

1. félévi beszámoló

D’Intino Eugenio (ead518@student.elte.hu)

Csillagászat és Űrfizika PhD program

Témavezető: Szabó Róbert, Bódi Attila

A dolgozat címe: Gépi tanulási módszerek alkalmazása
csillagászati mérések kiértékelésére

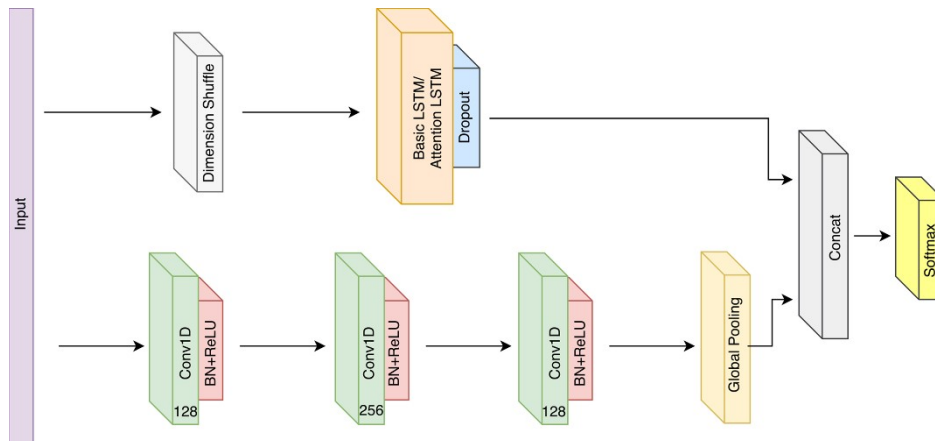
Bevezetés

Az időbeli változást mutató asztrofizikai objektumok tömeges, gyors és hatékony osztályozása elengedhetetlen mind a későbbi észlelések megtervezéséhez, mind a csillagok analíziséhez. Az új generációs égboltfelmérő programoknak köszönhetően az elérhető adatok mennyisége napról napról növekszik, amelyek feldolgozása automatizált szoftverek fejlesztését igényli. A jelenleg is futó Zwicky Transient Facility (ZTF), vagy a Vera C. Rubin Observatórium 2023-ban induló Legacy Survey of Space and Time (LSST) égboltfelmérő programok nem csak a már jól ismert "klasszikus" változócsillagok számát fogják növelni, hanem várhatóan számos új, eddig meg nem figyelt jelenség felfedezését is lehetővé fogják tenni. A különböző objektumok ismert kategóriákba való besorolására ma már számos klasszifikációs módszer elérhető, azonban azon változócsillagok felfedezése, amelyek eddig nem ismert változékonyságot mutatnak, még most is a szakértők által végzett manuális vizsgálatokon alapszanak, felfedezésük pedig esetleges, holott sokszor éppen a legnagyobb újdonság faktorral és tudományos értékkel bírnak. Célunk különböző gépi tanulási módszerek felhasználása ezen objektumok felfedezésének elősegítésére.

Az aktuális félévben elvégzett kutatások ismertetése

Kezdeti adathalmaznak a Kepler űrtávcső (Borucki és mtsai, 2010) idősorait használtuk, mivel az itt található változócsillagokat széles körben vizsgálták már. Ezeket a tanulmányokat felhasználva egy katalógust hoztunk létre, amely a már vizsgált objektumok azonosítószámához (KIC ID) hozzárendelt osztályokat tartalmazza.

A Kepler adatok fluxus idősorait különböző negyedekből úgy választottuk ki, hogy egységes időhosszúságúak legyenek, a lehető legkevesebb hiányzó adatponttal. Ahol csak egyetlen adatpont hiányzott az idősorból, de a szomszédos adatok megvoltak egyszerű lineáris extrapolációval pótoltuk azokat. Ennél nagyobb hiányzó idősávokat eldobtunk minden idősorból. A klasszifikációhoz az LSTM-FCN (F. Karim, 2018) architektúrát választottuk. A



ábra 1. LSTM-FCN architektúra

háló két ágból áll össze, amelyek eredményei össze vannak fűzve. A két ág egy konvolúcióból és egy LSTM-ből áll. Továbbá ahogyan az 1. ábra is szemlélteti egy opcionális figyelem réteget is tesztelték az LSTM ágon, amit mi nem használtuk.

Az adathalmazban a következő változócsillag osztályok szerepeltek: forgási változók, vörös óriások, fedési kettősök, exobolygók, delta Scutik, M óriások. Ezeket a modell minimális adattisztítás után az alábbi teljesítménnyel sorolta be az adott osztályokba:

- Pontosság: 91.17%
- Kiegyenlített pontosság: 91.14%
- F1 macro: 85.53%
- F1 micro: 91.17%

A klasszifikációs hálót arra használjuk, hogy kivegyük az osztályozáshoz optimalizált látens teret és egy izolációs erdő algoritmust alkalmaztunk rá. Ez a modell megmondja melyek azok a látens terek, amelyek a legjobban eltérnek a legtöbb látott mintától – anomália érték. Az anomália értékeket kompakt, RR Lyrae, nap-típusú és a Tabby's csillagon értékeltük ki. A nap-típusúak, amelyek hasonlóan viselkednek a látott osztályokhoz, azokat nem soroltuk be anomáliáknak, míg a kompaktoknál egy 33%-os hibaráttát értünk el. Az RR Lyrae-et minden esetben anomáliának találtunk, ami biztató, mivel ezek szignifikánsan eltérnek a látott mintáktól. A Tabby's csillagot is megtaláltuk, mint anomália, amennyiben az adatsor jól ismert anomáliát mutató része is megtalálható volt az elemezett mintában.

Mindezt úgy értük el, hogy minimális előkészítést csináltunk az adatokon. Kiszórtuk a szórás háromszorosánál jobban kiugró egyedi pontokat és standardizáltuk a fénygörbéket.

Tanulmányi tevékenység az aktuális félévben

- A gépi tanulás új eredményei szeminárium (FIZ/3/092)
- Adatmodellek és adatbázisok a tudományban (FIZ/3/086)

Hivatkozások

F. Karim és mtsai, 2018, "LSTM Fully Convolutional Networks for Time Series Classification", *IEEE Access*, 6, 1662-1669.

Borucki és mtsai, 2010, "Detection Mission: Introduction and First Results", *Science*, 327, 977