

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Entropy maximization under entropic constraints
Jméno autora:	Anna Ibatullina
Typ práce:	bakalářská
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Department of Computer Science
Oponent práce:	Ing. Mgr. Jaroslav Hlinka, Ph.D.
Pracoviště oponenta práce:	Ústav Informatiky AV ČR, v.v.i.

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Kombinací požadavku na znalosti oborů a teoretické postupy i programovou implementaci se týká o náročnější zadání práce.	

Splnění zadání	splněno s menšími výhradami
<i>Posudte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.</i>	
Zadání bylo splněno. V principu mohlo být rozsáhlejší reportování druhé části bodu 3. (evaluace škálování implementace) a 4. validace na reálných datech, ale základní informace je v práci a online jsou numerické experimenty.	

Zvolený postup řešení	správný
<i>Posudte, zda student zvolil správný postup nebo metody řešení.</i>	
Zvolený postup řešení je správný, za úvahu stálo rozšíření o validaci samotného použitého odhadu entropie a jeho případného bias, případně rozšíření výsledků o další metody odhadu (nezáporné atomy, Ingletonovy nerovnosti), které jsou zmíněné jako možnost, avšak bez konkrétních výsledků výpočtů.	

Odborná úroveň	B - velmi dobře
<i>Posudte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Práce má velmi dobrou úroveň, využívá znalosti získané studiem a nad jeho rámec, zejména v oblasti teorie informace a optimalizace. Oceňuji aplikaci na vzorek reálných dat získaných od autorů originální publikace. Určitá nejistota a nepřesnosti prosvítají spíše v pokročilejších oblastech vybočujících z úzkého cíle práce, například interpretační tvrzení "Similarly, negative mutual information indicates redundant interactions, where knowing a subset of variables reduces the information gained from knowing the others.", případně "As we can conclude from the information diagram and example, the more independent the random variables, the bigger is the joint entropy value and the smaller is the mutual information value." (jak to vyplývá z informačního diagramu, nebo dokonce příkladu XOR, kde "mutual information" je dokonce záporná, což by striktně vzato mělo snad naznačovat vyšší nezávislost veličin než v produktové distribuci?)	

Formální a jazyková úroveň, rozsah práce	B - velmi dobře
<i>Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku.</i>	
Po formální a jazykové stránce je práce velmi dobře vypracována. Obsahuje jen velmi malé množství překlepů (strana 15: "refed" místo "referred", rovnice 2.12. chybějící znak "=", strana 22 "technics" místo "techniques", vysázení časové náročnosti pro "Setting constraints", používání členů v angličtině.) Obrázek 2.1. neobsahuje popisky os, popis, ani na něj není odkázáno z textu. Obrázek 2.5. je nejasný ohledně označení oblasti odpovídající hodnotám $I(X_1, X_2)$, resp. $I(X_1, X_3)$, $I(X_2, X_3)$. Na Obrázek 3.1 rovněž nevede odkaz a chybí popisky.	

Pozornému čtenáři obeznámenému s tematikou to ovšem neznemožňuje orientaci v tématu. Rozsahem je práce spíše kratší (tělo textu tvoří 24 číslovaných stran včetně ukázek kódu, seznamu referencí a seznamu online příloh), zásadním problémem to ale není, byť by pomohlo rozšíření širší zasazení do kontextu nebo extenzivnější popis numerických validačních experimentů. Rušivě působí například drobnosti, že definice teorie informace se v různých místech práce liší, nebo nekonzistentní způsob oddělování odstavců (nový řádek, nebo volný řádek mezi odstavci).

Výběr zdrojů, korektnost citací

B - velmi dobře

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posudte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Práce zahrnuje základní relevantní zdroje, které obecně správně cituje, poněkud chybí širší zasazení do kontextu - například citace na práce, které zavedly samotný pojem "Connected information" a detailnější motivace řešené úlohy včetně aplikačního významu. Hranice mezi novými a známými postupy je mírně nezřetelná, zejména v kapitole 3. Autorka fakticky prezentuje originální, elegantní a implementačně praktičtější variantu jiné podobné transformace problému na lineární programování prezentované v Martin et al. 2016 a Martin et al., 2017; v postupném výkladu problému ale není zřejmá hranice vlastního přínosu.

Podobně některá tvrzení, zejména v kapitole 2, jsou předkládána bez konkrétního důkazu či reference, například "It can be shown that ... monotonicity and submodular inequalities are enough to characterize Γ_n ", " Γ_n is polyhedral since it is defined by finitely-many linear inequalities." Přestože se může jednat o relativně známé/triviální výsledky, v bakalářské práci která složitější teorii už neobsahuje bych očekával detailnější vysvětlení už jen z důvodu prokázání seznámení se s problematikou.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Celkově práce obnáší solidní popis problému, návrh teoretického řešení a jeho přiměřenou softwarovou implementaci aplikovanou na testovací data ze simulací i vzorek reálných dat, a má tak potenciál rozšíření do formy časopiseckého článku.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Práci hodnotím napříč jejím aspektům jako velmi dobrou.

Navrhuji tyto otázky k zodpovězení při obhajobě:

1. What method was used to estimate the entropy from finite sample sizes in the case of real-world data? How could the results be affected by the selected approach, as compared with for instance the methods used in the original publication of Martin et al., 2017?
2. What exactly is meant by the statement "As we can conclude from the information diagram and example, the more independent the random variables, the bigger is the joint entropy value and the smaller is the mutual information value."?
3. Can you discuss in more detail the computational complexity of the algorithms, and to whether (if at all) could the precomputation time complexity be a limiting factor?

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **B - velmi dobře.**

Datum: 06/17/23

Podpis:

