

הצעת פרויקט

דרור פורת dgingi@gmail.com , אורי יונאי oryjonay@campus.technion.ac.il

תיאור הבעיה:

הבעיה שאנו מעוניינים לפתור הינה הבעיה:

בהניתן משחק בין שתי קבוצות כדורגל, חיזוי תוצאת המשחק (ניצחון, תיקו, הפסד). או, באופן פורמלי: בהינתן (t_1, t_2) משחק בין הקבוצה t_1 ("קבוצת הבית") לבין הקבוצה t_2 ("קבוצת החוץ"), נרצה לסווג האם המשחק יסתיים בניצחון ל- t_1 (ערך 1), תיקו (ערך 0), או בניצחון ל- t_2 (ערך -1).

תיאור דרך פתרון הבעיה:

את הבעיה נפתור כפי שפותרים בעיות למידה וסיווג, כאשר המשחקים (t_i, t_j) הם הדוגמאות ו- $\{1, 0, -1\}$ הם הסיווגים האפשריים. נפרט את השלבים הדרושים לפתרון הבעיה:

1. איסוף בסיס נתונים - איסוף סטטיסטיקות על ליגות שונות בשנים שונות ע"מ ליצור אוסף דוגמאות למידה.
2. חיפוש תכונות (features) המתאימות לבעיה (למשל, מס' שערים כולל של החלוצים בכל קבוצה לעומת מס' השערים שספגו וכו').
3. מציאת התכונות אשר יביאו למסווג מדויק ביותר (ע"י אלג' חיפוש בגרפים אינסופיים כמו אלג' גנטיים).
4. יצירת מסווג ע"פ התכונות שנמצאו בשלב 3.

תאור המערכת:

חלקי המערכת המוצעת הינם:

1. *Crawler* לאיסוף מידע מהרשת.
2. *Database* אשר יחזיק את כל הנתונים, כמו כן אנו נרצה שה-*Database* יוכל לגדול באופן דינאמי.
3. המסווג.

תיאור הניסויים:

הניסויים אותם נבצע מתחלקים לשתי קבוצות:

1. ניסויים המתעסקים בכיול הפרמטרים, לדוגמא:
- מציאת מס' המשחקים הקודמים הדרושים על מנת לקבל את התוצאות האופטימליות.

- מציאת קבוצת התכונות האופטמליות (ניסוי מרכזי אשר יבוצע לצורך השוואה עבור שני אלג').

2. בדיקת תוצאות, לדוגמא:

- לקיחת ליגה אקראית אשר קיימים עבורה כל הנתונים מהעבר, ביצוע למידה על תקופה מסויימת ובחינה על העונה העוקבת.
- בחירת משחקים מהתקופה הנוכחית ממגוון ליגות ובדיקת אחוז ההצלחה בניחושים.

שונות:

לוח זמנים:

1. כשבועיים לאיסוף המידע הרלוונטי מהאינטרנט ולכתיבת ה-*crawler*.
2. כעשרה ימים לבחירת התכונות בהן נרצה להשתמש.
3. יצירת המסווג והרצת הניסויים שתיארנו עד סיום תקופת בחינות.
4. הגשת דו"ח מסכם עד ה-30/9.

שפת תכנות וסביבת תכנות:

את הפרוייקט נכתוב ב-*Python 2.7*, סביבת העבודה תהייה מבוססת *Windows* תחת *PyDev*.

ספריות ומודולים:

1. *Sklearn* - ספרייה המממשת אלגוריתמי למידה.
2. *PyMongo* - ספרייה המאפשרת התממשקות עם שרת *NoSQL*.
3. *Numpy, Scipy* - ספריות מתמטיות.
4. *Sphinx* - ספרייה ליצירת תיעוד אוטומטי.