Posudek diplomové práce

Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce	Jan Babušík		
Název práce	Detekce anomalit v log datech		
Rok odevzdání	2021		
Studijní program	Informatika	**Studijní obor**	Umělá inteligence

Autor posudku	Mgr. Martin Pilát, Ph.D.	**Role**	oponent
Pracoviště	KTIML MFF UK		

Text posudku:

Práce se zabývá důležitým problémem vyhledávání anomalit v datech z logů. Cílem práce bylo najít model, který by byl schopen tyto anomálie detekovat automaticky a porovnat ho s jinými existujícími přístupy. Tento cíl se podařilo splnit, autor navrhl metodu pro zpracování logů a otestoval ji na několika datasetech.

Práce je rozdělena celkem do pěti kapitol (kromě úvodu a závěru). V první kapitole autor popisuje existující práce věnující se analýze logů, ve druhé kapitole jsou potom popsány základní principy hlubokých neuronových sítí a jejich trénování. Tyto kapitoly jsou napsané celkem dobře a nechybí v nich žádné informace podstatné pro zbytek práce.

Ve třetí kapitole jsou potom popsány datasety, které se používají ve zbytku práce. Jde o dva běžně používané datasety a jeden dataset vytvořený autorem práce ve spolupráci se společností HAVIT s.r.o. Popis datasetů je dostatečně podrobný a nechybí zde ani základní statistiky.

Čtvrtá kapitola potom obsahuje popis modelů dále testovaných v práci. Jsou popsány celkem 4 různé modely, od jednoduchých modelů založených na vyhledávání specifických zpráv, až po modely založené na LSTM sítích. Zde je škoda, že autor také nepřišel s nějakým vlastním, složitějším modelem. Většina zmíněných modelů jsou přímé aplikace zvolených metod bez vlastního přínosu autora. Také bych očekával srovnání například s modelem DeepLog zmíněným v úvodu práce.

V páté kapitole je uvedeno srovnání modelů popsaných ve čtvrté kapitole. Srovnání je provedeno na několika datasetech a výsledky jsou podrobně diskutovány z různých úhlů pohledu. Nechybí ani závěrečné srovnání všech modelů a doporučení pro použití modelů v budoucnu.

Celkově je práce dobře napsána, nechybí v ní žádné podstatné informace a srovnání modelů je také dobře provedeno. Za slabší stránku práce se potom dá považovat spíše menší vlastní přínos autora, kdy především aplikoval jednoduché existující metody na existující a nová data. Student nicméně ukázal, že je schopný kvalitní samostatné práce a práci doporučuji k obhajobě.

K práci mám jen několik otázek:

1. Proč se u LSTM modelu také používají okénka pevné délky? Tento model by měl být schopný pracovat i s posloupností libovolné délky na vstupu.
2. Budou výsledky práce někde reálně nasazeny?

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

V Praze dne 23. srpna 2021

Podpis: