

Ormándi Róbert

Szegedi Tudományegyetem TTK

Konzulens: Dr. Kocsor András

Tudományos főmunkatárs

SZEMANTIKUS TÉR KERNEL-ALAPÚ TANULÁSA

Párhuzamos adatbázisnak nevezzük az olyan adatbázisokat, amelyek elemei egy jelenséget két vagy több reprezentációval is ábrázolnak. A szemantikus tér kernel-alapú tanulásának célja, az adott párhuzamos tanuló adatbázis minden reprezentációja számára olyan vetítési irányok megkeresése, amelyek mellett a különböző reprezentációbeli leírások vetítettjeinek korrelációja maximális. Az eljárás alkalmazható különböző nyelvű, természetes nyelvi szövegekből álló párhuzamos tanuló adatbázisokra/párhuzamos korpuszokra. Ebben az esetben a vetített elemeket tartalmazó tér interpretálható, mint szemantikus jelentéstér, hiszen a jelentésükben megegyező párok korrelációja feltételezhetően ekkor lehet maximális.

A szemantikus tér tanulás fontos szerepet kap a természetes nyelvi feldolgozás számos területén, hiszen alapvető a nyelvfüggetlen feldolgozás szempontjából. A módszer sikerrel alkalmazható akár eltérő típusú objektumok (például képek és szövegek) összerendelésének tanulására is.

A dolgozat célja a szemantikus tér tanulására alkalmazott eddig ismert kernel módszerek matematikai és implementációs részleteinek bemutatásán túl, alkalmas szöveg és karakterlánc kernelek definiálása és vizsgálata. Bevezetésre kerül továbbá egy újszerű hibrid eljárás is, amely tanuló módszerek kombinációját és kernel-paraméter optimalizációt alkalmaz. A hatékonyság mérésére párhuzamos adatbázisok egyik reprezentációján véletlen permutációt hajtunk végre és a módszertől a helyes permutáció visszaállítását várjuk el. A hatékonyság mérése permutációs mértékek használatával történik magyar-angol párhuzamos korpuszokon.