

Nagy adathalmazok osztályozásához használt aktív tanulás lekérdezési stratégiái

Papp Dávid¹, Szűcs Gábor¹

¹ *Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék, H-1111, Budapest, Magyarország*

Napjainkban az aktív tanulást egyre intenzívebb kutatómunka övezi, valamint alkalmazásának száma is jelentősen megnőtt. Aktív tanulás esetén a tanulórendszernek lehetősége van kérdéseket küldeni egy mindent tudó entitás, úgynevezett órakulom (pl. adott témában szakértő emberi erőforrás) felé. A kérdés általában egy címkézetlen entitás valós kategóriájára irányul, a válasz ismeretében pedig azt várjuk a tanulórendszertől, hogy pontosabb osztályozásra legyen képes, az aktív tanulás témáján belül pedig a képosztályozásra fókuszáltam. Célszerű mindig azt a címkézetlen képet lekérdezni, mely valódi címkéjét megismerve a tanulórendszer a lehető legtöbb információhoz jut. A kérdés már csak az, hogy miként határozzuk meg egy címkézetlen kép információtartalmát. Ez a legfontosabb építőeleme egy aktív tanulórendszernek, hiszen ez alakítja ki a címkézetlen képek lekérdezésének sorrendjét. Meghatározására számos lekérdezési stratégiát dolgoztak ki a szakirodalomban, például bizonytalansági mintavételezés, várható modell változása, verziótér csökkentés és várható hiba minimalizálása. Kutatási munkám során különböző lekérdezési stratégiákat dolgoztam ki és teszteltem, melyek az előbb felsoroltak közül az első kettő lekérdezési stratégia családba tartoznak. Az elvégzett mérések megmutatták, hogy az általam javasolt stratégiák képesek olyan sorrend kialakítására, melyet követve magasabb osztályozási pontosság érhető el azonos számú címkézetlen kép lekérdezését követően. Ez azért fontos, hiszen nem minden esetben van lehetőségünk az összes címkézetlen kép megtanulására, így érdemes egy olyan részhalmazt kiválasztani, mely alapján minél pontosabb becslést tudunk tenni.

[1] Burr Settles: Active learning. Synthesis Lectures on Artificial Intelligence and Machine Learning 6(1), 1-114 (2012).

[2] Papp D. and Szűcs G.: Balanced Active Learning Method for Image Classification. Acta Cybernetica 23(2), 645-658 (2017).

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg (EFOP-3.6.2-16-2017-00013).