

Posudek oponenta na bakalářskou práci Jurije Ružejnikova
„**Bayesovské odhadování pro adaptivní dynamické rozhodování**“

V předložené práci se autor zabývá problematikou tzv. *Real-time bidding (RTB)* tedy mechanismem pro prodej a nákup online reklamy. Konkrétně se pak autor zabývá řešením problému metodami z oboru učení se zpětnou vazbou, tedy *reinforcement learning*. Autor užívá formalismu Markovských rozhodovacích procesů pro formulaci úlohy a dynamického programování k její řešení.

Práce je středního rozsahu o čtyřiceti stranách. Je členěna do šesti kapitol, které se postupně věnují úvodu do problematiky, přehledu použité teorie, formulace rozhodovací úlohy, implementace navrženého algoritmu, experimentální ověření výsledků a závěr.

Ve stručném úvodu se autor snaží popsat fungování RTB, motivaci k řešení této úlohy a podat přehled o stávajících řešeních. Autor se dopouští nesprávného popisu role webového prohlížeče jako platformy pro RTB. Ve skutečnosti ve většině případů probíhá RTB zcela mimo webový prohlížeč. Postrádám také detailnější popis procesu RTB a porovnání s jinými přístupy, jako třeba statická aukce. Závažnější nedostatek je však zcela nedostatečný přehled dosavadních řešení a oboru ve kterém se autor angažuje. Velice stručně komentuje použití statických metod ze dvou zdrojů. Nezmiňuje však významnou práci autorů *Han Cai et. Al.* která taktéž užívá dynamické rozhodování v jazyku Markovských procesů a dynamického programování. Tento příspěvek se také nachází v doporučené literatuře v zadání práce.

První kapitola tvoří přibližně polovinu práce a autor v ní zavádí potřebné značení a vztahy. Autor zavádí základní pravděpodobnostní pojmy, od náhodné veličiny, Bayesovi věty a bayesovského učení až po Markovské procesy. Navzdory místy zmatečnému a nejednotnému značení je předvedený přehled informativní a adekvátně rigorózní. Kapitola o dynamickém programování je však příliš stručná. Postrádá diskusi o motivaci užití dynamického programování na řešenou úlohou.

Ve druhé a třetí kapitole autor formuluje úlohu RTB v jazyce dynamického rozhodování a navrhne implementaci algoritmu bayesovského učení s využitím dynamického programování s iterací hodnot. V začátku druhé kapitoly autor zcela milně konstatuje nedostupnost měřených dat z důvodu firemního tajemství. Ve zmíněné publikaci *Han Cai et. Al.* se nachází volně dostupný kód s předpřipravenými evaluačními procedurami na reálných datech společnosti *iPinYou*. Tyto data jsou v oboru považovány za standard, na kterém se RTB modely evaluují. Repositář také obsahuje implementaci stávajících algoritmů. Užití tohoto kódu by autoři jistě ušetřilo čas s generováním syntetických dat a také by poskytlo možnost srovnání se stávajícími řešeními na reálných datech. Autor ovšem úspěšně formuluje úlohu a navrhne algoritmus k jejímu řešení.

Ve čtvrté a poslední kapitole autor experimentálně ověřuje navrženou implementaci řešení úlohy. Jako jeden z hlavních nedostatků experimentální části vnímám nedostupnost kódu implementace. V návrhu implementace byl zvolen na první pohled nestandardní postup a postrádám možnost ověřit implementační detaily. Autor provádí tři experimenty s různou konfigurací. V prvním experimentu autor simuluje pět tisíc aukcí s rovnoměrným apriorním rozdělením. Ve druhém experimentu je prvních dvě stě kroků prováděných rovnoměrným náhodným výběrem za účelem prozkoumávání stavového prostoru. Ve třetím experimentu autor přidává šum z normálního rozdělení za účelem simulace lepší aproximace reálné situace.

V každém experimentu zvolil jako hlavní metriku akumulovanou odměnu agenta převedenou na české koruny. Autor také ukazuje okamžitou odměnu agenta pro malý počet iterací. Autor v těchto metrikách porovnává tři agenty: agenta který volí přiřazovanou částku rovnoměrně náhodně, agenta který má plnou znalost systému a třetího, navrženého agenta který odhaduje systém bayesovsky. Na první pohled může působit podezřele skutečnost, že akumulovaný zisk agenta který vybírá rovnoměrně náhodně roste téměř lineárně. Také je podezřelá závislost okamžité odměny, kdy pozorujeme výraznou nesymetrii ve znaménku odměn. Může to být způsobeno nesprávnou implementací, systematickou chybou v experimentu nebo nedorozuměním v interpretaci sledovaných metrik které autor formálně nezavádí.

V závěru autor konstatuje provedení simulace a budoucí vylepšení ve formě víceřadkové optimalizace, přesnější odhad modelu a evaluace na reálných datech. Chybí jakákoliv sumarizace dosažených výsledků a jejich diskuze. Postrádám diskusi o zjednodušení, které byly zavedeny za účelem simulace, diskusi o vlivu volby apriorního rozdělení a také diskusi o jiných hodnotách volených parametrů simulace.

Předložená práce působí silně nevyváženě. Jako největší problém vnímám absenci jakékoliv diskuse výsledků a závěr. Za další vážný problém pokládám nedostatečnou rešerš kdy autor nezohlednil kritický příspěvek z doporučené literatury, který nabízí řešení RTB pomocí dynamického programování a možnost evaluaci na reálných datech. Z formální stránky lze vytknout těžkopádnou dikci ve třetí a čtvrté kapitole a také příliš stručné popisky obrázků. Navzdory uvedeným výtkám se domnívám, že autor demonstroval schopnost pochopit matematickou formulaci rozhodovací teorie a také schopnost vytvořit výpočetní simulaci. Cíle zadání bakalářské práce byly: seznámit se s RTB, bayesovským odhadováním a dynamickým programováním, vytvořit a ověřit algoritmus na řešení RTB. Tyto cíle autor splnil a po komentáři k otázkám doporučuji k obhajobě a navrhuji ohodnotit stupněm D – (uspokojivě).

- 1) Proč experimenty neběželi po deklarovanou potřebnou dobu padesát tisíc iterací ?
- 2) Jaké je vysvětlení asymetrie ve znaménku okamžité odměny na Obr. 4.1 ?