

3. félévi beszámoló

Borkovits Bendegúz (borbende@phys-gs.elte.hu)

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája

PhD program

Témavezető: Pollner Péter

A dolgozat címe: Képdiaosztika adattudományi alapú fejlesztése

2025.01.13.

1. Bevezetés

Folytattam a Budapesti Műszaki és Gazdaságtudományi Egyetemen (BME), az Alkalmazott Biotechnológia és Élelmiszertudományi Tanszék laborjában közép-infravörös tartományban a spektrumok felvételét. A spektrumokat felügyelt tanításnak és multiple instance learningnek vetettük alá. Ugyanakkor, egy eoizinofil sejt és szöveti régió detekciós projekten is dolgoztam a patológusok közreműködésével. Emellett betekintést nyertem a Komondor szuperszámítógép infrastruktúrájába.

2. Az aktuális félévben elvégzett kutatások ismertetése

A vastagbélrákos szövetekről felvett közép-infravörös (MIR) spektrumokból álló adatbázisom kétharmada áll már rendelkezésre. Ez Kilim segítségével felhasználtunk egy újabb adatfeldolgozó módszert (multiple-instance learning-et, avagy gyengén felügyelt tanítást), amely ezúttal nem magukhoz a spektrumokhoz, hanem az általuk meghatározott szöveti területhez rendelt hozzá egy globális címkét (normál szövet, primer carcinoma, májrák) és a korongokat klasszifikálta. Itt egy listát kell elképzelni, amiben a lista egy eleme egy adott koronghoz tartozó több ezer spektrumból álló táblázat egyetlen címke alatt. A módszert összehasonlítottuk az immár kibővített adatsoron lefuttatott felügyelt tanítási módszereinkkel és azt találtuk, hogy teljesítményben felülmúlja azokat. Ugyanakkor, ez az új módszer arra is képes, hogy ún. attention map-eket készítsen a prediktált korongokhoz, ezzel jelölve a korong hő térképén, hogy melyik belső régió spektruma alapján hozta meg prediktálási döntését és mekkora súllyal. Ez hasznos lehet a korongok belső régióinak feltérképezése, illetve a hangsúlyos adatok kiszűrése miatt.

Patológusokkal folytatott, együttműködésünk során eozinofil sejteket kellett detektálnunk gyermekgyógyászati szövetmintákról készített Whole Slide Image-eken (WSI). A projektet három nagyobb részre osztottuk: eozinofil sejtek detekciója, epithelium szöveti régiók szegmentálása, és szöveti eozinofil sűrűség kiszámítása. A képi és annotációs adatok feldolgozása és megfelelő formátumra konvertálása után a detekciós részhez elkészítettem egy objektum detekciós framework-öt, ami vegyes eredményeket produkált. Habár szépen el tudta különíteni az eozinofil sejteket a háttértől és sok más sejtípustól, egy vörös hámsejtet ugyanúgy magas konfidenciával jelölt eozinofilnak a modell, emiatt erre a feladatrészre később vissza kell térni. Ezután a szegmentálási feladat már sokkal jobb eredményeket hozott. A kereszt-validálás során egy független teszt adathalmazon végzett prediktálások eredménye stabil volt és a pontossági metrikák is biztató eredményeket mutattak. Jelenleg a projekt harmadik feladata előtt állunk, amit remélhetőleg a következő év elején be tudunk fejezni, hogy a projekt publikálásra kerüljön.

A múlt félévben elkezdett tömegspektrometriai projektünk jelenleg adatbeszerzési fázisban van.

3. Publikációk

Nem publikáltam ebben a félévben.

4. Tanulmányi tevékenység az aktuális félévben

A következő kurzusokat végeztem el ebben a félévben:

- Klaszterezés hálózatokkal, (FIZ/3/064E)
- Evolúciós játékelmélet. (FIZ/3/059E)

5. Konferenciák az aktuális félévben

Nem vettem részt konferencián ebben a félévben, mint előadó. A Komondor szuperszámítógép kezelését segítő két online workshopon és a Kifü Óbudán tartott előadás-sorozati napján voltam, mint a közönség egy tagja.

6. Oktatási tevékenység az aktuális félévben

- Valószínűségszámítás és statisztika a fizikában (valszamf19va, gyakorlat, 6 kredit) gyakorlattartás (5-ször 2 óra) és ellenőrző zárthelyi dolgozat felügyelése (2 óra) és javítása (3 óra).

7. Elismerések

Továbbra is részesülök a DKÖP-2023 Doktori Kiválósági Ösztöndíj Program által nyújtott támogatásban.