



MODELL ALAPÚ IDŐSOR-OSZTÁLYOZÓ FEJLESZTÉSE ÉS KITERJESZTÉSE

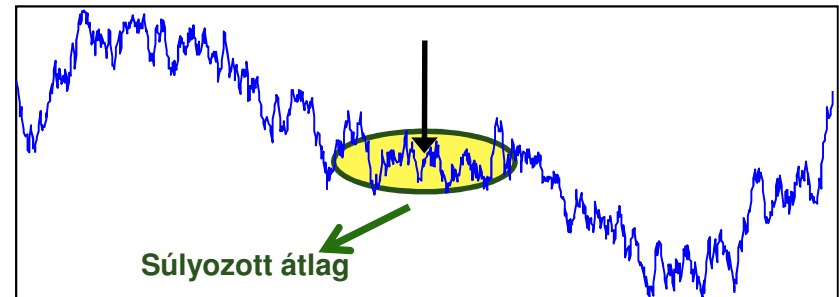
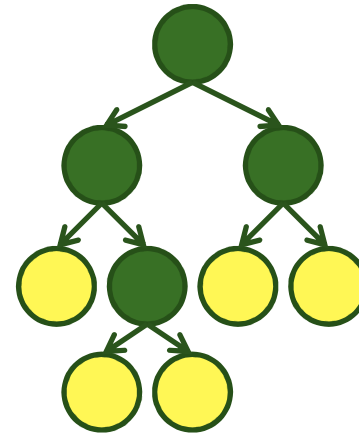
Hidasi Balázs (hidasib@freemail.hu)

Konzulens: Gáspár-Papanek Csaba (gaspar@tmit.bme.hu)

2011. JÚNIUS 21, BUDAPEST

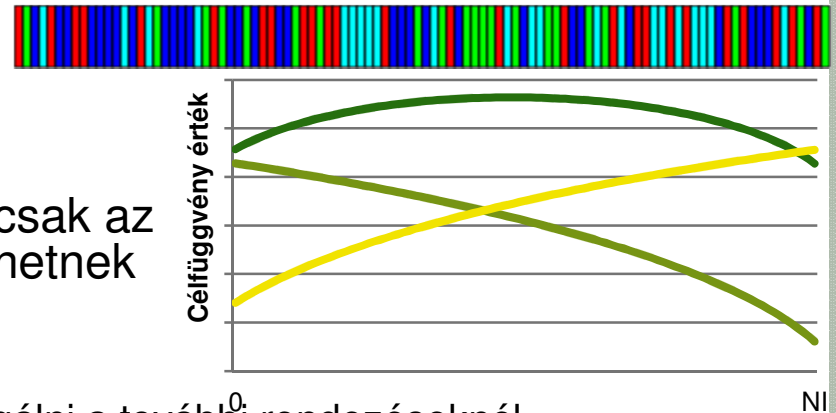
SHIFTTREE ALGORITMUS

- Idősor-osztályozó algoritmus
- Bináris döntési fa alap
- Dinamikusan előállított attribútumok
 - Hol nézzük? (időtengely)
 - Mit nézzünk? (számítások)
- Előnyök
 - Modell alapúság előnyei
 - Értelmezhető
 - Szakértői tudás nélkül is pontos
 - Szakértői tudással pontosabbá tehető
- Hátrányok
 - Modell alapúság hátrányai



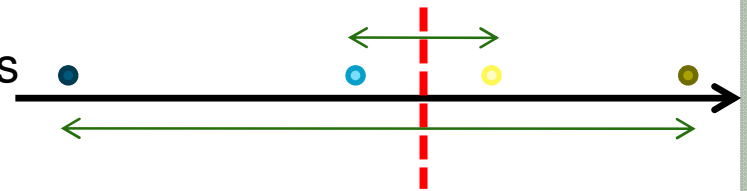
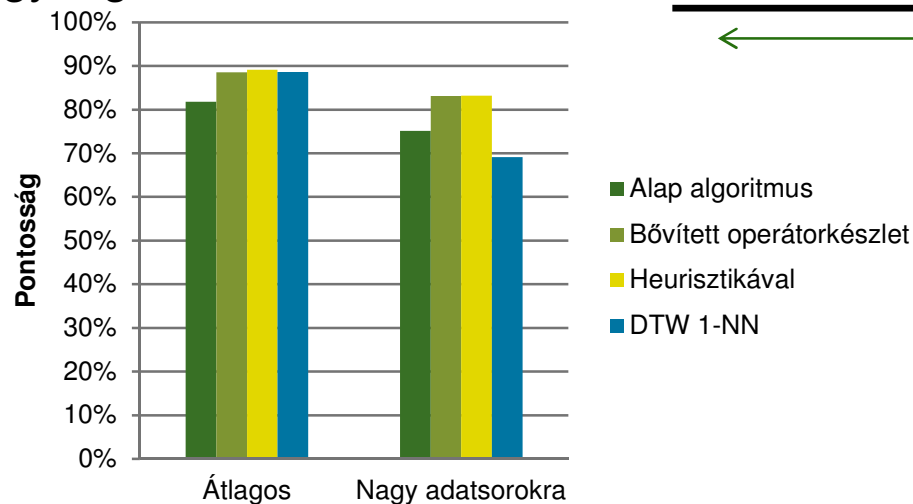
MODELL ÉPÍTÉS IDEJÉNEK CSÖKKENTÉSE

