

Megoldási stratégiák azonosítása gépi tanulási módszerekkel az OECD PISA 2012 adatai alapján

Megyeri István

II. évf. programtervező informatikus MSc

Szegedi Tudományegyetem, Természettudományi és Informatikai Kar

Témavezetők:

Dr. Farkas Richárd, adjunktus, SZTE TTIK Számítógépes Algoritmusok és Mest. Int. Tanszék

Hegedűs István, PhD hallgató, SZTE TTIK Számítógépes Algoritmusok és Mest. Int. Tanszék

A papír alapú tesztelés mellett napjainkra megjelentek a számítógépes tesztek és egyre szélesebb körben alkalmazzák a tanulók tudásának, kompetenciáinak mérésre. Ennek több előnye is van a papír alapúval szemben, mint például költséghatékonyabb; újfajta, komplex feladatok implementálását teszi lehetővé és a kitöltők interakcióit log fájlok formájában tárolja. Ezen log bejegyzések alapján a kitöltők viselkedése pontosan reprodukálható és hatékonyabban elemezhető. Azonban a log fájlok manuális feldolgozása kihívásokkal teli, különösen több ezer vagy tízezer kitöltő esetén. Így nagymintás felmérések esetén a kutatási cél eléréséhez különösen szükséges a számítógéppel támogatott információkinyerés. A dolgozatomban az OECD-PISA vizsgálat problémamegoldás tesztjének egy feladatához tartozó log bejegyzések alapján mutatom be, hogyan nyerhetjük ki ezen információkat, azaz hogyan azonosíthatunk megoldási stratégiákat gépi tanulói módszerekkel.

Két elemzési módszert mutatok be és hasonlítok össze az elért eredmények alapján. Az első egy automatikus módszer, mely során egy klaszterező algoritmussal azonosíthatóak a különböző megoldási stratégiák. A második egy fél automatikus módszer, mellyel a kitöltők egy kis részhalmazát felcímkézve egy osztályozó algoritmus képes a viselkedési minták azonosítására.

A diplomamunka fő eredménye, hogy a fél automatikus módszer által készített modell, emberek számára értelmezhető és 93.27% pontossággal képes a kitöltőket az előre definiált osztályokba sorolni.