с една слабо изследвана област от математиката. Тази област можем да наречем „Търсене на модел“.

Защо това е слабо изследвана област?

Интересно ли е търсенето на модели?

Какво представлява търсенето на модел?

Това е половината от Изкуствения Интелект. Другата половина е търсенето на стратегия.

Все пак, защо това е слабо изследвано?

Четири основни грешки, които масово се допускат

1. Търси се модел без памет.

Тоест, предполагаме че светът е постоянен и че не се променя или че виждаме всичко.

2. Не се търси модел, а стратегия.

Не си задаваме въпроса „Какво се случва?“, а само въпроса „Какво да правя?“

3. Търси се съвършен модел.

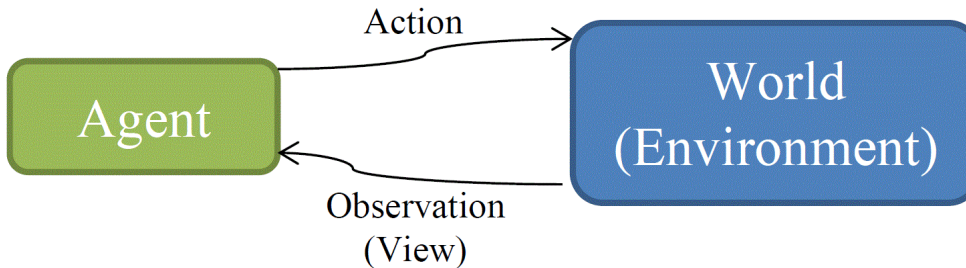
Търсим модел, който описва света напълно и еднозначно.

4. Предсказва се бъдещето, но не се предсказва миналото.

Вървим по посока на стрелките, но не и в обратна посока.

Постановка на задачата

1. Какво е дадено?



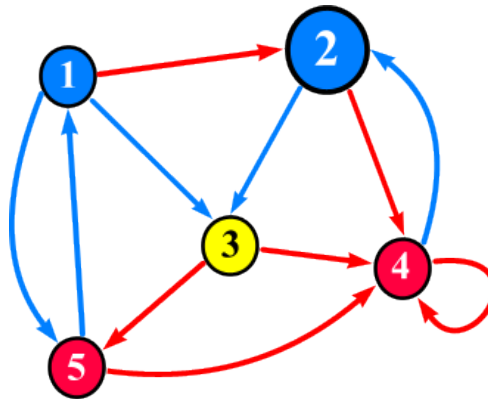
$\dots, V_{-2}, a_{-2}, V_{-1}, a_{-1}, V_0, a_0, V_1, a_1, V_2, a_2, \dots$

Това взаимодействие можем да си го мислим като *наблюдение-действие* или като *въпрос-отговор*.

Постановка на задачата

2. Какво търсим?

Търсим модел на света. Моделът ще бъде ориентиран граф.



На стрелките ще съответстват някакви причини за промяната на състоянието на света. Тези причини ще наречем „събития“.

Четири основни грешки

1. Търси се модел без памет.

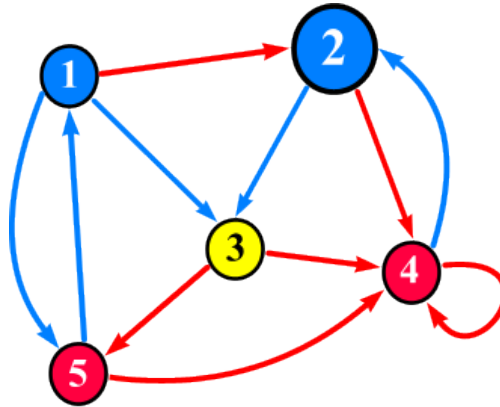
Динамична памет

Каква е разликата между постоянна и динамична памет?

В седмицата има седем дни. (постоянна памет)

Днес е петък. (динамична памет)

Full Observability vs Partial Observability



Колко голяма е динамичната памет на модела? Колко бита?

$$\log_2 |S| \text{ бита}$$

2. Не се търси модел, а стратегия.

„Какво се случва?“ – търсим модел.

„Какво да правя?“ – търсим стратегия.

