

פרויקט בבינה מלאכותית – הצעת פרויקט

מערכת כוללת לכוונון פרמטרים במערכות לומדות

שלומי לוי, 308025659

מטרה ראשית:

מערכת לומדת, אשר בהינתן הגבלת זמן וקבוצת אימון, מחזירה (כקופסא שחורה) את המסווג הטוב ביותר.

תכנון הפרויקט בשלבים:

הגדרות ראשוניות

1. הגדרת 10 קבוצות אימון שעליהם אעבוד. טווח דוגמאות בכל דאטה סט הינו 1000-5000.
2. הגדרת 4-5 אלגוריתמי למידה.
3. הגדרת 3-5 פרמטרים לכל אלגוריתם למידה. לכל פרמטר נגדיר קבוצת ערכים אפשרית (ערכי הפרמטר יכולים להיות רציפים או בדידים).

feature selection:

1. בחירת התכונות הוא שלב באלגוריתם שאותו ננסה לפני כיוונון הפרמטרים ואחרי כיוונון הפרמטרים.
2. ככל שנשקיע בשלב זה יותר זמן, בחירת התכונות תהיה טובה יותר מצד אחד ומצד שני, יישאר פחות זמן לשאר שלבי המערכת.
3. בשלב זה אני רוצה לבחון מה יותר אפקטיבי למערכת שלנו, האם לתת הרבה זמן לבחירת תכונות והבחירה תהיה איכותית, אבל תשאיר מעט זמן לכל השלבים האחרים, או שמא לתת מעט זמן לשלב הזה (בחירה פחות איכותית), ולהשקיע יותר זמן בשלבים האחרים.
4. כאן יבחנו מספר אפשרויות של בחירת תכונות שלמדנו בקורס בבינה כמו למשל: חיפוש מקומי קדמי, חיפוש מקומי אחורי וחיפוש מקומי כללי עם Restart.

כוונון פרמטרים:

1. יבוצע בניסיון ראשון לאחר בחירת התכונות ובניסיון שני לפני בחירת התכונות.
2. כאמור, לכל אלגוריתם למידה יוגדרו מספר פרמטרים, ולכל פרמטר כזה יוגדרו מספר ערכים אפשריים.
3. כיוונון התכונות יבוצע באמצעות cross validation שיבוצע בדיוק ע"י אותן הדוגמאות לכל האלגוריתמים הקיימים במערכת, ע"מ לשמור על אחדות ולמנוע שונות בין הבדיקות של כל האלגוריתמים.
4. גם כאן באה בחשבון חלוקת הזמנים, שכן, צריך לבדוק האם לתת לשלב זה זמן ממושך ע"מ שיבוצע בצורה איכותית, אבל יותר פחות זמן לשאר השלבים במערכת, או שמא לתת לשלב זה מעט זמן, ולהקטין את מספר הדוגמאות המשתתפות בתהליך ה cross validation, ולקבל כיוונון פרמטרים פחות איכותי.

בסוף, תיווצר לנו מערכת לומדת, **הלומדת** מהו המסווג הטוב ביותר עבור אותה קבוצת אימון (בהינתן הגבלת הזמן שניתנה לנו). כלומר, המערכת שלנו, תבחר אלגוריתם למידה, תסנן תכונות, תבחר פרמטרים ותכוון אותם, תאמן מסווג עבור קבוצת האימון, ומסווג זה יהיה המסווג הטוב ביותר עבור קבוצת האימון.

בדיקת התוצאות:

1. הדיוק של המערכת שלי על קבוצת המבחן ייבדק אל מול תוצאות שהתקבלו ונכתבו במאמרים שהשתמשו באותו הדאטה (במידה וקיים), או אל מול מפרסם הדאטה והדיוקים שהוא הצליח להגיע אליהם.

הערות נוספות:

מספר גרפים שאני מתכוון להראות, שיענו על מספר שאלות המעניינות אותי:

1. גרף 1 – דיוק על קבוצת המבחן בהינתן הזמן שניתן לכל התהליך.
2. גרף 2 – דיוק על קבוצת המבחן כפונקציה של הזמן הניתן לכוונון הפרמטרים.
3. גרף 3 – דיוק על קבוצת המבחן כפונקציה של הזמן הניתן לבחירת התכונות.
4. גרפים נוספים – דיוק על קבוצת המבחן בהינתן חלוקת זמנים שונה בין השלבים השונים בתהליך.