

Posudek vedoucí diplomové práce

Jméno a příjmení autora posudku: doc. RNDr. Iveta Mrázová, CSc.

Jméno a příjmení autora práce: Marek Behún

Název práce: Graph neural networks and their application to social network analysis

Vlastní text (sem prosím napište text posudku, délka textu posudku není omezena):

Diplomová práce Marka Behúna se zabývá problematikou grafových neuronových sítí a jejich využití při analýze sociálních sítí. Cílem práce tedy bylo popsat a navzájem porovnat relevantní modely grafových neuronových sítí. Dále měl diplomant zrekapitulovat známá paradigmatata a metody používané pro analýzu sociálních sítí. Věnovat se měl i technikám vhodným pro interpretaci a vizualizaci znalostí extrahovaných pomocí umělých neuronových sítí. Na některou z výše uvedených oblastí se měl student zaměřit podrobněji a měl navrhnout vhodnou strategii použitelnou pro analýzu dat ze sociálních sítí pomocí grafových neuronových sítí. Vybrané modely grafových neuronových sítí a přístupy použitelné pro jejich analýzu měl uchazeč implementovat. Součástí práce mělo být i zhodnocení vlastních výsledků a zkušeností.

V kapitole týkající se oblasti sociálních sítí se autor soustředil na míry centrality používané pro identifikaci vlivných jedinců a na problematiku komunit v sociálních sítích a jejich detekci (algoritmy Kerninghana a Lina, Girvanové a Newmanova a Lovaňský algoritmus). Těžiště práce pak spočívá v podrobném popisu nejpoužívanějších variant grafových neuronových sítí, tedy rekurentních grafových neuronových sítí (GNN* a GGNN), konvolučních grafových neuronových sítí (a to jak jejich spektrálních variant, které zahrnují spektrální CNN, ChebNet a GCN, tak také jejich prostorových modelů, které zahrnují MPNN, GraphSAGE a grafové sítě s modelem pozornosti GAT), grafových autoenkodérů (GAE* a VGAE) a časoprostorových grafových neuronových sítí (GCRN-LSTM a GCRN-GRU). Popisovanými vizualizačními technikami jsou metody t-SNE a UMAP.

Podstatná část práce je dále věnovaná popisu dat z portálu Tripadvisor, jejich scrapování, předzpracování a prvotní analýze pomocí technik navržených pro analýzu sociálních sítí. Protože scrapovaná data vedou na sociální síť s různými typy uzlů (hotely, resp. autoři recenzí), je pro řešení příslušných úloh použitý model tzv. heterogenní grafové neuronové sítě. Experimentální ověření vlastností navržených řešení autor představil na třech úlohách: predikce ohodnocení v recenzi hotelu (pomocí grafových autoenkodérů), odhad hvězdičkové kategorie hotelu (pomocí rekurentních grafových sítí) a vývoj významu/skóre hotelu ve zkoumané sociální síti (pomocí časoprostorových grafových neuronových sítí). K vyhodnocení výsledků student použil 10-násobnou křížovou validaci, resp. 5-násobné opakování experimentu v případě poslední řešené úlohy. Pro vizualizaci interní reprezentace znalostí vytvořené v naučené grafové síti student zvolil techniku UMAP.

Ke scrapování dat z platformy Tripadvisor, otestování vybraných algoritmů pro analýzu sociálních sítí, ověření vlastností porovnávaných modelů grafových neuronových sítí a vizualizaci extrahovaných znalostí pomocí techniky UMAP uchazeč použil vlastní software přiložený k práci na CD. Stručnou charakteristiku implementovaných utilit obsahuje Kapitola 10 předkládané práce. Implementované prostředí bylo navrženo pro spuštění z příkazového řádku. Za úvahu by jistě stál návrh a implementace uživatelsky příjemnějšího rozhraní, popřípadě jednoduché demoverze programu. Text sám je napsaný poměrně dobrou angličtinou, je přehledně členěný a má hezkou grafickou úpravu (výjimkou je např. reference [9] na str. 113). Základní ideje a výsledky pořízených měření v práci zohledňuje velký počet ilustrativních obrázků, grafů a tabulek. Seznam referencí čítající 60 položek dokládá dobrou orientaci autora v řešené problematice.

Předkládaná práce splňuje svůj původní cíl. Diplomant pronikl dostatečně hluboko do problematiky umělých neuronových sítí i analýzy sociálních sítí. Prokázal schopnost samostatně řešit i poměrně náročné úlohy spojené s analýzou reálných dat získaných z portálu Tripadvisor. Popsané a implementované přístupy uchazeč podrobně otestoval v rámci virtuální organizace české Národní Gridové Iniciativy MetaCentrum. Vlastní výsledky a zkušenosti dokázal objektivně vyhodnotit.

Doporučení k obhajobě:

Z výše uvedených důvodů práci doporučuji k obhajobě.

Vynikající práce vhodná pro soutěž studentských prací	NE [x]
-------------------------------------------------------	--------

Seznam soutěží studentských prací, viz <http://www.mff.cuni.cz/studium/bcmgr/prace/>

Pokud jste výše zaškrtnli ANO, zdůvodněte prosím svůj návrh, případně uveďte konkrétní soutěž, pro kterou je práce vhodná (rámeček lze nechat prázdný, pokud za dostatečné zdůvodnění považujete text posudku):

V Praze dne: 25. 1. 2023

Podpis:**

* *nehodící se škrtněte (vymažte)*

** *do SISu vkládejte formulář nepodepsaný (ve formátu PDF), podpis je potřeba doplnit až na vytištěný posudek.*