

3. félévi beszámoló

Báskay János (baskayj@student.elte.hu)

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája PhD program

Témavezető: Pollner Péter

A dolgozat címe: Hálózat-statisztikák adattudományi alkalmazásai

Bevezetés

A 21. század információs forradalma megteremtette a lehetőséget, hogy a természeti és társadalmi jelenségek tanulmányozása során korábban elképzelhetetlennek tűnő adatmennyiséget gyűjthessünk össze. Hamar világossá vált, hogy ezeket manuálisan feldolgozni, majd előrejelzéseket alkotni lehetetlen feladat, erre a problémára nyújtott megoldást a különböző gépi tanulási módszerek megjelenése. Ezek segítségével az adatok közötti korrelációk könnyedén összegyűrhetőek egy robosztus modellé, mellyel klasszifikációs vagy regressziós problémák oldhatók meg.

Ezzel párhuzamosan gyors fejlődésen ment keresztül a Hálózattudomány is, melynek keretein belül rendelkezésre áll számos eszköz a komplex rendszerek megértésére és modellezésére.

PhD tanulmányaim során szeretnék a Hálózattudomány eszközeinek felhasználásával és gépi tanulási módszerek interdiszciplináris alkalmazásán keresztül átfogó ismereteket elsajátítani az Adattudomány területén.

Aktuális félévben elvégzett kutatások ismertetése

Túlélés jóslás gépi tanulással

Ebben a félévben igyekeztem a meglévő kód technikai adósságát csökkenteni, a megfelelő részek újraírásával, illetve a korábbi fa alapú, illetve logisztikus módszerek helyett egy neurális háló fogja jóslni a betegek túlélési görbáját. Ennek a hálónak az adatok mellett szüksége van a különböző tünetek közötti kauzális kapcsolatra, amelyet egy Bayesi hiedelemhálózattal lehet leírni. Ennek az előállítása hagyományos statisztikai módszerekkel elképesztően időigényes: Exponenciálisan skálázódik a csúcspontok számának függvényében. Emiatt megvizsgáltam annak a lehetőségét, hogy egy variációs autoenkóder (VAE) segítségével állítsam elő a szomszédossági mátrixot. Yu et. al „*DAG-GNN: DAG Structure Learning with Graph Neural Networks*”¹ publikációját használtam fel és implementáltam tensorflow-ban a megoldásukat, mivel az általuk publikált pytorch implementáció nem volt jól átlátható. Sajnos ez a megoldás nem vezetett stabil szomszédossági hálókhoz, ezért a hiedelemhálózatot végül a bnlearn python csomag segítségével állítom elő a szükséges hiedelemhálót.

¹ YU, Yue, et al. Dag-gnn: Dag structure learning with graph neural networks. In: *International Conference on Machine Learning*. PMLR, 2019. p. 7154-7163.

- Preklinikai modellek a daganatkutatásban (FIZ/3/082)

Oktatási tevékenység az aktuális félévben

Az alapszakos Fizika hallgatók számára tartott *Valószínűségszámítás és statisztika a fizikában* tárgy kiadott gyakorló feladatainak egy részét javítottam, illetve a mesterszakos Fizikus hallgatók számára tartott *Adatmodellek és adatbázisok a tudományban* nevű előadáson két hálózat-analízissel kapcsolatos témát adtam ki a hallgatóknak, melyek megoldásában velem konzultálhattak.