UNIVERSITAT DE LLEIDA
Escola Politècnica Superior
Grau en Enginyeria Informàtica
Models de Computació i Complexitat

Restless Bandit

Joaquim Picó Mora, Sergi Simón Balcells

PraLab2

Professorat : M.Valls

Data : Divendres 27 de Març

Contents

1 Introducció

2 Context

El teu estomac gruny. Vas al restaurant Italià que t'encanta, o proves el nou Thailandes que acaba d'obrir? Hi vas amb el teu millor amic o amb una persona que no coneixes tant i que vols coneixer millor? Masa dificil, millor quedar-se a casa. Cuines aquella recepta que t'encanta, o optes per buscar-ne alguna de nova per Internet? Saps que, potser millor demanar una pizza a domicili. Demanes la teva preferida o preguntes per les especials? Tant dubtar entre una o l'altra te n'hauràs cansat abans de poguer fer la primera mossegada.

Cada dia ens veiem forçats a fer decisions entre dos opcions, que difereixen en dos dimensions: Ens quedem amb les nostres coses preferides, o n'explorem de noves? Intuitivament podem pensar que la vida és un balanç entre les dues, pero la pregunta és: Quin és el balanç?

Molts matemàtics i informàtics han estat treballant en aquest balanç des de fa més de 50 anys donant-li el nom de explore/exploit tradeoff.

En computació, la tensió entre explorar o explotar pren la seva forma més concreta an l'escenari anomenat multi-armed bandit, o k-armed bandit. Aquest nom li és donat degut a que es la forma coloquial de referir-se a les màquines escura-butxaques. Imagina entrar a un cassino ple de màquines escura-butxaques, cada una amb les seves possibilitats de fer una tirada guanyadora. Naturalment, s'està interessat en maximitzar els guanys. Està clar que hi haura una fase d'exploració en la qual testegarem les màquines, i una altra d'explotació tirant d'aquelles que creiem que són més beneficioses.
La primera passa cap a la solució va ser l'algorisme Win-Stay, Lose-Shift, proposat per Herbert Robins. Aquest consisteix en triar a una màquina aleatoria, mentres s'obtingui profit jugant en aquella màquina, es continua jugant en la

mateixa i, si després d'una certa tirada la màquina deixa de ser profitosa, es canvia a una altra. Aquesta tot i estar lluny d'una solució optima, es va demostrar que els resultats eren millors que els de la pura sort.
No va ser fins al 1970 que John Gittins va trobar una solució optima que resolia el problema. Gittins va enfocar el problema en termes de maximitzar els guanys per un futur que és interminable pero amb 'descontes'. Fent així la asumsió de que el valor assignat als guanys decreixia geométricament. Per exemple, es creu que hi ha un 1% de probabilitats de ser atropellat per un autobus un día, aleshores s'ha de valorar el sopar del seguent día un 99% del valor del d'aquesta nit, només perque l'endemà potser mai s'arriba a sopar. D'aquesta forma, va arribar a la conclusió de que cada màquina de la qual en sabem una mica o res, té un nombre que ens indica la probabilitat de guany que ens farà decidir si tornar a jugar en ella o no. Aquest nombre és conegut com l'index de Gittins.

Una variació d'aquest problema(multi-armed bandit), és que cada una de les màquines escura-butxaques es comporta com una màquina Markov. Es a dir, cada cop que una màquina en particular és jugada, l'estat d'aquesta canvia a un nou escollit dacord a l'evolució de probabilitats dels estats d'aquesta màquina de Markov. I si a la variació anterior se li aplica, que l'estat de les màquines no jugades pot evolucionar al llarg del temps, apareix el problema del restless bandit.

3 Desenvolupament de la temàtica

Dins de les complexitats que poden tenir els problemes, la complexitat de PSPACE compleix:

$$NP \subseteq PSPACE \subseteq EXP$$

De la mateixa manera que no es sap si $P = NP$ tampoc es sap si $NP = PSPACE$. La demostració que NP es contingut dins de PSPACE es realitzar per reducció a l'absurd:

Sigui M una màquina de Turing NP, és a dir, donada una instància d'un problema ens diu en un temps polinomial si aquest pertany al problema. Si l'espai per a desenvolupar l'algorisme fos més gran a polinòmic, llavors forçosament per a llegir o escriure aquesta informació es necessitaria aquest temps, pel que arriba a l'absurd amb la definició de la màquina M.

El problema de *Restless bandit* es troba dincs la complexitat PSPACE. Això va ser demostrat l'any 1999 en l'article [].En aquest s'explica un problema de xarxes que es demostrar ser exponencial. Al mateix temps, un problema relaxat d'aquest es demostra ser PSPACE-complet, és a dir, tots els problemes de PSPACE poden ser reduïts a aquest problema i aquest pot ser reduït a tots els problemes de PSPACE. Finalment es dóna una fórmula de cost del problema de *restless bandit* i es redueix el problema relaxat a aquest, demostrant que és PSPACE-hard (tots els problemes de PSPACE són reduïbles a aquest).

4 Conclusions

5 Bibliografia