

TEMATIKA

2010/2011 tavaszi félév

SZTE Eötvös Loránd Kollégium



1. **Dombi József: Fuzzy elmélet és alkalmazásai**
2011. március 3. 19:00
2. **Móra György:**
Információkinyerés természetes nyelvű szövegekből
2011. március 31. 19:00
3. **Nagy István: Tanulás pozitív és jelületlen példák**
2011. április 1. 8:30
4. **Krész Miklós: Gráf alapú adatbányászat**
2011. április 15. 8:00
5. **Harka Győző: Internetes kataklizmák**
2011. április 19. 19:00 (Eötvös Esték keretében)
6. **Berend Gábor: Kulcsszókinyerési technikák**
2011. április 21. 19:00
7. **Jelasity Márk: Elosztott adatbányászat**
2011. április 29. 9:00
8. **Bánhalmi András:**
Neurológiai betegek adatainak feldolgozása
2011. május 12. 19:00

Dombi József:

Fuzzy elmélet és alkalmazásai

Időpont: 2011. március 3. 19:00

A fuzzy logika – a többértékű logika egy formája –, 1965-ben Zadeh egy tanulmányával született meg. A fuzzy logikában nem pusztán "igen" és "nem" létezik, hanem átmenet is e kettő között. Maga a fuzzy szó *elmosódottat* jelent, utalva a bizonytalanságra.

Ez a tudományterület az elmúlt fél évszázadban nagy teret nyert nem csak az informatikában, hanem a nyelvtudományi és logikai szemantikában, a matematikai logikában és a valószínűségelméletben is.

Az előadás kitér a fuzzy logika alapjaira és a fuzzy logikán alapuló irányításra, mely egyre gyakrabban használt napjainkban elsősorban gépek, robotok illetve háztartási készülékek irányítására.

Móra György: Információkinyerés természetes nyelvű szövegekből

Időpont: 2011. március 31. 19:00

- Mi az a természetes nyelvfeldolgozás
- Milyen feladatok vannak (információ-visszakeresés, kivonatolás, illetve információkinyerés)
- Mi a kettő közötti különbség, melyik mire jó
- Hol van szükség automatikus információkinyerésre
- Mik azt alapvető technikái, milyen módszereket használnak (szakértői rendszerek, gépi tanulás, hibrid rendszerek, nyelvészeti eszközök)
- Gépi tanulási megközelítések
- Hogyan mérhető a teljesítmény (kiértékelési metrikák, versenyek)
- Tudomány vs. ipar

Nagy István: Tanulás pozitív és jelöletlen példákból

Időpont: 2011. április 1. 8:30

Szövegek osztályozása egy igen fontos és inenzíven kutató terület az információkinyerésen és a gépi tanuláson belül. Ahhoz, hogy egy szövegosztályozót konstruálhassunk, szükséges tanítópéldák gyűjtése, melyek fel vannak címkézve az előre definiált osztályok címkéivel. Ez a címkézés gyakran manuálisan történik, ami igen nehéz és költséges folyamat eredménye.

Ahhoz, hogy egy osztályozó modellt kapjunk egy osztályozó algoritmust kell tanítani ezeken a tanítópéldákon. Az ilyen típusú osztályozók a felügyelet gépitanulás osztályába tartoznak mivel, minden tanítópélda fel van címkézve az algoritmus számára.

Előadásomban speciális, nem felügyelt tanulás osztályba tartozó megközelítéseket fogok ismertetni. Kétsztályos (pozitív és negatív) problémákra fókuszálva, mely során csupán pozitív tanítópéldákból tanítjuk az algoritmusokat.

Krész Miklós:

Gráf alapú adatbányászat

Időpont: 2011. április 15 8:00

Az utóbbi években a nem struktúrált adatok elemzése központi kérdéssé vált mind az ipari (telekommunikáció, pénzügyi szféra), mind az akadémiai (bioinformatika, szociológia, közgazdaságtan) kutatás terén. A fenti típusú adatok reprezentációjában a rejtett gráfstruktúra rendkívül hasznosnak bizonyult az elemzések során.

Az előadás során bemutatásra kerülnek olyan gráfstruktúrára épülő modellek és hatékony algoritmusok, melyek különböző adatbányászati elemzések alapját képezhetik nagyméretű (milliós nagyságrendű), strukturálatlan adathalmazokon is.

Szintén ismertetjük azon projekteket (biológia, környezet-tudomány, telekommunikáció, pénzügyi szféra), ahol a kifejlesztett módszerek alkalmazást nyertek.

Berend Gábor: Kulcsszókinyerési technikák

Időpont: 2011. április 21. 19:00

Az elektronikus, azon belül is az online szöveges tartalmak méretének robbanása újszerű megközelítéseket tesz szükségessé kategorizálásukra. Egy ilyen elterjedt módszer az ún. címkézés eljárása, amely során dokumentumainkat az azokat tömören és jól leíró kulcskifejezésekkel látjuk el. A szöveges dokumentumokból történő kulcsszókinyerés számos alkalmazási területen hasznosítható, a katalogizáló- és kivonatoló rendszerektől kezdve egészen az információ-visszakereső módszerekig.

A kulcsszavakkal el nem látott dokumentumok mennyiségének folyamatos és markáns növekedése miatt azonban feldolgozásuk csupán automatikus eszközök segítségével képzelhető el. A kulcsszavak egy része egzaktul a szövegben is megtalálható, de kulcskifejezések lehetnek olyan absztrakt címkék is, amik a dokumentumban magában nem fordulnak elő, mégis szemantikus kapcsolatba hozhatók a leírtakkal.

Előadásomban az absztrakt és nem absztrakt kulcsszavak eltérő doménekből (tudományos publikációk, hírdokumentumok) történő kinyerésére fókuszálok majd.

Jelasity Márk:

Elosztott adatbányászat

Időpont: 2011. április 29 9:00

A nagyméretű elosztott rendszerekben, mint pl. a peer-to-peer hálózatok, vagy szenzorhálózatok, nagy mennyiségű adat keletkezik ill. áll rendelkezésre teljesen elosztott formában, azaz a hálózat egyes csomópontjaiban szétszórva. Ennek a hatalmas adatmennyiségnek a központi feldolgozása költséges, és adatvédelmi problémákat is felvet.

Az előadásban ismertetek olyan algoritmusokat, amelyek képesek tolerálni ezeknek a hálózatoknak az ellenséges körülményeit: a nagy üzenetvesztési valószínűséget, a csomópontok esetleges meghibásodásait, a késleltetéseket, és a nagyon nagy méretet is. A csomópontok csak lokálisan kommunikálnak, nincs egy csomópontnak sem kitüntetett szerepe, mégis hatékonyan konvergál a rendszer állapota az általunk megkívánt állapotba, ami pl egy gépi tanulási modell az adatok felett, vagy valamilyen globális számítási feladat eredménye.

Bánhalmi András: Neurológiai betegek adatainak feldolgozása

2011. május 12. 19:00

Egy, a neurológiai betegek otthoni ápolását, monitorozását célzó projekt kapcsán felmerülő néhány összetettebb adatelemzési, kiértékelési feladatot vázlok fel, ezek közül:

- nyelvi tesztek, feladatokhoz kapcsolódóan a beteg kiejtésének kiértékelése
- kognitív tesztekhez kapcsolódóan érintőképernyős rajzoló feladat kiértékelése
- otthoni torna végzése (csuklóra rögzített gyorsulásmérő)
- alvás monitorozása (nyomá szenzor-mátrix jeleinek kiértékelése)
- mozgásérzékelő hálózattól származó adatsorok