

Posudek oponenta bakalářské práce

Autor práce: **Jaroslav Lehačka**

Název práce: **Aplikace metod pro generování EEG dat**

Obsah práce

Autor v bakalářské práci popisuje problematiku zpracování EEG signálu, evokovaných potenciálů a metody pro rozšiřování datové sady. Hlavním cílem bakalářské práce je dogenerování augmentovaných dat pro zvětšení robustnosti klasifikačních algoritmů. Autor využil různých metod (přidání šumu, časový posun, kombinace šumu a posunu, GAN sítě). Autor v práci popisuje metriky měření úspěšnosti samostatné (Accuracy, Precision) a kombinované (Confusion matrix, Precision - Recall, F - skóre). V další kapitole je popsán experiment (charakteristika vstupních dat, obsažené stimuly, velikost a formáty dat, vyvážený a nevyvážený datový soubor) a analýza dat. Dále autor popisuje technologie využívané pro strojové učení. Na základě zvolených technologií (Jupyter Notebook, Google Colab, Python, MNE, DAGsHub.com, MLFlow, DVC, Keras) provádí návrh řešení a realizuje rozšíření datového souboru. Nakonec výsledné řešení otestuje a provede zhodnocení výsledků.

Kvalita řešení a dosažených výsledků

Autor se dostatečně seznámil s problematikou zpracování EEG signálu a evokovaných potenciálů. Dále prozkoumal metody pro rozšiřování datové sady. Autor úspěšně vybral metody pro rozšíření datové sady (posun, šum, kombinace posun a šum, GAN sítě) a vytvořil další datové soubory rozšiřující původní datový soubor. Autor tyto data použil v klasifikační úloze, kde použil klasifikátor studenta Romana Kalivody. Při implementaci vlastního řešení v Jazyce Python autor pracuje v prostředí Jupyter Notebook. Při porovnání původních a vytvořených dat využil vybraných metrik (accuracy, precision). Nejlepší výsledky vycházely autorovi s daty vygenerovanými v rámci GAN sítě. Zlepšení při této metodě odpovídalo výsledkům referenčního článku (3,1%). Úroveň kvality zdrojových kódů je dobrá. Zdrojové kódy jsou okomentovány. Dále autor vytvořil jednotkový test pro ověření GAN sítě, provedl manuální testování komponent a správnost fungování metod pro rozšíření datasetu ověřil jednotkovými testy. Autorem mi bylo předvedeno jeho řešení a nemám žádné výhrady.

Formální úroveň

Po formální stránce je kvalita práce vyhovující. Práce je logicky strukturovaná. Počet překlepů je v toleranci vzhledem k rozsahu práce (cca 37 normostran). Dále je v bakalářské práci 32 zdrojů. Tištěná příloha obsahuje model generátoru, model diskriminátoru, uživatelskou příručku a obsah elektronické přílohy. Elektronická příloha obsahuje text bakalářské práce, Jupyter Notebooky pro generování dat (kombinované kanály, GAN sítě, s přidaným šumem, s časovým posunem), testovací utility pro ověření správnosti řešení, použitý klasifikátor Romana Kalivody, zdrojové kódy a návod na instalaci.

Práce s literaturou

Seznam literatury obsahuje 32 zdrojů (odborné články, knihy, studijní materiály, tutoriály, dokumentaci knihovny MNE). Uvedené zdroje jsou vhodně zvolené vzhledem k tématu práce.

Splnění zadání

Všechny body zadání byly splněny.

Dotazy k práci

Na začátku jste měl datasety nevyvážené a Vaším úkolem bylo dogenerování augmentovaných dat pro větší robustnost klasifikačních algoritmů. Co to znamená vygenerování 100% dat v kontextu vyvážení datasetů?

Jak si stojí Vaše dosažené výsledky oproti výsledkům jiných autorů?

Navrhuji hodnocení známkou **výborně** a práci doporučuji k obhajobě.

V Plzni 16.08.2022

Ing. Petr Brůha