

Osztályozási algoritmusok kiválasztásának vizsgálata tanuló algoritmus felhasználásával

Orosz Mónika

III. évf. programtervező informatikus BSc

Témavezető: Dr. Dombi József Dániel

SZTE TTIK Szoftverfejlesztés Tanszék

A Mesterséges Intelligencia napjaink egyik leggyorsabban fejlődő tudományterülete és talán az egyik legfontosabb, mind az informatikában, mind más tudományokban. Az MI egyik ága, a Gépi tanulás, azt jelenti, hogy a számítógépes rendszerek képesek anélkül tanulni, hogy közvetlen programozva lennének. A Felügyelt tanulás, a Gépi tanulás egyik fajtája. A Felügyelt-, azaz Induktív tanulás során, egy probléma esetén, adottak megoldások, melyek tanárként, egyfajta útmutatóként viselkednek a tanulás során. Így jön létre az általánosítás, ami ismeretlen bemenetre megfelelő megoldást tud nyújtani. Az Osztályozási módszerek is ezen kategóriába sorolhatók.

Az Osztályozási-, avagy Csoportosítási algoritmusok egyik előnye, hogy számos területen alkalmazhatók, szinte minden területen, például adatbányászat, szövegelemzés, időjárás előrejelzés, orvosi diagnózis felállítása. Az Klasszifikációs módszerek lényege, hogy olyan adathalmaz felhasználásával történik a tanulás, amely példáinak osztálya ismert. A tanulás után, korábban nem látott példák osztályát képes meghatározni. Tehát korábbi megfigyelések alapján, ismeretlen attribútum meghatározása a cél, a jelenlegi attribútumok megfigyelésének segítségével. Ennek a módszernek a megvalósítására több algoritmus is ismert. Legismertebbek közé tartoznak a Döntési fák, Bayes hálók, K-legközelebbi szomszéd algoritmus, valamint a Mesterséges Neurális hálók. Más-más működésű algoritmusok, adott adathalmaz esetén különböző teljesítményt nyújthatnak. Az osztály meghatározásának pontossága nem csak a felhasznált algoritmustól, hanem az adathalmaztól is függ. Egy adathalmaz mérete, tehát példáinak- és attribútumainak száma, valamint hiányos példák fellelhetősége, attribútumok típusai egyaránt hatással vannak arra, hogy mennyire jó eredményt kapunk.

A dolgozat célja, hogy egy választ adjon arra a kérdésre, mit célszerű használni egy adott osztályozási probléma esetén. A szakirodalomban megjelentek cikkek, melyek szintén ezzel a kérdéssel foglalkoznak. A dolgozat során ezekhez hasonlóan szeretnék választ adni arra, hogy megadható-e előre egy probléma esetén, hogy mely algoritmussal lehet a legjobb eredményt elérni. A korábban említett algoritmusokat és nyers adathalmazokat használva, a kapott eredményeken egy tanuló algoritmust végrehajtva, vizsgálható, hogy mely algoritmus használata lehet a leghatékonyabb adott esetekben.