Reinforcement Learning без reinforcement

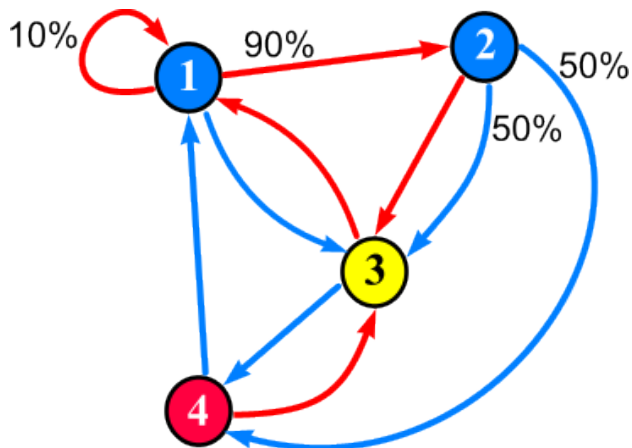
Повечето автори дефинират MDP специално за Reinforcement Learning, като добавят и цел. Тоест добавят rewards и discount factor.

Ако търсим стратегия, ние не можем без цел, но ако търсим само модел тогава целта не ни е нужна. Няма да имаме цел и единствената ни цел ще е познанието.

3. Търси се съвършен модел.

Модел, който описва света напълно и еднозначно.

Markov decision process (Марковски процес на взимане на решения)



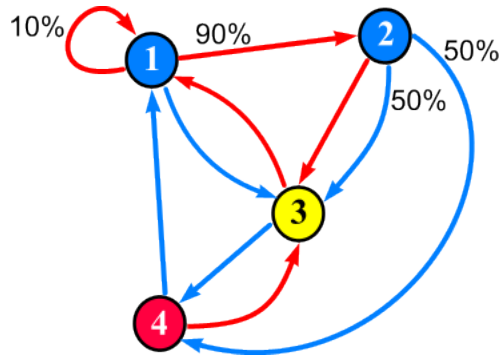
На всяка стрелка трябва да съответстват по две вероятности.

Моделът MDP е съвършен!

Дефиниция. Съвършен модел ще бъде този, който ни дава съвършено описание на бъдещето, при условие че стратегията на агента е фиксирана.

Дефиниция. Съвършено описание на бъдещето ще наричаме множеството *Future* всеки един от елементите, на което има вида $\langle \omega, p \rangle$, където ω пробягва възможните развития на бъдещето, а p е вероятността това развитие да се случи ($p > 0$).

$Future = \{ \langle \omega, p \rangle \mid \omega \text{ е възможно развитие на бъдещето, } p \text{ вероятността } \omega \text{ да се случи} \}$



Какъв модел на света търсим?

Дали търсим генератор или дескриптор?

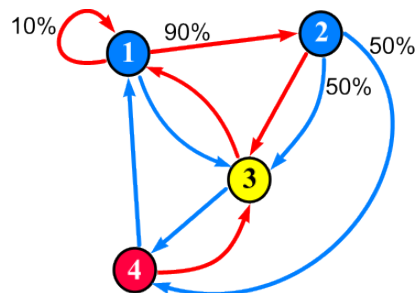
Ако търсим генератор, то той трябва да е съвършен и да описва света напълно. За да създадем света (да го генерираме) на нас ни е нужно пълното му описание. Да, но на нас не ни е нужно да създаваме света, защото той вече е създаден и ние трябва само да го разберем.

Ние ще се опитаме да намерим дескриптор, който описва света частично, като казва нещо, без да каже всичко. Ще се откажем от намирането на генератор (пълното описание на света) и ще търсим дескриптор (някакво частично описание).

а) Предполага се, че светът няма свободна воля.

Агентът прави каквато си поиска (избира следващия си ход с вероятност в интервала $[0, 1]$).

Светът за всеки възможен ход има точно определена вероятност.



Предполагаме, че агентът има свободна воля, а светът е задължен да следва една точно определена стратегия.

Дефиниция. Стратегия на агента ще наричаме функция, която на всяко състояние и възможно действие ни дава вероятността това действие да бъде извършено от агента.

$$Policy : S \times \Sigma \rightarrow \mathbb{R}$$

Дефиниция. Детерминирана стратегия ще наричаме такава, която дава вероятност единица за една от възможностите и вероятност нула за останалите.

Дефиниция. Ограничение на агента ще наричаме функция, която на всяко състояние и възможно действие ни дава интервал, в който се намира вероятността това действие да бъде извършено от агента.

$$\textit{Constraint} : S \times \Sigma \rightarrow \mathbb{R} \times \mathbb{R}$$

Напълно свободна воля имаме когато интервалът е $[0, 1]$. Свободната воля липсва напълно, когато интервалът е с дължина нула (тоест, когато вероятността е точно определена).

Ще обобщим MDP и ще получим един модел, който ще наречем MDP плюс свободна воля и непрогнозируема случайност (за по-кратко MDP Plus).