- Tanítás során attribútum választás
 - Célfüggvény minimalizálása
- Célfüggvény tulajdonságai
 - Adott rendezés mellett minimumok csak az egybefüggő intervallumok szélén lehetnek
 - Minimum előre meghatározható
 - Nem léphetünk ki minimumnál, de...
 - Ha eléri, akkor csak 2-2 helyet kell vizsgálni a további rendezéseknél
- Jelentősen csökken a célfüggvény értékének meghatározásának száma
- Futási idő átlagosan 22,33%-kal csökkent
 - Pl.: FordB – 3636 tanítóminta:
 - 214,94s → 173,52s (-19,27%)
 - Pl.: CBF – 30 tanítóminta:
 - 0,246s → 0,145s (-41,18%)
 - Pl.: Beef – 30 tanítóminta
 - 0,574s → 0,517s (-9,9%)



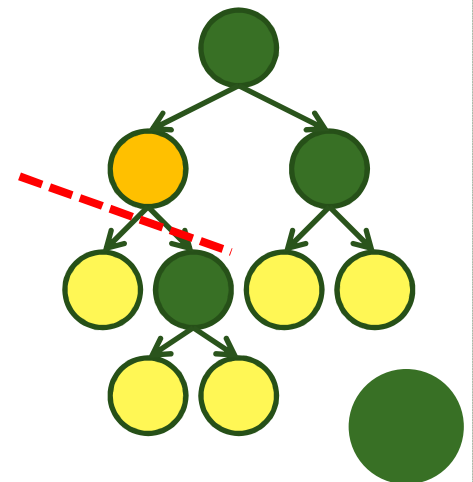
FEJLESZTÉSEK A PONTOSSÁG NÖVELÉSÉRE

- Operátorkészlet jelentős kibővítése
- Heurisztika egyezően jó attribútum jelöltek közüli választásra
 - Legyen maximális a határsáv
 - Eltérő nagyságrendek miatt normalizálás



○ Nyesés

- Több módszer kipróbálása
- Romlott a pontosság
- Más módszerekhez viszont szükség van rájuk



