

בינה מלאכותית תרגיל 3

1.

number of users are : 5000

number of products ranked are : 500

number of ranking are: 399664

2.

minimum number of ratings given to a product : 737

maximum number of ratings given to a product : 883

3.

minimum number of products ratings by user : 53

maximum number of products ratings by user : 114

4. א. מערכת המלצות מסוג CF בעצם לקוחת את הדירוגים הקיימים ומסתמכת עליהם בבניית ההמלצות שלה. ישנם מקרים בהם יש מוצרים או משתמשים בעלי דירוגים מועטים או ללא דירוגים כלל ומקרה כזה מערכת המלצות מסוג CF תתקשה מכיוון שלא יהיו לה מספיק דירוגים קיימים להסתמך עליהם בבניית ההמלצות שלה, והיא תצטרך לבצע כל מיני הנחות או יוריסטיקות על מנת לפצות על כך ולהשלים את הדירוגים החסרים או לנסות בעזרת יוריסטיקות מסוימות להעריך דירוגים של משתמשים שבכלל לא דירגו. בכך למעשה מערכת ההמלצות עשויה להיות פחות מדויקת מאשר אם היו הרבה דירוגים קיימים ומעט כאלה שמערכת ההמלצות צריכה לחזות. בעצם במערכת המלצות מסוג CF המערכת מחשבת את מטריצת הדמיון באמצעות מרחקים בזוגות (למשל בתרגיל שלנו בין משתמשים או בין מוצרים) ובחישוב זה עבור מוצרים שלא דורגו כלל על המערכת להשלים ערכים לדירוגים אלו (אצלנו בתרגיל למשל שמנו אפסים) ומילוי זה פוגע בדיוק מערכת ההמלצות שכן ככל שיש יותר מוצרים שלא דורגו כלל כך למערכת יש יותר אופציות לבחור ערכים (בעזרת יוריסטיקות כמו ממוצע של משתמשים/מוצרים וכו') לדירוגים אלו ויש לה יותר מקום לטעות. בעיה זו נקראת בעיית cold start. ניתן להתמודד יותר טוב עם בעיה זו למשל ע"י שימוש במערכות היברידיות שמשלבות גם מערכת המלצות מסוג CF וגם מערכות מבוססות תוכן, כלומר עם ידע מקדים על משתמש/מוצר שעוזרים במצבים של cold start שכן אנו יכולים להעריך יותר טוב בעזרת ידע מקדים על המוצר/משתמש (כמו למשל גיל, אזור מגורים, סוגי מוצרים וכו') את הדירוגים שחסרים לנו. נציין שישנם עוד רבים לטפל בבעיית cold start ולפעמים נצטרך שילוב של כמה מהם על מנת להגיע למערכת המלצות יותר מדויקת.

4. ב. כדי לטפל במוצרים חדשים, או מוצרים עליהם אנו מעוניינים להמליץ למרות מיעוט נתוני הדירוג נשתמש במערכות המלצה מבוססות תוכן או היברידיות (שמשלבות גם מערכות המלצה מסוג CF וגם מערכות המלצה מבוססות תוכן). מערכות המלצה שמבוססות תוכן למעשה ממליצה למשתמש על פריטים שדומים (לפי סוג הפריט וכו') שלפריטים שאהב, לדוגמה מערכת ההמלצה יכולה לקחת פריט חדש ובעזרת כל מיני תכונות של הפריט שיש למערכת עם קבלת הפריט החדש היא יכולה להמליץ עליו למשתמשים שאהבו פריטים דומים לפי התכונות שהגדרנו. ובעצם מערכות המלצה היברידיות למעשה משלבות את מערכות ההמלצה מבוססות התוכן (על מנת לפתור בעיות כמו cold start למשל) ובנוסף מערכות המלצה אשר מבוססות CF אשר חוזות המלצות ע"י שימוש בדירוגים קיימים על מנת לחזות דירוגים של אותם משתמשים/על אותם מוצרים. ולמעשה מערכת המלצה היברידית יכולה לקחת מוצר שיש לו מעט דירוגים ובאמצעות שילוב של דמיון למוצרים אחרים לפי תכונות ותוכן שידוע על המוצר ושל הדירוגים הקיימים כבר המערכת תבצע שכלול שמבוסס על שני אלו על מנת לחזות דירוגים למוצר זה וע"י כך להמליץ למשתמשים על מוצר זה לפי החיזוי.

עבור משתמשים חדשים, להם אין דירוגים כלל נשתמש במערכות המלצה מבוססות דמוגרפיה ותכונות של משתמשים או היברידיות (שמשלבות גם מערכות המלצה מסוג CF וגם מערכות

המלצה מבוססות דמוגרפיה ותכונות של משתמשים). מערכות המלצה שמבוססות דמוגרפיה ותכונות של משתמשים למעשה ממליצה למשתמש על פי מידע שיש לנו על המשתמש- גיל, אזור מגורים, תחומי עניין, השכלה וכו'. לדוגמה מערכת ההמלצה יכולה לקחת משתמש חדש ובעזרת מידע על המשתמש שיש למערכת עם קבלת המשתמש החדש היא יכולה להמליץ לו על פריטים שמשתמשים שדומים לו לפי התכונות והמידע עליו אהבו/ לא להמליץ על פריטים שהם לא אהבו. ובעצם מערכות המלצה היברידיות למעשה משלבות את מערכות ההמלצה אשר מבוססות דמוגרפיה ותכונות של משתמשים (על מנת לפתור בעיות כמו cold start למשל) ובנוסף מערכות המלצה אשר מבוססות CF אשר חוזות המלצות ע"י שימוש בדירוגים קיימים על מנת לחזות דירוגים של אותם משתמשים/ על אותם מוצרים. ולמעשה מערכת המלצה היברידית יכולה לקחת משתמש שדירג מעט פריטים ובאמצעות שילוב של דמיון למשתמשים אחרים לפי תכונות ומידע שידועים על המשתמש ושל הדירוגים הקיימים כבר המערכת תבצע שכלול שמבוסס על שני אלו על מנת לחזות דירוגים של המשתמש וע"י כך להמליץ לו על מוצרים. כמו כן ניתן ל כמו כן ניתן להשתמש במערכות המלצה מסוג non personalized אשר ממליצה המלצות גנריות שלא מותאמות בהכרח אישית וממליצה למשל על הפריטים הפופולאריים או הטרנדיים החדשים.

7. 5 המוצרים המומלצים עבור משתמש "AQWF3BBBDL4QJ" על פי user-user cf הם :

['B0001Y74VS', 'B0002B0R14', 'B0009EXONC', 'B0002JKQ8K', 'B0009PVV40']

8. 5 המוצרים המומלצים עבור משתמש "A3EO0WA7R3LVBQ" על פי item-item cf הם :

['B000052WYN', 'B000052YQ2', 'B000052ZSS', 'B0000534VO', 'B0000536M2']

9.

	RMSE
User-based CF	0.54751
Item-based CF	0.55211
Mean based (benchmark)	0.60155

10.

	Precision@20	Recall@20
User-based CF	0.04735	0.24685
Highest-ranked (benchmark)	0.04028	0.2128