

Iván Szilárd, Ormándi Róbert
IV. éves programtervező matematikus hallgatók
Szegedi Tudományegyetem TTK

Konzulens: Dr. Kocsor András
Tudományos főmunkatárs

Magyar mondatok újszerű kernel-alapú szintaxis elemzése

Motiváció:

A szintaxis elemzés célja az input mondat nyelvtani struktúrájának meghatározása. A struktúra egy hierarchikus szerkezet, ahol a legnagyobb egység a mondat, legkisebb egységei pedig az alapszimbólumok (például a szavak szófajai, vagy akár maguk a szavak is lehetnek). Az ilyen típusú szerkezetet általában egy fával szokás reprezentálni, melynek gyökerében az egész mondatnak megfelelő csúcs áll, levelei az alapszimbólumokat, a belső csúcsok pedig az egyéb nyelvtani egységeket reprezentálják.

Az ilyen típusú elemzésnek rendkívül fontos szerepe van a természetes nyelvi feldolgozás számos területén, hiszen egy mondat szintaxisfájának helyes meghatározása alapvető fontosságú a magasabb szintű szövegfeldolgozáshoz (például szemantikai elemzés, vagy gépi fordítás).

Alapvetően kétféle megközelítés létezik, az egyik a szakértők által megadott összefüggéseken alapuló, a másik a gépi tanulást előtérbe helyező eljárások. Manapság a figyelem az utóbbi módszerekre összpontosul, angol nyelvre igen hatékony algoritmusok kerültek kidolgozásra, de a magyar nyelv sajátosságai (nyelvi variabilitás) miatt ezek változatlan formában történő alkalmazása jelentős hatékonyságvesztéssel jár.

Eredmények:

Jelen dolgozatban bemutatunk egy gépi tanuláson alapuló szintaxis-elemző eljárást, amely a manapság intenzíven kutatott kernel-alapú módszerek gondolatmenetét követi. A kidolgozott eljárás a szintaxisfákat mint derivációs fákat értelmezi egy előre rögzített CF nyelv szabályai alapján. Ezeket a fákat jól jellemzi, hogy a deriváció során az egyes szabályok hányszor lettek alkalmazva. Voltaképpen a szabályok alkalmazásának eloszlására definiáltunk egy újszerű kernel-alapú tanuló eljárást. A módszer algoritmikus részleteinek bemutatásán túl, egy releváns gyakorlati feladaton keresztül igazoljuk a bevezetett eljárás létjogosultságát.