

1. félévi beszámoló

**Pataki Bálint Ármin** ([patbaa@gmail.com](mailto:patbaa@gmail.com))

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája PhD program

Témavezető: Csabai István

A dolgozat címe: Gépi tanulás a tudományokban

2019.01.15.

### **Bevezetés:**

2012-ben a vizuális, kép alapú problémák számítógépes, algoritmikus megoldása mérföldkőhöz érkezett, ugyanis ebben az évben először nyerték deep learning modellel a legismertebb képfelismeréssel kapcsolatos versenyt, az ImageNet-et. Ettől kezdve a deep learning számos más területen is messze felülmúlta a korábbi módszerekkel elért eredményeket, úgy, mint beszédfelismerésben, táblajátékokban, arcfelismerésben vagy például az orvosi diagnosztikában.

A gépi tanulás, mesterséges intelligencia napjainkban egyre szélesebb körben kezd elterjedni. Ezen módszerek alapfelvetése, hogy csak a predikciós pontosság számít. Nem cél a problémakör megértése (pl.: hogy arcfelismerésnél melyek a legfontosabb szempontok, ami alapján azt lehet mondani, hogy két képen ugyanaz az arc található-e), kizárólag az számít, hogy a módszer minél pontosabban működjön. Sok esetben ezek a modellek annyira összetettek, hogy kevés esély mutatkozik a részletes megértésükre, hisz több millió (több tízmillió vagy akár százmillió) paraméterrel is rendelkezhetnek. A tudomány azonban hagyományosan arra alapoz, hogy megértsük a problémák gyökerét, és ezen megértésre alapozva prediktáljuk a jövőbeli kísérletek kimenetét lehetőség szerint minél egyszerűbb modellel.

Az utóbbi évtizedekben számos tudományág kezd átalakulni adat intenzív tudománnyá, ahol rengeteg mérési adat keletkezik, melyeket hagyományos eszközökkel sokszor reménytelen feldolgozni.

A PhD évei során szeretnék minél inkább elmélyülni a különböző gépi tanulási algoritmusok rejtelmében, és ezen algoritmusok tudományokban való alkalmazásainak lehetőségében.

### **Az aktuális félévben elvégzett kutatások ismertetése:**

A félév során főleg biológiai és kozmológiai adatokon dolgoztam, ezen adatokra alapozva igyekeztem gépi tanulási algoritmusokat fejleszteni, tanítani.

Egyik kutatás gravitációs gyenge lencsézéses szimulációkból származó képek alapján a szimuláció során beállított kozmológiai paraméterek ( $\Omega_m$  és  $\sigma_8$ ) meghatározása volt. Ebből a kutatásból cikk is született a félév során, melyről részletesebben alább írok.

A másik kutatás, amin dolgoztam biológiai eredetű adatokat használ. Baktériumok DNS-e alapján szeretnék megmondani, hogy egy adott antibiotikumra (ciprofloxacín) milyen mértékben ellenállóak. Nagyságrendileg 500 eColi baktérium teljes genom szénvenálása áll rendelkezésünkre, melyekre lemérték az említett antibiotikumra való ellenállóképességét is, így egy supervised gépi tanulási problémával állunk szemben. Hogyha a DNS alapján meg tudnánk mondani a pontos ellenállóképességet, akkor a DNS szekvenálási módszerek fejlődését figyelembe véve egy, a jövőben jóval gyorsabb módszerre tennénk szert az eddig használthoz képest. A kutatáshoz kapcsolódó kézirat írása folyamatban van.

**Publikációk:**

2018.10.29-én megjelent a Nature Astronomy folyóiratban egy cikk „An improved cosmological parameter inference scheme motivated by deep learning” címmel, melyen társszerző vagyok Ribli Dezső és Csabai István mellett. A cikk fő témája kozmológiai paraméterek becslése gravitációs gyenge lencsézés hatásán keresztül. A Columbia Lensing Team gyenge lencsézéses szimulációkat készített, mely szimulációk különböző kozmológiai paraméterekkel lettek elkészítve. Ha a szimulációk módszerében kellően közel állnak a valósághoz, akkor egy olyan modell által, ami a szimulált képek alapján megmondja, milyen kozmológiai paraméterekkel készült a szimuláció, becslést mondhatunk valódi észlelésből származó képek alapján a valódi kozmológiai paraméterekre. Hiszen a modell számára a valós mérés és a szimuláció között nincs különbség, ha módszerében valóban helyes a szimuláció.

A cikkben bemutattuk, hogy konvolúciós neurális hálókkal jobb becslés adható a szimulált képekre, mint a hagyományos módszerekkel. Arra a meglepő eredményre jutottunk, hogy a konvolúciós neurális hálóból merített ötlet alapján egy egyszerűbb modellel is hasonló eredmény érhető el.

**Tanulmányi tevékenység az aktuális félévben:**

A félév során négy kurzust vettem fel, melyek közül hármat végzek el. Az elvégzett kurzusok:

- Modern képkalkulációs technikák a biológiában FIZ/3/077E
- Klaszterezés hálózatokkal FIZ/3/064E
- Haladó informatika a csillagászatban I. EA FIZ/2/031E

A tárgyak teljesítése a beszámoló írásának időpontjában folyamatban vannak.

A Haladó informatika a csillagászatban kurzus a Rézecskefizika és Csillagászat PhD program része, de a doktori szabályzatnak megfelelően a doktori programom (Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája) vezetőjétől, Kürti Jenőtől írásos igazolást kaptam, hogy elfogadható a tárgy.

Ezen felül a fast.ai online kurzusát végzem, mely gépi tanulásról, deep learningről illetve az általuk fejlesztett, PyTorch-ra épülő ezen algoritmusokat implementáló Python csomagról szól.

**Konferenciák az aktuális félévben:**

A félév során egy projektmeetingen vettem részt Lyonban november 7-én. Ezen a meetingen angol nyelven prezentáltam a munkám a kollaboráció többi tagjának az antibiotikum-rezisztencia prediktálásával kapcsolatosan. Egyéb konferencián nem vettem részt.

**Szabadalmak:**

A fentebb említett cikk megjelenését követő sajtóközlemény hatására megkeresett az ELTE Innovációs Központja, hogy a sajtóközleményben említett Dream Challenge versenyen elért, Parkinson-kór mobilszenzor alapú detektálását elősegítő munkám szellemi tulajdon kezelésével, védelmével kapcsolatosan konzultáljunk. A szabadalmi eljárás/bejelentés elindult, a kimenetele egyelőre megjósolhatatlan számomra.

**Oktatási tevékenység az aktuális félévben:**

A félév során az alábbi három tárggyal kapcsolatosan végeztem oktatási tevékenységet:

- Számítógépes alapismeretek számalapf18la 4 kredit. Gyakorlat tartása heti szinten
- Adatbányászat és gépi tanulás dsminingf17vm 4 kredit. Néhány előadás tartása, valamint gyakorlatokon való besegítés heti szinten
- Deep learning és gépi tanulás a tudományokban 2. deplea2f18em 4 kredit. Konzultációk tartása kb havi szinten