

Druga seminarska naloga

Jernej Vivod
Komunikacija človek računalnik

10. januar 2021

1 Uvod

Za drugo seminarsko nalogo pri predmetu Komunikacija človek računalnik smo implementirali vmesnik možgani-računalnik, ki z uporabo klasifikatorjev strojnega učenja in metod ekstrakcij značilk skuša klasificirati zamišljene aktivnosti na podlagi podanih posnetkov EEG. Vmesnik smo implementirali kot pregledno in modularno zasnovano programsko opremo napisano v jeziku Python. Rezultate klasifikacij predstavimo z uporabo standardnih metrik za klasifikacijsko uspešnost kot tudi z matriko zmot. Značilke pridobimo tako z uporabo metode skupnih prostorskih vzorcev (common spatial patterns), kot tudi z izračunom lastnosti signala v časovni in frekvenčni domeni.

2 Metode

Vmesnik smo implementirali v obliki programske opreme oziroma paketa za programski jezik Python. Paket je zgrajen modularno. Modul *data_utils* je namenjen delu s podatki in z segmentacijo signalov na enakomerne intervale, modul *feature_extraction* je namenjen funkcionalnosti ekstrakcij značilk, modul *results_utils* pa je namenjen predstavitvi rezultatov.

Primer zagona evalvacije v terminalu z uporabo klasifikacijske metode naključnih gozdov, uporabo frekvenčnih značilk, 10% prekrivanjem med intervali in intervali dolžine 4 sekund je prikazan spodaj:

```
python3 main.py --method rf --use-spec --overlap 0.1 --interval-len 4.0
```

Posnetke signalov smo najprej segmentirali na enakomerne intervale, katerih dolžino v sekundah in prekrivanje lahko določi uporabnik. Oznako takšnih intervalov smo določili kot oznako, kateri v izvornem signalu pripada največ vzorcev znotraj vsakega intervala. Večje prekrivanje med intervali pomeni večje število primerov, vendar moramo biti pozorni na možnost, da bosta učni in testni množici v primeru velikega prekrivanja posledično lahko vsebovali zelo podobne primere, kar lahko močno dvigne klasifikacijsko uspešnost.

Za ekstrakcijo značilk smo uporabili tako metodo skupnih prostorskih vzorcev, kot tudi množico drugih značilk signala v časovni in frekvenčni domeni, ki jih izluščimo z uporabo skripte *engineer_features.m*, ki se nahaja v direktoriju *eeg_classification* znotraj korenskega direktorija.

3 Rezultati in diskusija

Evalvacijo izvedemo z razbitjem množice primerov na učno in testno množico ter z uporabo specificiranega klasifikatorja (uporabimo bodisi LDA, naključne gozdove ali množico podpornih vektorjev). Rezultati se izpišejo v obliki klasičnih performančnih metrik v datoteko *classification_reports.txt* znotraj direktorija *eeg_classification/results*. Izrisi matrike zmot pa se shranijo v direktorij *./eeg_classification/results/confusion_matrices/*. Značilke dobljene z uporabo skupnih prostorskih vzorcev zamišljenih aktivnosti v večini ne ločujejo dobro, kar lahko vidimo z vizualizacijo parov ali pa z vizualizacijo komponent dobljenih z LDA. Problematična je tudi relativno majhno število učnih primerov, ki jih lahko dobimo iz posameznega posnetka, saj lahko pri uporabi velikega števila značilk, ki je v slabem razmerju z številom učnih primerov, pride do težav povezanih s t.i. prekletstvom dimenzionalnosti. Razvite metode predstavljajo izhodišče za primerjavo kompleksnejših metod, ki smo zmožne tudi povsem preskočiti fazo ekstrakcij značilk in jo vključiti v samo fazo učenja.