



Posudek oponenta závěrečné práce

Oponent práce:	doc. Ing. Kamil Dedecius, Ph.D.
Student:	Bc. Anna Husieva
Název práce:	Detekce anomálií v časových řadách pomocí pravděpodobnostních modelů a metod strojového učení
Obor / specializace:	Znalostní inženýrství
Vytvořeno dne:	29. května 2024

Hodnotící kritéria

1. Splnění zadání

- ▶ [1] zadání splněno
- [2] zadání splněno s menšími výhradami
- [3] zadání splněno s většími výhradami
- [4] zadání nesplněno

Zadání diplomové práce považuji za splněné. Studentka sepsala stávající stav poznání v míře vyhovující zadání, implementovala statistické a ML modely pro modelování časových řad, navrhla jejich kombinaci a vše otestovala na vybraných datových sadách.

2. Písemná část práce

92 /100 (A)

Diplomová práce čítá celkem 87 stran, z toho 69 arabsky číslovaných. Sestává z jedné nečíslované kapitoly (Introduction) a pěti kapitol číslovaných. V těch je pojednáno postupně o základních pojmech v oboru čas. řad (kap. 1), statistických přístupech k modelování čas. řad a odhalování anomálií (kap. 2), ML a DL metodách pro modelování čas. řad a detekci anomálií (kap. 3), experimentech (kap. 4), závěrečná 5. kapitola shrnuje dosažené výsledky. Jednotlivé kapitoly jsou informačně vcelku pěkně vyvážené a dobře na sebe navazují. Autorka čerpala z několika odborných zdrojů, což je velmi chvályhodné, nicméně z toho plynou občasné drobné inkonzistence (např. požadavky na vlastnosti šumové veličiny v AR, MA a ARMA procesech).

Práce je napsána v anglickém jazyce o velmi dobré úrovni, gramatických či stylistických chyb je naprosté minimum. V sazbě mi občas chyběla interpunkce v matematice, typicky pokud šlo o více rovnic pod sebou, nebo rušilo násobení pomocí \times . Jinak je sazba pěkná. Grafická úprava práce je povedená, grafy jsou většinou velmi dobře srozumitelné a čitelné. Snad jen ve vyhodnocení experimentů by bylo vhodné obrázky trochu zvětšit a tabulky vysázet jinak, než vedle sebe. To je nicméně drobnost. Autorka cituje celkem 35 pramenů. Osobně bych raději některé nahradil primárními, typicky "Towards Data Science" či články v MDPI.

Celkově hodnotím písemnou část práce jako velmi povedenou.

3. Nepísemná část, přílohy 95 /100 (A)

Součástí práce jsou zdrojové kódy v pythonu, využívající běžné knihovny (sklearn, statsmodels...). Kód - překvapivě stručný - je přiměřeně okomentován. Implementace ensemble modelů mi není úplně jasná, nicméně jako programátorský laik se necítím dostatečně erudovaný toto posuzovat.

4. Hodnocení výsledků, jejich využitelnost 93 /100 (A)

Diplomová práce přináší nové a zajímavé výsledky v oblasti detekce anomálií v časových řadách. Studentka nasadila několik modelů a jejich kombinací na tři datové sady (Yahoo S5, Numenta a GutenTAG) a dosáhla určitých poznatků z hlediska jejich fungování. Celkově je nutno podotknout, že řešený problém je velmi komplikovaný, neboť je zpravidla těžké definovat, co je přesně anomální pozorování.

Celkové hodnocení 91 /100 (A)

Diplomovou práci považuji za povedenou, doporučuji ji k obhajobě a navrhuji hodnocení A.

Otázky k obhajobě

1. Experimenty vyhodnocujete pomocí F1 metriky. Přináší to s sebou nějaká rizika či nevýhody?
2. Mohly by využití transformace dat nějakým způsobem zamaskovat anomálie existující v původním datasetu?

Instrukce

Splnění zadání

Posudte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posudte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Písemná část práce

Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posudte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti.

Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posudte správnost používání formálních zápisů obsažených v práci. Posudte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 52/2021, článek 3.

Posudte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Nepísemná část, přílohy

Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů.

Hodnocení výsledků, jejich využitelnost

Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Celkové hodnocení

Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.