

Posudek diplomové práce pana Bc. Michala Struny nazvané „Detekce a analýza exoplanet s využitím distribuovaných výpočtů a umělé inteligence“
Recenzent práce doc. Dr. Ing. Tomáš Brandejský

Diplomant na základě své znalosti programování, algoritmizace a umělé inteligence se ujal řešení velmi zajímavého problému vytvoření prostředí pro distribuované hledání exoplanet v dostupných astronomických datových archivech a jejich analýzu s využitím umělé inteligence (neuronových sítí). Práce využívá distribuované výpočty v podobě gridových výpočtů, kombinuje klasické Pythonovské aplikace, Jupyter notebooky a bash kód. Rozsah dat, která je třeba při hledání exoplanet prohledat se blíží k oblasti Big data. Diplomant vytvořil analytické prostředí pro toto hledání exoplanet. Diplomant projevil dobrou znalost programování, zpracování rozsáhlých externích datových archivů, projevil solidní znalost astronomie, která rozhodně není předmětem studia v našem oboru a ověřil je na vhodném a zdokumentovaném souboru příkladů.

Práce z formálního hlediska odpovídá doporučené šabloně a čítá 82 stran textu a 5 stran příloh. Práce je strukturována přehledně a logicky členěna do jednotlivých kapitol. Po stručném úvodu seznamujícím s vlastní strukturou práce následuje první kapitola věnovaná vysvětlení problému hledání exoplanet. Jsou zde vysvětleny jednotlivé přístupy (metody) hledání exoplanet i zhodnocena jejich dosavadní úspěšnost.

Druhá kapitola vysvětluje přístupy umělých neuronových sítí volené s ohledem na zamýšlené využití. Další kapitola pak s vlastním algoritmem přímo nesouvisí, ale popisuje další přístupy umělé inteligence. Chybí snad již jen inteligentní roboti a agenti.

Čtvrtá kapitola popisuje použité SW technologie. Pátá kapitola pak již popisuje návrh a vývoj aplikace včetně její struktury. Protože se jedná o gridové počítání, je logicky zvolena architektura klient-server s komunikací s uživatelem přes webovské rozhraní. Je zde navíc rozlišeno mezi administrací a přístupem běžných uživatelů zajišťujících vlastní zpracování dat.

K práci přiložený ZIP soubor obsahuje vedle vlastního textu práce i zdrojové kódy a dokumentaci experimentů. Na můj návrh nejsou součástí užití knihovny, které jsou volně dostupné na Internetu a jejich zahrnutí do předávané práce by jen vedlo na nesmyslné výsledky kontroly plagiátů. Na místo toho diplomant v příloze uvedl seznam těchto knihoven včetně identifikace jejich verzí. Kontrola se tedy soustředila na vlastní diplomantovu práci a zdrojový kód a nikde nedetekovala shodu nad 5%. Student pracoval samostatně, projevil značnou iniciativu a odvedl značný objem práce. Sám se seznámil s velkým množstvím nástrojů a knihoven.

Poslední, šestá kapitola nazvaná Rozvržení aplikace se nezabývá jejím grafickým rozvržením, jak by mohl implikovat její název, ale jejím popisem a popisem práce s ní.

Závěr přináší autorovo zhodnocení práce.

Seznam literatury je přiměřený a odkazuje na 29 prací. Přílohy přinášejí popis architektury neuronové sítě, databáze, dokumentaci REST API a statistické vyhodnocení verifikace výsledků.

Zadané cíle práce byly beze zbytku splněny. Po typografické stránce je práce zdařilá.

Kontrola plagiátů jako obvykle „nezklamala“ a kontrolovala vedle textu a zdrojových kódů i další doprovodné soubory, takže našla až 100% shodu v mnoha souborech json, README.md a pod. Přitom aplikace kontroly plagiátů je sama napsána hanebně, kdy nereaguje na dvojklik hlavní lišty okna, který má vést na jeho maximalizaci, nedovoluje skrolování pomocí křozorových šipek atd. Do budoucna je lepší předejít kontrole standardních souborů buď jejich zabalením do vnořeného archivu, nebo jejich vynecháním a nahrazením odkazem na jejich www stránky.

Vlastní pdf práce má shodu pod 5%. Zdrojové kódy nebyly kontrolou hodnoceny. Práci lze tedy jednoznačně považovat za původní.

Práci považuji za zdařilou, rozsahem textu velmi obsáhlou a zabývající se na diplomovou práci velmi zajímavou problémovou oblastí, kterou se z pohledu nároků na diplomovou práci podařilo diplomantovi zvládnout.

Proto navrhuji diplomovou práci Bc. Michala Struny doporučuji k obhajobě a hodnotím ji vzhledem k její neobvyklosti, rozsahu a náročnosti známkou A, tedy Výborně.

Připomínky

Obr. 1 – zobrazený trend je posunut vůdči většině údajů daty z let 2014 a 2016. Čím jsou způsobeny jejich extrémní hodnoty?

Obr. 2 – proč má osa x anglické popisky? Obdobný dotaz by bylo možno zformulovat i pro obr. 3, název kapitoly 1.1.1?

Tab. 1: “naleznutí” – není správný tvar slova.

Str. 35: “Hlavními oblastmi, ... , jsou např.:

Kap. 2.4: Při učení sítí bez učitele se ovšem využívají často jiné učící algoritmy, jako Hebbovo učení Hopfieldových ANN. - ale to jen připomínka, ne výhrada

Proč nebyl využit middleware BOINC, původně sice vyvinutý pro analýzu dat Ceti, ale posléze využívaný i v mnoha dalších gridových projektech, často z oblasti Astrofyziky, jako Einstein (gravitační vlny neuronových hvězd), Asteroids, MilkyWay atd.?

Jak je řešena spolupráce normálních Pythonovských a Bash aplikací s Jupyterovskými notebooky?

V Pardubicích 1.9.2021

doc. Dr. Ing. Tomáš Brandejský