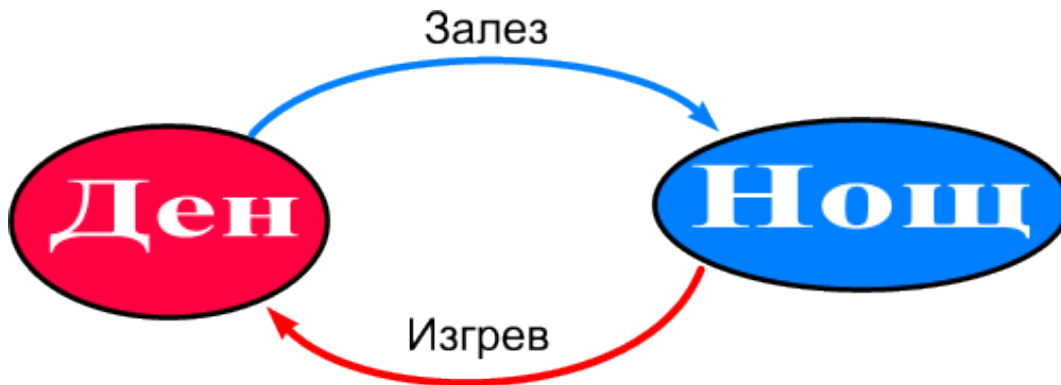
В този модел и светът и агентът имат свободна воля, но тя не е неограничена (т.е. поставена е в някакви граници).

б) Предполага се свойството на Марков, което гласи:

„Бъдещото развитие зависи само от текущото състояние и не зависи от това как сме стигнали до това текущо състояние.“

Това свойство се нарича още „безпаметното свойство“ и означава, че повече памет не ни е нужна. Всичко, което е било нужно да се запомни, е запомнено в текущото състояние. Това означава, че моделът не може да се подобри. Ако имаше значение как сме достигнали до определено състояние, щяхме да разделим състоянието на две и по този начин да подобрим модела.

в) Предполага се, че само действията на агента ни интересуват. Освен действията на агента, може да има и други интересни събития.



Състоянията се наблюдават за известен период от време, а не само за една стъпка.

Така получените event-driven модели ще описват света много грубо. Например, дали е ден или нощ, но в нашия свят има още много други интересни неща.

г) Предполага се, че моделът се наблюдава постоянно.

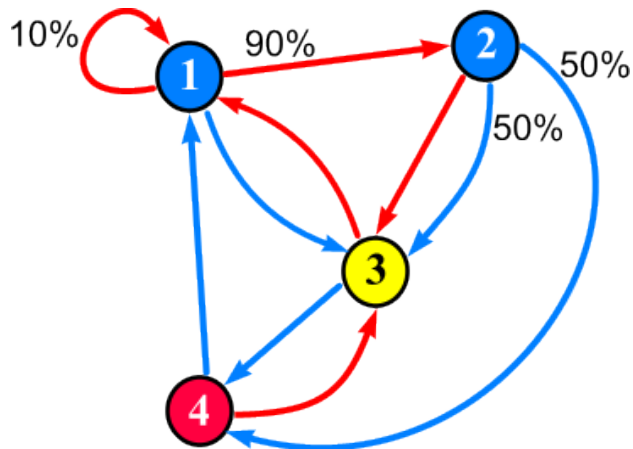
Една зависимост може да се наблюдава постоянно, а може да се наблюдава от време на време.



Дефиниция. Явление ще наричаме модел, който не е валиден непрекъснато (през целия живот), а само през определени интервали от време.

д) Предполага се, че следата е съвършена.

Във всяко състояние нещо се случва и знаем точно какво се случва.



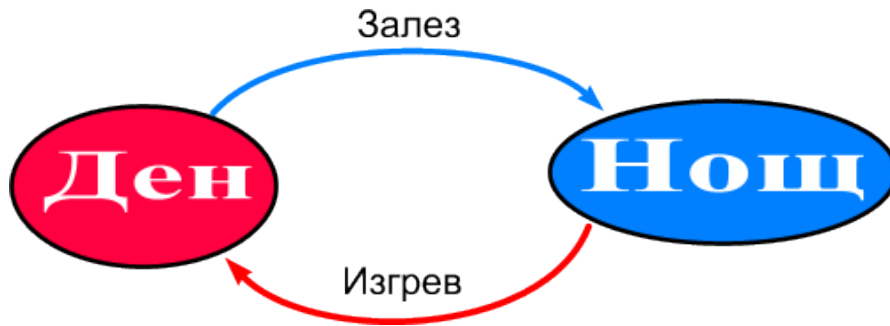
Съвършената следа е тотална и пълна.

