

Hodnocení vedoucího bakalářské práce

Autor práce: **Karel PAUKNER**

Název práce: **Automatické rozdělení DNA do funkčních oblastí**

Jazyková a grafická úprava

Podprůměrné

Samostatnost zpracování tématu

Průměrné

Vhodnost použitých metod

Průměrné

Způsob zpracování a vyhodnocení

Podprůměrné

Správnost získaných výsledků

Průměrné

Vlastní přínos

Průměrné

Doplnění hodnocení, připomínky:

Práce se věnuje velmi aktuální tematice využití umělých neuronových sítí pro zpracování genetických dat. Cílem práce byla klasifikace sekvencí DNA do 3 funkčních tříd: promotor (sekvence obsahující regulační mechanismy), kódující sekvence (informace pro překlad DNA do proteinů) a offgene (definovaný zde jako cokoliv ostatního) a porovnání vlivu různých typů vstupních dat na trénování neuronové sítě. Student uvádí, jak vytvořil vlastní sadu datasetů pro trénování. Poté se v práci nachází vyobrazení výsledků trénování, a nakonec je představen nástroj sloužící k vizualizaci odhadů vybraného natrénovaného modelu pro libovolnou sekvenci DNA. Praktická práce studenta, kde prošel všemi stanovenými kroky a z hlediska trénování provedl většinu požadovaných experimentů, je však zastíněna velmi podprůměrným zpracováním.

V textu se nachází nespočet faktických, pravopisných a gramatických chyb, v úvodní kapitole „Důležité pojmy“ se popisované architektury neuronových sítí z hlediska obsahu postupně přesouvají mezi třemi úrovněmi, podsekce 1.2.1.4 na straně 8 neobsahuje vůbec žádný text a v sekci 1.7 se nachází proklikávatelné odkazy na tabulky mimo práci. Problémy přetrvávají i v představení a analýze natrénovaných modelů. Podoba celé sekce 2.3 „Výsledky“ je v podstatě 16x nakopírovaná stránka obsahující graf „accuracy“ přes epochy trénování a matici záměn. Matice záměn má nepřehlednou barevnou škálu a hodnoty udávající škálu odpovídají nejspíše četnostem, zatímco v jednotlivých polích jsou pak zobrazeny pravděpodobnosti. Graf „accuracy“ odhaluje chyby, kdy například na straně 34 student uvádí dosaženou validační přesnost 37%, která odpovídá fázi se zřetelným overfittingem (problém se v práci opakuje). Analýza modelů je prováděna jednou větou bez jakéhokoliv rozboru. Srovnání modelů z jedné sady experimentů neobsahuje rovněž žádnou přidanou informaci nebo diskusi o relevanci výsledků vzhledem k testovanému parametru.

Výsledná práce doplácí na nedostatek času, které jí bylo v průběhu akademického roku věnováno. Rozdíl 1-2 týdnů mohl zásadně změnit podobu celé práce a tím výrazně i hodnocení.

Dotazy

Na výstupu z transformeru se v práci nachází average pooling. Jaké jsou další možnosti, jak z transformerů přejít ke klasifikační vrstvě?

Splnění bodů zadání

úplně

Doporučení k obhajobě

ANO

Hodnocení: 3 - Dobře

V _____ dne _____

Ing. Lukáš Kuhajda