

בינה מלאכותית – דוח תרגיל 3:

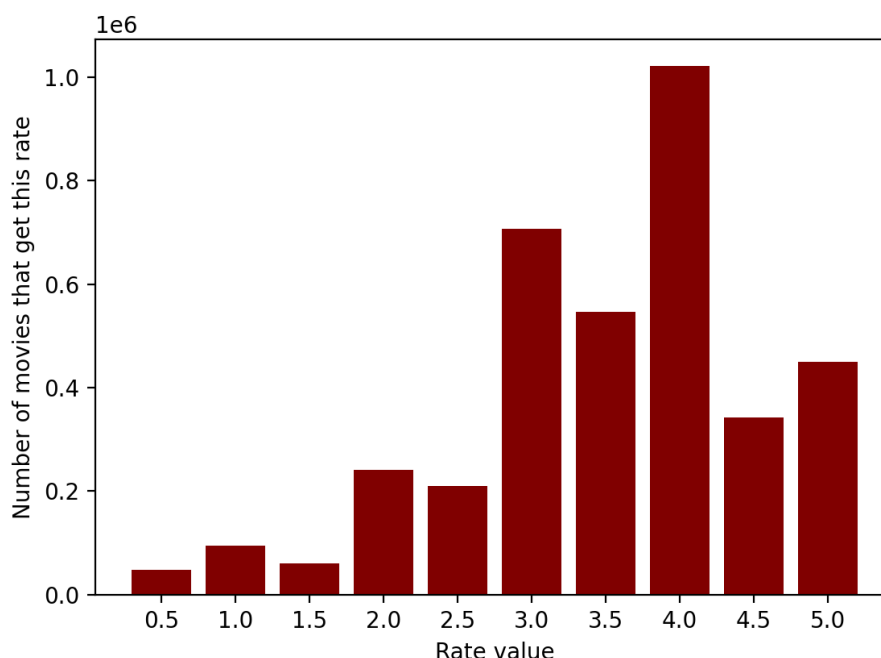
חלק ראשון – נתונים

(כל המימושים נמצאים בקובץ `data.py`)

1. מספר המשתמשים הייחודיים שדרגו את הסרטים – נבדוק בקובץ `ratings`, כמה `userId` ייחודיים קיימים בו. ונקבל: 21244.
מספר הסרטים הייחודיים – נבדוק בקובץ `ratings`, כמה `movieId` ייחודיים קיימים בו. ונקבל: 1342.
מספר הדירוגים ברובץ `ratings` – פשוט מספר השורות שבקובץ (ללא שורת הכותרות של העמודות). נקבל: 3720552.
2. מספר הדירוגים המינימלי שניתן עבור סרט כלשהו – 462.
מספר הדירוגים המקסימלי שניתן עבור סרט כלשהו – 9416.
3. מספר הדירוגים המינימלי שמשתמש כלשהו דירג – 46.
מספר הדירוגים המקסימלי שמשתמש כלשהו דירג – 432.

גרף –

נראה את התפלגות הדירוגים. כלומר, נציג בגרף כמה סרטים דורגו עבור כל ערך דירוג. נקבל את הגרף הבא:



הסקה על צורת ההתפלגות – כפי שניתן לראות, רוב הערכים מקבלים ערך גבוה, ולכן מדובר בהתפלגות א-סימטרית שלילית.

חלק שני - collaborating - filltering

4. נציג את התוצאות עבור משתמש "283225", נחלק את התוצאות ל:2:
- כאשר אנו במצב user based, כלומר ערך הדגלים שווה לTrue:

נציג את תוצאות החיזוי שהתקבלו עבור $k=5$:

(1995) Babe
(1993) Mrs. Doubtfire
(1993) Sleepless in Seattle
(1990) Pretty Woman
(1997) Titanic

נרצה לבדוק האם התוצאה שהתקבלה, הגיונית לאופי המשתמש.
לשם כך, נשווה את התוצאות עם תוצאות המשתמשים האחרים שדומים לו.

נבחר להסתכל על 22 המשתמשים הדומים (לפי מטריצת הדמיון שחישבנו). עבור משתמשים אלו – נרצה לראות אילו סרטים הם דירגו גבוה, בעוד המשתמש שלנו לא דירג אותם בכלל (כלומר, מופיע אצלו הערך Nan).
אם המשתמשים הדומים דרגו גבוה סרט מסוים שהמשתמש שלנו לא דירג -נצפה לראות סרטים אלו בחיזוי של המשתמש שלנו. כלומר, מערכת ההמלצה אמורה להמליץ לנו על סרטים שהמשתמשים הדומים אהבו (דרגו גבוה).

