

Некоректни ходове и проверими състояния

Димитър Добрев
Пролетна научна сесия на ФМИ
25 март, 2017

Reinforcement learning (Обучение с утвърждение)

Имаме една последователност:

ИЗХОД, ВХОД, ИЗХОД, ВХОД ...

действие, наблюдение, действие, наблюдение, ...

action, view, action, view ...

$a_1, v_1, a_2, v_2, \dots$

a_i и v_i са вектори от скалари

Искаме да разберем тази последователност.

Играем игра и играта има вътрешно състояние
(позиция на таблото)

$s_0, s_1, s_2, s_3, \dots$

$s_{i+1} = \mathbf{World}(s_i, a_{i+1})$

$v_i = \mathbf{View}(s_i)$

Разглеждат се два случая:

full observability (виждаме всичко) $s_i = \mathbf{View}(s_i)$

partial observability (виждаме частично) $s_i \neq \mathbf{View}(s_i)$

Проверими състояния (това е резултата от някакъв експеримент)

Вратата е заключена

Ако натисна дръжката \Rightarrow вратата ще се отвори.

Котлона е горещ

Ако пипна котлона \Rightarrow ще се опаря.

Покрива е за поправка

Ако завали \Rightarrow ще протече.

Проверимото състояние е нещо реално



Кога е: И да и не?

Хлабав експеримент или недетерминирана функция *World*.

Проверимото състояние не го получаваме наготово

Не можем директно да го измерим, а трябва да го апроксимираме с теория.

Какво представлява теорията?

При известни обстоятелства дадено проверимо състояние е истина.

Ако проверимото състояние е $A \Rightarrow B$ и известните обстоятелства са S , то импликацията $S \& A \Rightarrow B$ може да е част от теорията.

Ние събираме статистика за определени конюнкции. Как получихме импликация?

Ще означаваме с $|A|$ колко пъти конюнкцията A е била истина.
Ако $|S \& A| = |S \& A \& B|$, тогава е вярна импликацията $S \& A \Rightarrow B$.
За да имаме вяра на тази импликация ще искаме $|S \& A| > 10$.

Аналогично, при известни обстоятелства дадено проверимо състояние може да е:

лъжа,

или истина с вероятност по-голяма от някакво число **a**,

или истина с вероятност по-малка от някакво число **b**

Какви са възможните стойности, които теорията може да даде (да прогнозира) за проверимото състояние:

вярно,

невярно,

вярно с вероятност по-голяма от някакво **a**,

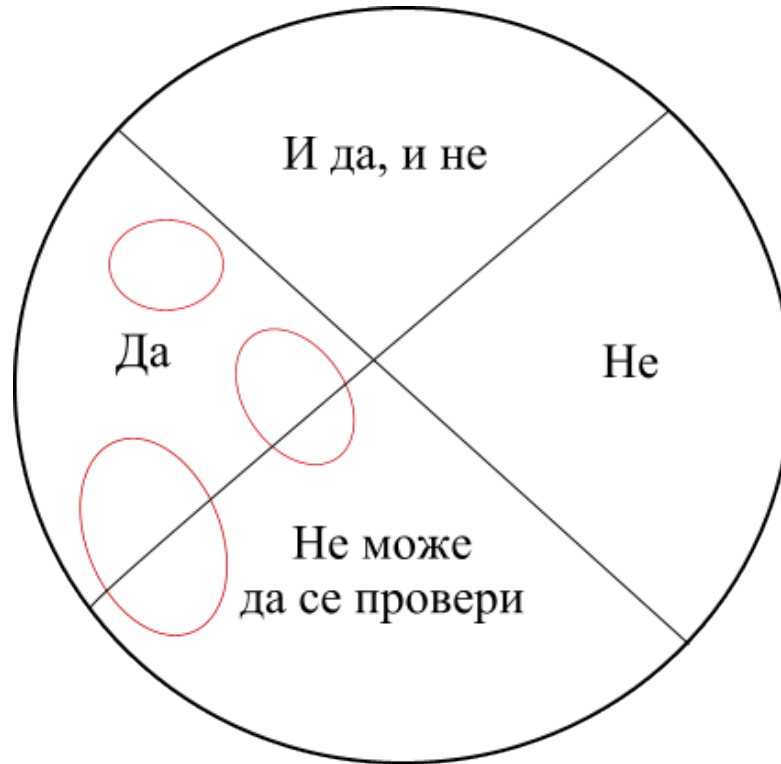
вярно с вероятност по-малка от някакво **b**,

