

תקציר

חילוץ תלויות תחביריות (Dependency Parsing) הינה משימה חשובה בתחום עיבוד השפה הטבעית, המשמשת כאבן דרך ביישומי קצה רבים. המטרה המרכזית במשימה זו, הינה מציאת יחסים בין מילים במשפט נתון. יחסים אלה כפופים לאילוצים מבניים, אשר חלקם "קשים" (כלומר מתקיימים לכל משפט) וחלקם "רכים" (מתקיימים ברב המשפטים).

בעבודתנו, אנו מציגים מסגרת כללית לשימוש בידע מבני מוקדם עבור חילוץ תלויות, תוך שימוש במשפטים לא מתתויגים. הרעיון המרכזי הינו להעניש הפרת האילוצים במסגרת פונקציית המטרה. אנו מתרכזים בשני סוגי אילוצים מוכרים על מבנים אלו – עליהם להיות חסרי מעגלים ופרוג'קטיביים. הפונקציות שבעזרתן נבצע זאת, הינן בלתי-תלויות בתיוג הנכון למשפט, ולכן מתאימות למסגרת של למידה מונחית למחצה (Semi Supervised Learning). שיטה זו יכולה להשתלב בכל מערכת חילוץ מבוססת-גרפים מסדר ראשון (first-order graph-based parser) הנלמדת על ידי שיטות גרדיאנט.

אנו מראים אמפירית כי השיטה משפרת את ביצועי המערכת ביחס למערכת הבסיס, בייחוד בעת שימוש במספר קטן של דוגמאות מתויגות. בנוסף, אנו מדגימים כי תדירות המבנים ה"חוקיים" בזמן טסט עולה משמעותית. לסיכום, אנו משתמשים בתהליך היסק חמדן, ומראים שניתן להפחית את זמן הריצה שלו על ידי הגישה המוצגת.