Може да виждаме явление или обект.

Следата може да има памет.

Събитийни модели

Event-Driven Models



Вместо един съвършен MDP модел, ние ще търсим голям брой простички event-driven модели.

Най-ранните работи относно event-driven моделите:

1. Xi-Ren Cao (2005). Basic Ideas for Event-Based Optimization of Markov Systems. *Discrete Event Dynamic Systems: Theory and Applications*, 15, 169–197, 2005.
2. Xi-Ren Cao, Junyu Zhang (2008). Event-Based Optimization of Markov Systems. *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, VOL. 53, NO. 4, MAY 2008.

Предишната дефиниция на събитие:

Това е произволна булева функция $Event(t)$.

Новата дефиниция на събитие:

Това е подмножество от множеството на стрелките на някой модел-генератор.

Как описваме пукнатина в статуетка?

Предполага се, че моделът генератор е единствен.
Не се интересуваме от това какво „знаят“ състоянията.

Минимален модел ще бъде този, при който състоянията знаят минималното (не знаят нищо излишно).

Максимален модел ще бъде този, при който състоянията знаят всичко за миналото и за бъдещето.

Бихме могли да допуснем съществуването на модели, в които състоянията знаят дори повече от тези в максималния модел.

Как отчитаме събитията?

Имаме теоретичната дефиниция на събитие, която не ни върши работа.

Всеки event-driven модел има няколко наблюдавани събития. Трябва да се научим да отчитаме когато тези събития се случват. Това ще става по два начина – пряко и косвено.

Пряко ще ги откриваме с характеристична функция, а **косвено** ще ги откриваме чрез следата (тоест, по това което се случва, ще заключим, че сме преминали в друго състояние и оттам, че се е случило събитие).

Прякото откриване ни дава точния момент, в който събитието се е случило и ни дава възможност да описваме модели с примки.

При косвеното откриване точния момент на настъпване е по-трудно да се определи и не можем да забележим примка.

Какво е обект?

Предишната дефиниция на обект:

Обектът е явление.

Тоест, зависимостта, която се наблюдава, когато наблюдаваме обекта.

Дефиниция. Свойство ще наричаме явление, което се случва, когато наблюдаваме обект от групата обекти имащи това свойство.

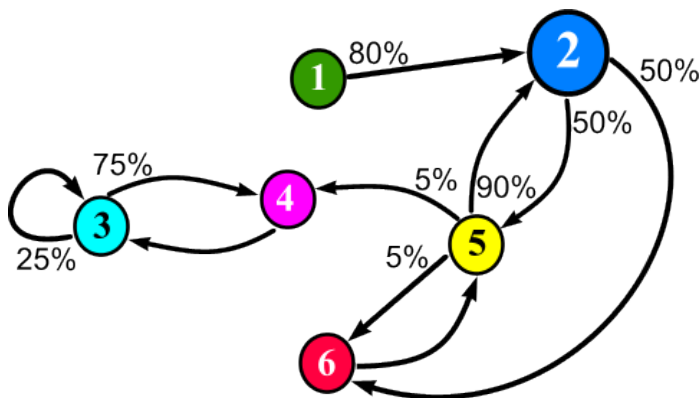
Новата дефиниция на обект:

Обектът е абстрактно понятие. Обектът има някакви свойства. Свойствата на обекта не го характеризират. Тоест, два различни обекта могат да имат еднакви свойства.

4. Предсказва се бъдещето, но не се предсказва миналото.

Можем ли да обърнем стрелките и да получим модел предсказващ миналото?

Можем, ако нямаме „бели върхове“.



Дефиниция. Черна дупка ще наричаме множество от състояния, при което няма път от състояние от множество до състояние извън множеството. Ще искаме още множеството да е непразно и да не съдържа началното състояние.

Можем ли да обърнем стрелките на MDP модела и да получим модел предсказващ миналото?

Можем да обърнем MDP Plus модела и да получим MDP Plus модел, който да предсказва миналото. (Пак ще предположим, че нямаме бели върхове, защото в противен случай обратният MDP Plus модел няма да е единствен.)

Обратния на MDP модела не е MDP модел (има обратен, но той е MDP Plus модел). Затова обикновено не се търси обратния на MDP модела.

Нека агентът има напълно свободна воля, а светът е ограничен да спазва една точно определена стратегия (това е MDP модела). Тогава като обърнем стрелките получаваме агент, който има някакви ограничения и свят който има известна свобода (тоест получаваме MDP Plus, а не MDP модел).