MODELLEK KOMBINÁLÁSA

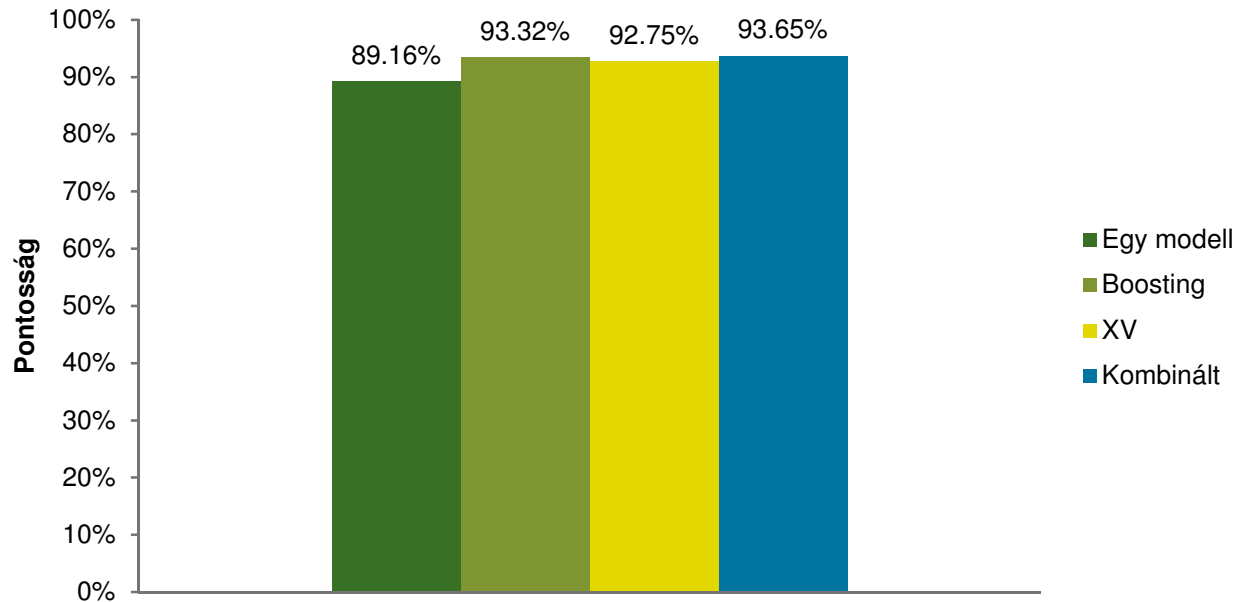
- Több modell építése
 - Címkézéskor súlyozott szavazás
 - Hatékony, ha a modellek önmagukban pontosak és az egyes modellek eltérőek (más mintákat osztályoznak jól)
- Boosting
 - Súlyok a tanítómintákhoz
 - Felépített modell kiértékelése a tanítómintákon
 - Modell súly
 - Rosszul osztályozott tanítóminták relatív súlyának növelése (nyesés!)
 - Klasszikus AdaBoost követelmény: pontosság a tanítómintán minimum 50% (és kevesebb, mint 100%)
 - SAMME: minimális pontosság követelmény csak $1/(\text{osztályok száma})$
- XV
 - Egyszerű kombinálás, keresztvalidáción alapul
 - Tanítóhalmaz véletlenszerű szétbontása, csak egy részén tanul, másik részén kiértékel
 - Pontosság \rightarrow modell súlya



KOMBINÁLT MODELLEK EREDMÉNYEI

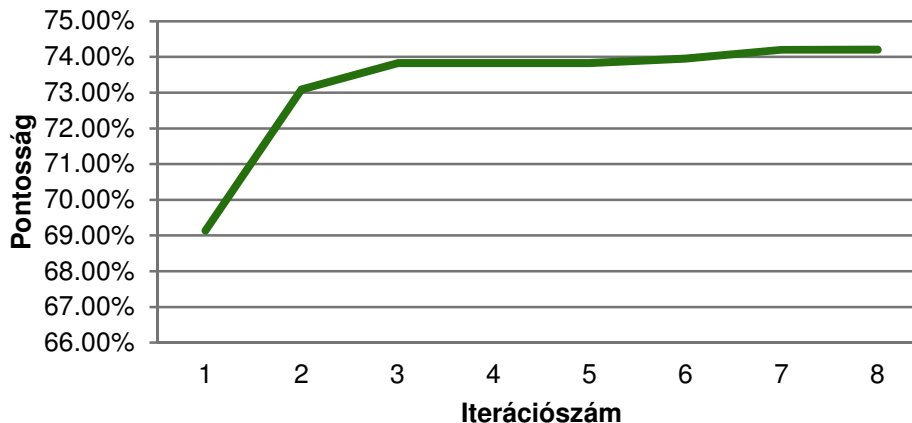
○ Boosting vagy XV?

- Bizonyos esetekben a boosting leáll
- „Tökéletes” modell a tanítóhalmazon
- Főleg kis tanítóminták mellett
- Ilyenkor az XV még segíthet



OSZTÁLYOZÁSI KONFIDENCIA

- Mennyire biztos a modellünk a kimenetben
 - Levél (csomópont) konfidencia
 - Pl. többségi osztály aránya a levélben (nyesés!)
 - Útvonal konfidencia
 - Osztályozási útvonalon a konfidenciák (súlyozott) összegzése
 - Egyfajta nyesési eljárás
- On-line tanulás (újratanulás)
 - Teljes modellépítés helyett a modell kisméretű megváltoztatása
 - Útvonal konfidencia, mint dinamikus nyesés
 - Arányok változásával változik a nyesés



ÖSSZEFOGLALÁS

- Modell alapú idősor-osztályozó döntési fa alapon
- Tanítás gyorsítása: -22% futási idő
- Pontosság növelése
 - Új operátorok
 - Attribútum választási heurisztika
 - Kombinált modellek
- On-line tanulás teljes modellépítés helyett
 - Útvonal konfidenciák használatával
- Fejlesztési lehetőségek
 - Streamben jelfelismerés (csúszóablakos megoldás)
 - Az elv alapján bármilyen félig-struktúrált/struktúrált adat osztályozására kiterjeszhető