כעת נציג את חלק מהתוצאות של המשתמשים הדומים, וננתח אותם:

Similar Number 9: [('Pretty Woman (1990)', 'Comedy|Romance', 5.0)]

Similar Number 10: [('Titanic (1997)', 'Drama|Romance', 5.0)]

Similar Number 11: [('Babe (1995)', 'Children|Drama', 5.0)]

Similar Number 19: [('Mrs. Doubtfire (1993)', 'Comedy|Drama', 4.0)]

Similar Number 21: [('Sleepless in Seattle (1993)', 'Comedy|Drama|Romance', 5.0), ('Mrs. Doubtfire (1993)', 'Comedy|Drama', 4.0), ('Pretty Woman (1990)', 'Comedy|Romance', 5.0)]

כפי שניתן לראות, המשתמשים הדומים דירגו בדירוג גבוה את הסרטים שקיבלנו כחיזוי עבור המשתמש 283225.
(מספר הדמיון שהדפסתי מציין "כמה קרוב" המשתמש הנ"ל למשתמש 283225. כלומר, כמה הוא דומה לו).
כיוון שהסתכלתי על 22 משתמשים שהכי דומים – רוב החיזויים באמת כמעט זהים שהרי מדובר בדמיון ממש קרוב (כיוון שקיימת קבוצה גדולה של משתמשים).
מסקנה – ההמלצות תואמות לאופי המשתמש, כפי שניתן לראות בדירוג של המשתמשים שדומים לו.

5. הוספת משתמש פיקטיבי:

לפי בקשת התרגיל, נוסף משתמש פיקטיבי עם id 283238. אגדיר את הדירוגים שלו לסרטים השונים לפי הז'אנרים. את הז'אנרים Comedy ו-Action הוא מאוד אוהב, הם יקבלו דירוג מאוד גבוה (ניתן לראות את שאר דירוגי הז'אנרים בקוד שלי). הוספתי לratings 50 שורות של המשתמש הזה. הסרטים שהוספתי עבורו, קיבלו דירוג של ממוצע הדירוגים לפי הז'אנרים של אותו הסרט. כלומר, סכום דירוגי הז'אנרים (לפי מילון הדירוגים שעשיתי), חלקי מספר הז'אנרים שיש לאותו הסרט.

כעת, נראה את תוצאות החיזוי (שמות הסרטים והז'אנרים) עבור משתמש זה:

"Nutty Professor, The (1996)", Comedy|Fantasy|Romance|Sci-Fi
 Waterworld (1995), Action|Adventure|Sci-Fi
 Ace Ventura: Pet Detective (1994), Comedy
 Batman & Robin (1997), Action|Adventure|Fantasy|Thriller
 Batman Forever (1995), Action|Adventure|Comedy|Crime

כפי שציפינו, החיזויים שקיבלנו הם מהז'אנרים Comedy ו-Action בעיקר. כעת, נראה מה הדירוגים שדירגו המשתמשים הדומים (כפי שראינו בשאלה הקודמת):

Similar Number 4: [('Waterworld (1995)', 'Action|Adventure|Sci-Fi', 3.5), ('Batman & Robin (1997)', 'Action|Adventure|Fantasy|Thriller', 3.0)]

Similar Number 13: [('Batman Forever (1995)', 'Action|Adventure|Comedy|Crime', 4.0), ('Waterworld (1995)', 'Action|Adventure|Sci-Fi', 3.0)]

Similar Number 12: [('Ace Ventura: Pet Detective (1994)', 'Comedy', 5.0), ('Waterworld (1995)', 'Action|Adventure|Sci-Fi', 4.5)]

Similar Number 17: [('Batman Forever (1995)', 'Action|Adventure|Comedy|Crime', 2.0), ('Waterworld (1995)', 'Action|Adventure|Sci-Fi', 5.0)]

Similar Number 19: [('Nutty Professor, The (1996)', 'Comedy|Fantasy|Romance|Sci-Fi', 4.0), ('Batman & Robin (1997)', 'Action|Adventure|Fantasy|Thriller', 2.0)]

כפי שניתן לראות, גם המשתמשים הדומים דירגו את הסרטים הללו בדירוגים גבוהים.

.6

	precision@10	ARHR	RMSE
User based	0.24671	0.79846	0.89039
Item based	0.22940	0.72531	0.88815