מערכות המלצה, תרגיל מספר 1

Rating Prediction

כללי: בתרגיל זה נממש שיטות שונות לחיזוי rating. במהלך התרגיל נריץ ניסויים על מאגר דירוגים של דירוגי עסקים על ידי משתמשים. רצוי מאוד לכתוב קוד מודולרי ככל האפשר, מאחר והתרגילים הבאים יתבססו על הקוד שכתבתם בתרגיל זה. הינכם רשאים לממש את התרגיל בכל דרך שתבחרו, תוך הקפדה על הממשק המוגדר במחלקה RecommenderSystem.

1. הורידו את בסיס הנתונים -

<https://github.com/sidooms/MovieTweetings/tree/master/latest>

המכיל דירוגים של משתמשים על סרטים (אנו מעונינים בעיקר ב-rating.dat). הסברים על בסיס הנתונים ניתן למצוא ב- <http://recsyswiki.com/wiki/Movietweetings>.

1. ממשו את השיטה Load הטוענת את המידע מהקובץ לבסיס נתונים. רצוי לממש את בסיס הנתונים באופן יעיל כפי יכולתכם על מנת לסייע במימוש השלבים הנוספים.
2. ממשו את השיטה GetRating במחלקה RecommenderSystem – השיטה מחפשת בבסיס הנתונים את הדירוג (הקיים, ללא תחזיות) שנתן משתמש לפריט (פרופיל). במידה והמשתמש לא נתן לפריט דירוג התנהגות השיטה אינה מוגדרת.
3. ממשו את השיטות
   1. GetUsers – מחזירה רשימה של כל המשתמשים במערכת.
   2. GetItems – מחזירה רשימה של כל הפריטים במערכת.
   3. GetRatedItems – מחזירה את רשימת הפריטים שמשתמש דירג.
4. ממשו מערכת תחזיות החוזה ציונים לסרטים על ידי ציוני המשתמשים הדומים לו:
   1. Pearson Correlation – השתמשו בנוסחת קורולצית פירסון על מנת לחשב דומות בין משתמשים. חזו את ציוני הסרטים באמצעות ממוצע משוקלל של ציוני המשתמשים הדומים. שימו לב לא לחשב דומות בין המשתמש לעצמו, כיוון שזה יכול ליצור סטיות גדולות בחישובים. יש לבצע התיחסות מיוחדת למשקלות שליליים, כפי שנלמד בכיתה.
   2. Cosine Similarity – חזרו על הסעיף הקודם, כאשר חישוב הדומות בין משתמשים נעשה על ידי מדידת הזוית בין הוקטורים.
   3. עליכם למממש את השיטה PredictRating המקבלת שיטה, משתמש, ופריט, וחוזה את הדירוג של הפריט על ידי המשתמש באמצעות השיטה המתאימה. ערכים אפשריים לשם השיטה –Pearson, Cosine.
5. Random – ממשו מערכת החוזה ציונים באופן רנדומלי, על ידי בחירת ציון מתוך התפלגות הציונים של המשתמש (זה לא מסובך כמו שזה נשמע). הוסיפו לשיטה PredictRating את השיטה תחת השם Random.
6. ממשו השוואה בין אלגוריתמי החיזוי על ידי MAE. כדי לקבל ציון עבור אלגוריתם חיזוי נשתמש בפרוטוקול הבא:
   1. בחרו משתמש u באקראי.
   2. בחרו פריט i שהמשתמש דירג באקראי.
   3. חשבו תחזית לפריט i עבור משתמש u באמצעות כל אחד מהאלגוריתמים.
   4. חשבו את הטעות eu,i=ru,i-r̂u,i
   5. חזרו על החישוב למעלה מספר פעמים כנדרש וחשבו את הטעות הממוצעת
   6. ממשו את השיטה ComputeMAE המקבלת רשימה של שמות שיטות ומחזירה מילון ובו לכל שיטה רשום ה-MAE שלה.

הגשה: עליכם להגיש עבור תרגיל זה את קבצי הקוד בלבד (ללא קבצי מידע כגון ratings.dat). את תיקיית הפרויקט יש לכווץ באמצעות תוכנת zip ולהגיש לשרת <https://subsys.ise.bgu.ac.il/submission>. שם קובץ ה-zip צריך להכיל את מספרי תעודות הזהות של המגישים. ניתן להגיש מספר עבודות כרצונכם, אך הציון יחושב רק עבוד העבודה האחרונה.

את העבודות יש להגיש לא יאוחר מיום ראשון, ה-3 לאפריל ב-11:59 בלילה. הארכות ניתן לקבל בפניות אישיות.

את העבודה ניתן לעשות בזוגות. אל תעתיקו!

שאלות, בעיות, טענות ומענות – [shanigu@bgu.ac.il](mailto:shanigu@bgu.ac.il)

בהצלחה!