

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	Využití velkých jazykových modelů v inženýringu proteinů
Jméno autora:	Matouš Soldát
Typ práce:	díplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Vedoucí práce:	Doc. Ing. Jiří Kléma, PhD.
Pracoviště vedoucího práce:	Katedra počítačů FEL ČVUT

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání	náročnější
<i>Hodnocení náročnosti zadání závěrečné práce.</i>	
Zadání předpokládá seznámit se hned s několika obory (velké jazykové modely, protein inženýring, aktivní učení). Současně vyžaduje schopnost analytického myšlení, integrace několika existujících přístupů a dobré technické a programovací schopnosti při zvládnutí jazykových modelů. Pokládám jej tedy spíše za náročnější.	

Splnění zadání	splněno
<i>Posuďte, 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 předpokládá vypracovat řešerši, navrhnout vlastní algoritmus pro rychlý návrh proteinu s danou funkcí, porovnat jej s existujícími algoritmy a srovnání odpovídajícím způsobem interpretovat. Všechny tyto cíle práce Matouše Soldáta plní beze zbytku.	

Aktivita a samostatnost při zpracování práce	A - výborně
<i>Posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven. Posuďte schopnost studenta samostatně tvůrčí práce.</i>	
Diplomant na zadání pracoval systematicky po dobu zhruba jednoho roku. Naše konzultace byly pravidelné, student na ně přicházel připraven a postup prací byl plynulý. V pozdějších fázích řešení byl student iniciativní a přicházel s vlastními nápady založenými na průběžném studiu literatury. Dle testů nejlepší navržené řešení založené na bayesovské optimalizaci v prostoru zanoření proteinů je výsledkem jednoho z těchto iniciativních návrhů.	

Odborná úroveň	A - výborně
<i>Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.</i>	
Student dle mého názoru nastudoval relevantní zdroje a využil je k návrhu tří vlastních algoritmů. Jeden z nich na základě testů doporučuje jako nejvhodnější. Tento algoritmus pojmenovaný BOES vychází z myšlenky využít současně předtrénované proteinové modely a bayesovskou optimalizaci s gaussovským procesem jako fitness modelem. Jde o agregaci dvou již použitých přístupů, výsledek je ale z pohledu maximalizace fitness proteinu při zachování fixního počtu evaluací lepší než u obou předchůdců. Současně diplomant demonstruje, že jeho algoritmus překonává i další základní benchmarky a jeden pokročilý referenční přístup pracující s předtrénovaným proteinovým modelem.	

Formální a jazyková úroveň, rozsah práce	A - výborně
<i>Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.</i>	
Práce je psána anglicky, dobře se čte a obsahuje minimum typografických chyb. Práce je dostatečně dlouhá na to, aby byly všechny základní kroky zdůvodněné a reprodukovatelné. Diplomant pracoval s rozumným časovým plánem a moje připomínky mohly být zapracovány do finální podoby textu.	

Výběr zdrojů, korektnost citací

A - výborně

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, 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 obsahuje 101 referencí. Mají standardní formát a vzhledem k zadání dobře pokrývají nutné zdroje. Práce jasně odděluje použitou literaturu a vlastní myšlenky, resp. přínos práce.

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.

Další komentář nemám..

III. CELKOVÉ HODNOCENÍ A NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení.

Využití velkých jazykových modelů v proteinovém inženýrství je moderní a perspektivní téma. Diplomant splnil všechny požadavky zadání. Dle mého názoru postupoval při svém řešení správně a podařilo se mu navrhnout algoritmus, který na dostupných proteinových datech prokazatelně překonává své předchůdce. Jeho práce má nepochybně publikační potenciál.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm A - výborně.

Datum: **10/06/2024**

Podpis: