

PREDIKTÍV MODELLEK VALÓSZÍNŰSÉGI BECSLÉSEINEK BINARIZÁLÁSI LEHETŐSÉGEI A NATÉR ELEMZÉSEK MEGALAPOZÁSÁRA

LEPESI NIKOLETT¹-BOTTA-DUKÁT ZOLTÁN² - SOMODI IMELDA²

¹ELTE TTK Biológia Doktori Iskola

²MTA Ökológiai Kutatóközpont Ökológiai és Botanikai Intézet

*e-mail: lepesiniki@gmail.com

Kulcsszavak: *prevalencia, vágáspont, kontingenciatábla*

A Nemzeti Alkalmazkodási Térinformatikai Rendszer (NATÉR) projekt a klímaváltozás káros hatásainak kiértékelési vizsgálataira összpontosít a vízgazdálkodás, biodiverzitás és földhasználat vonatkozásában. Ennek része, hogy a vegetáció válaszát előre jelezzük klímaváltozási forgatókönyvekre. A vegetáció klímaváltozáshoz való alkalmazkodóképességének becslésének feltétele a vegetáció-predikciós modellek valószínűségi becsléseinek jelenlét/hiány formába való konvertálása, azaz binarizálása. Ez általában úgy történik, hogy a valószínűségi skálán felfelé haladva megvizsgáljuk, hogy adott valószínűségi szinten elvágva, azaz binarizálva a becsült eloszlást, mekkora egyezést találunk a teszt adattal. Az egyezés mértékének megállapítására nagyszámú különféle indexet vezettek be. Ezeket felkutattuk, aminek során egyértelművé vált az átfedések nagy száma, ezért külön hangsúlyt fektettünk a szinonimák azonosítására. Az indexek leírása szintakikailag is meglehetősen változatos volt, ezért egységesítettük a képleteket a faj-előfordulás modellekhez általánosan használt kontingencia-tábla (két dichotóm változóból keletkező négy cellás táblázat) konvencióit használva. A kontingencia tábla az együttes gyakoriságok táblázatos elrendezése, mely lehetőséget ad a függőváltozók közötti összefüggések megállapítására. Ez esetben a megfigyelt és a becsült jelenlétek ill. hiányok gyakoriságának kereszttáblázatáról van szó. Mivel a valószínűségek folytonosan változnak, ezért mind az egyes kontingenciatáblázatokat mind az ebből eredő index-értékeket a valószínűségi eloszlás teljes skáláján egyenletesen elosztott vágáspontok mentén számoltuk ki.

R statisztikai környezetben. Emellett teszteltük az indexeknek:

a modellezéstől és a

prevalenciától (jelenlétek aránya a bemenő adatban) való függését

Bár binarizálási módszerek számos más tudományágban előfordulnak, áttekintésünket az adott ökológiai problémára alkalmazott megoldásokra szűkítettük. A legfontosabbak ezek közül: Kappa, True Skill Statistic, True Negative Rate, True Positive Rate, Jaccard, Minimal Predictive Area, Minimized Difference Threshold, Maximized Sum Threshold, F-score, Negative Predictive Power, Positive Predictive Power.

A vizsgált indexek között nagyfokú átfedést találtunk, számos szinonimát felismertünk. Az indexek tulajdonságai alapján ajánlásokat fogalmaztunk meg, hogy mely kutatási kérdések

esetén használhatóak leghatékonyabban. Azonosítottuk a NATÉR-ben meghatározott feladathoz legjobban illeszkedő indexet. Eredményeink szerint a legkonzekvensebb döntést az Minimized Difference Threshold hozza, amely egyben prevalencia függetlennek is bizonyult. A kappa már ismert prevalencia függését mi is tapasztaltuk, de emellett a TSS index is prevalencia függőnek bizonyult.