вярно с вероятност в интервала [**a**, **b**],

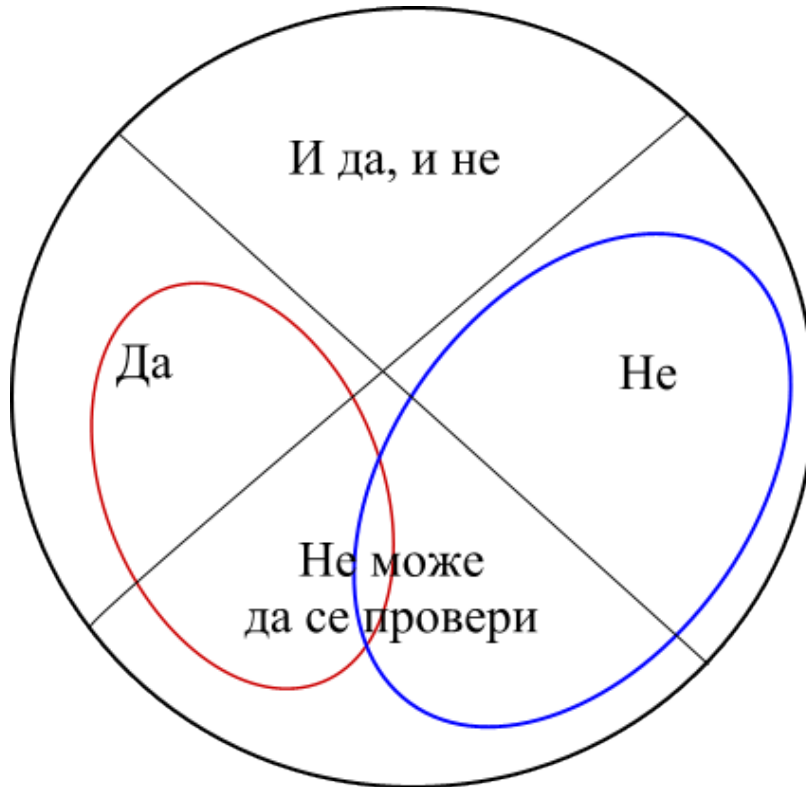
неизвестно (теорията нищо не предсказва)

неизвестно (теорията е противоречива, т.е. предсказва прекалено много)

Как от импликациите правим теория:



Обединяваме всички „известни обстоятелства“



Известните обстоятелства може да са конюнкция само от миналото, а може да са от миналото и от бъдещето.

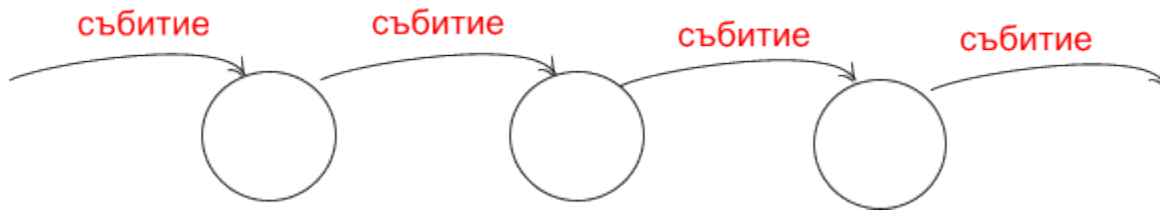
Пример: Виждаме, че една врата е заключена. Коя точно е заключената врата разбираме в последствие, след като прочетем какво пише на табелката.

Известните обстоятелства може:

да са проста конюнкция без памет – да се отнасят само за няколко стъпки преди и няколко стъпки след текущия момент

да са сложна конюнкция с памет – да се отнасят за дълъг период преди и след текущия момент

Простата конюнкция може да бъде разложена на събития, които се случват в няколко последователни стъпки, докато сложната конюнкция има вида:



Събитие, което се случва на някоя стъпка, след това период от време, когато определени събития не се случват, после пак събитие, период, събитие, период.

Пример: „След белия диван, надясно, вратата е заключена.“
Това правило означава, че в един момент сме видели бял диван, след което известен период от време не сме завивали нито на ляво, нито на дясно, след което сме завили на дясно.

Това правило може да описва една конкретна врата, а може да описва и няколко врати (които са след бели дивани). Може дори да описва една врата, но само понякога, когато белият диван си е на мястото.

Представете си, че се движим в замък, в който едни врати са заключени, а други отключени. Определените обстоятелства ни казват пред коя врата (или врати) се намираме. Картата на замъка ни казва какви са тези определени обстоятелства, но задачата е обратната. По определените обстоятелства трябва да съставим картата на замъка.

Нека си представим, че вратите не са постоянно заключени или отключени, а се отключват (заключват) от време на време. Тогава теорията, която ще използваме, ще предполага известна устойчивост.

Нека си представим, че в замъка живее едно същество, което отключва и заключва вратите по свое усмотрение. Тогава ще ни трябва теория, която да предвиди поведението на това същество.

Заключението е, че ние можем да представим състоянието на света чрез проверими състояния, които са различни в различните състояния на някакви крайни автомати. Освен това, че проверимите състояния са инертни и когато скоро сме ги проверявали е вероятно те да имат още същата стойност. Ще предполагаме, че в света има и други агенти, освен нас и че те могат да променят определени проверими състояния според собствената си воля и собствените си цели.

След като сме направили теория, която адекватно описва съответно проверимо състояние ние можем да събираме статистика на базата на тази теория. По този начин ние ще разберем света (играта).