

Magyar nyelvű Szentiment Analízis Projekt Terv

Név

2025. április 19.

1. Projekt Áttekintés

Ez a projekt célja egy magyar nyelvű szentiment analízis modell fejlesztése Pythonban, amely a HuSST adatkészletet használja. A modell felé elvárás, hogy képes legyen szövegeket negatív, semleges és pozitív kategóriákba sorolni.

2. Módszertan

A cél megvalósításához a huBERT betanított neurális hálót fogom felhasználni alapmodellként. Az előre betanított neurális háló nagyon jó kiindulási alapként szolgál, mivel magyar nyelvű adatokon tanították tehát általános magyar nyelvtudással rendelkezik. Képes a szövegek értelmezésére és feldolgozására, viszont általánosságban elmondható, hogy ezeket az alapmodelleket további tanítással kell kiegészíteni ha specifikusan egy bizonyos célra szeretnénk használni a tudását.

Jelen feladatban a HuSST adathalmazzal fogok további tanítást végezni a modellen. A HuSST mint korábban említésre került, magyar nyelvű kijelentéseket tartalmaz és az azokhoz tartozó címkét. A címke lehet negatív, semleges, vagy pozitív. Ezek alapján kerül besorolásra az adott szöveg.

3. Dataset

A bevezetőben ismertetett két forrást fogom használni a projekt megvalósításához.

- huBERT base model (Hungarian Universal Bidirectional Encoder Representations from Transformers)
- HuSST dataset (Hungarian Stanford Sentiment Treebank)

4. Implementáció

A modell Pythonban készül a következő könyvtárakkal:

- torch
- sklearn
- transformers
- datasets
- tqdm

5. Források

A dokumentumot az alább felsorolt források segítségével készítettem el.

Hivatkozások

- [1] SZTAKI-HLT. (2022). *hubert-base-cc*. Hugging Face.
<https://huggingface.co/SZTAKI-HLT/hubert-base-cc>
- [2] NYTK. (2022). *HuSST Dataset*. Hugging Face.
<https://huggingface.co/datasets/NYTK/HuSST>
- [3] SZTAKI-HLT. (2022). *huBERT - Hungarian BERT Model*. BME-HLT.
<https://hlt.bme.hu/hu/resources/hubert>

- [4] Orosz György. (2023). *Awesome Hungarian NLP Resources*. GitBook.
<https://oroszgy.gitbook.io/awesome-hungarian-nlp-resources>
- [5] Orosz György. (2023). *Awesome Hungarian NLP*. GitHub.
<https://github.com/oroszgy/awesome-hungarian-nlp>
- [6] Laki László J., Yang Zijian Győző. (2022). *huBERT - Hungarian BERT*.
Acta Universitatis Óbuda.
https://acta.uni-obuda.hu/Laki_Yang_134.pdf