

Adatbányászat – Szabálytanulás

Ellenőrző kérdések

- Hogyan értelmezhető a szabályhossz paraméter szabálytanulásnál?

Egy szabály általános alakban a következőképp néz ki: [törzs]=>[fej], azaz áll egy törzsből (mely akár több elemből is állhat) és egy fej elemből. Egy szabály hossza nem más, mint a törzs elemszáma +1 (vagyis a fej). Tehát az [A,B,C]=>[D] szabály hossza: 4 (= 3 törzselem +1 fej).

- Mit értünk asszociációs szabály alatt?

Az asszociációs szabálytanulás egyik népszerű alkalmazása a vásárlói kosarak elemzése, ahol a cél a vásárlói kosarak elemei közötti függések feltárása. Lényegében olyan asszociációs kapcsolatok (szabályok) megismerésére irányul, melyek alapján kimondható, hogy ha az elemhalmazban szerepel egy (vagy több) adott elem, akkor egy másik elem is nagy valószínűséggel szerepel benne. Tehát például a vásárlói kosarak elemzése esetén egy lehetséges szabály a következő: ha a vásárló vesz hot-dog kiflit és virslit, akkor vesz hozzá mustárt is. Ez strukturálisan a következőképp írható le: [‘hot-dog kifli’ + ‘virslit’]
=>[‘mustár’].

- Mit értünk szekvenciális szabály alatt?

Szekvenciális szabálytanulás esetén a feladat komplexebb, mivel az idő dimenziót is figyelembe kell venni. Célja az összetartozó tranzakciók láncolatának elemzése, illetve azok időbeliségét figyelembe véve időbeli kapcsolatok azonosítása.

//magának a szabálynak a def.ja még nincs itt

- Mit jelent a konfidencia szabálytanulásnál?

Konfidencia minimum: A konfidencia az adott szabály bizonyosságának a mértékét jelöli. *Egy 50%-os konfidencia érték azt jelenti, hogy ahol az adott szabálytörzs igaz (pl.: vásárlói kosárban jelen van egy 'A' termék), ott az esetek 50%-ban az adott szabály fejrésze is igaz lesz (jelen van a kosárban 'B' termék).* Tehát a konfidencia minimum egy küszöb, mely alapján csak azok a szabályok kerülhetnek be a végeredményként elálló szabálygyűjteménybe, melyeknél a konfidencia az adott küszöbnél magasabb.

Ha a szabály törzse X és feje Y, akkor konfidencia alatt az Y X-el vett feltételes valószínűségét értjük: $P(Y|X)$

- Definiálja a support (támogatottság) fogalmát!

Support minimum: *A support megadja, hogy az adott szabály a rekordok hány százalékában igaz. Az eredményhalmazba csak azon szabályok kerülnek bele, amelyek támogatottsága meghaladja az itt megszabott küszöbértéket.*

Ha a szabály törzse X és feje Y, akkor támogatottság alatt az X, Y előfordulási valószínűségét

értjük: $P(X,Y)$

- Mit jelent és hogyan értelmezhető a lift érték szabálytanulás során?

Lift: A lift a konfidencia és a támogatottság (support) hányadosa, azt reprezentálja, hogy egy adott szabály a várhatónál mennyi többletinformációt tartalmaz.

(Erre nem adható meg küszöb ezen a felületen, csak a teljesség kedvéért szerepel itt.)

- Lift > 1: a két termék kiegészíti egymást, tehát a kosár nagyobb valószínűséggel tartalmaz 'B'-t, ha van benne 'A' is.
- Lift < 1: a két termék helyettesíti egymást, a vásárlók jellemzőn vagy az egyiket, vagy a másikat veszik. Tehát ha 'A' szerepel a kosárban, akkor 'B' kisebb valószínűséggel fog szerepelni.
- Lift = 1: az 'A' nem befolyásolja 'B' konfidenciáját, vagyis attól, hogy 'A' benne van a kosárban, nem változik annak a valószínűsége, hogy 'B' szerepel-e benne. Tehát 'A' és 'B' függetlenek.

Hányszorosára növeli X jelenléte Y előfordulását: $P(Y|X) / P(Y)$

- Mi a különbség az oksági kapcsolatok és az asszociációs szabályok között irányultság szempontjából?

- Definiálja egy oksági kapcsolat ($X \rightarrow Y$) létezésének kritériumait!

1. erős statisztikai asszociáció
2. X időben előzze meg Y -t
3. plauzibilis magyarázat létezzen, úgy hogy ne legyenek alternatív, zavaró tényezőre is építő alternatív magyarázatok
4. szükségesség (az ok megszüntetésével a hatás is szűnjön meg)
5. elégségesség (az ok bekövetkeztével a hatás is erősödjön)

Egy másik Max Born-tól származó feltétel lista a következő:

1. Y előfordulása törvényszerű kapcsolatban függjön X előállításától
2. az oknak időben megelőzőnek, vagy egyidejűnek kell lennie
3. térbeli vagy fizikai törvényszerűségeken keresztüli szomszédosság

- Ismertesse Reichenbach: Common Cause elvének lényegét (4. old.)!

H.Reichenbach's: Common Cause Principle: a correlation between events A and B indicates either that A causes B, or that B causes A, or that A and B have a common cause.

magyarul: közös ok elve: A és B esemény közötti korrelációból következik, hogy vagy A

okozza B-t, vagy B okozza A-t, vagy A-nak és B-nek közös oka van.

- Mit jelent a lokális szabályalapú rendszerekben?

Lokális (locality): Logikai rendszerekben, ha van $A \Rightarrow B$ formájú szabály, akkor a B kikövetkeztethető, ha az A bizonyíték adott, bármely más szabályra való tekintet nélkül.

- Mit jelent a leválasztás (detachment) szabályalapú rendszerekben?

Leválasztás (detachment): Ha egyszer találtunk egy logikai bizonyítást B állítás kikövetkeztetéséhez, az állítást tetszés szerint felhasználhatjuk, a származtatásától függetlenül.

- Oksági kapcsolatok vagy asszociáció erősségének mérésére szolgál a kölcsönös információ (mutual information) mutató?