

**Budapesti Műszaki és Gazdaságtudományi
Egyetem**

Természettudományi Kar
Sztochasztika tanszék



**Bandita algoritmusok
sztochasztikus környezetben**

BSc Szakdolgozat kivonat

Kiss Botond

Témavezetők:

Dr. Csáji Balázs Csanád

MTA SZTAKI

Tudományos főmunkatárs

Dr. Pete Gábor

BME TTK Sztochasztika Tanszék

Docens

2019

A gépi tanulás egyik fő ága a megerősítéssel való tanulás, melyben a tanulás folyamata egy dinamikus, bizonytalan környezettel való interakció során valósul meg. A dolgozatban tárgyalt bandita problémakör a megerősítéssel való tanulás egyik alapvető esete, és hatékony keretrendszert nyújt olyan algoritmusoknak, melyek szekvenciálisan hoznak döntést bizonytalan környezetben. Emellett jól szemlélteti a megerősítéssel való tanulás egyik központi problematikáját, a felderítés és kihasználás dilemmáját. A tanulás kulcsa sztochasztikus banditák esetén az, hogy a környezet minden döntéshozatal (akció választás) után jutalmat mintavételez, és a tanuló célja, hogy ezen visszacsatolások által maximalizálja a kumulatív jutalmat. A sztochasztikus abban rejlik, hogy a teljes interakció során egy adott döntésre kapott jutalmak sorozata független és azonos eloszlású.

Az 1. fejezetben bemutatásra került azon keretrendszer, melyen belül a többkarú sztochasztikus bandita algoritmusokat tárgyaljuk. Megismerkedtünk a megbánás fogalmával, amely segítségünkre volt az algoritmusok hatékonyságának mérésében. Áttekintettük, hogy milyen lehetőségek vannak arra, hogy mit feltételezzünk arról a környezetről, amellyel interakcióba lépünk.

A 2. fejezetben átfogóan tárgyaltuk a szubgauss valószínűségi változókat, melyek gyorsan lecsengő farokvalószínűsége nyújtott segítséget hatékony algoritmusok tervezésében. Ismertettük az alapvető tulajdonságaikat, melyeket felhasználva független azonos eloszlású szubgauss minta alapján egyoldali konfidencia intervallumot szerkesztettünk a közös várhatóértékre.

Ezek után a 3. fejezetben négy különböző többkarú sztochasztikus bandita algoritmust tárgyaltunk. Elméleti szempontból a szubgauss tulajdonság segítségével az algoritmusok megbánását vizsgáltuk, melyekre szublineáris felső korlátot adtunk. Mind a négy algoritmust implementáltuk Python nyelven, és numerikus kísérletek során hasonlítottuk őket össze megbánás, és gyakorlati alkalmazhatóság szempontjából. A kapott eredményeket részletesen értékeltük. Az ETC algoritmus kitűnően szemléltette a felderítés és kihasználás közti átmenetet, és hatékonynak bizonyult két kar esetén a szuboptimális rések ismeretében. Az ETC felderítési fázisa randomizálással javítható. Az így kapott ε -mohó algoritmus szolgáltatja a legalacsonyabb megbánást a kísérletek során két kar esetén. A korábban említett konfidencia intervallum segítségével bemutattuk, hogy hogyan tervezhető meg a híres UCB algoritmus. Végül ennek az aszimptotikusan optimális változatát is tárgyaltuk. A kísérletek során láttuk, hogy az aszimptotikusan optimális UCB algoritmus kiküszöböli minden hiányosságot, melyet az ETC típusú algoritmusok és az UCB egyszerű változata magukban hordoznak, ugyanis az aszimptotikusan optimális UCB esetén nincsenek optimalizálandó hiper-

paraméterek, továbbá kettőnél több kar esetén is jól viselkedik megbánás szempontjából.

A 4. fejezetben általánosítottuk az 1. fejezetben bemutatásra került sztochasztikus bandita modellt. Így az kapott keretrendszerben az interakció során rendelkezésre áll körönként a kontextus, mint többlet információ. Bemutattuk, hogy miként adaptálható a 3. fejezetben látott legkisebb megbánással rendelkező UCB algoritmus lineáris jutalomfüggvénnyel és kontextussal rendelkező környezetre (LinUCB algoritmus). Míg a többkarú esetben konfidencia intervallumra volt szükségünk az UCB típusú stratégiákhoz, addig a lineáris UCB esetén konfidencia ellipszoidra, melynek középpontja az ismeretlen paraméter becslése, amit a regularizált legkisebb négyzetek módszere szolgáltatott a korábbi akciók és jutalmak függvényében. Az említett ellipszoidok sorozatának segítségével megmutattuk, hogy hogyan adható meg a LinUCB algoritmusra ismert szublineáris felső korlát a megbánásra. Összességében a sztochasztikus lineáris jutalomfüggvényre tervezett LinUCB algoritmus hatékony megoldást nyújtott a változó és nagyméretű akcióhalmazok kezelésére, melyről Python szimuláció segítségével különböző eloszlású zajok esetén bizonyosodtunk meg.