

אחזור מידע (67782) - ניתוח תרגיל 3

מוגש על-ידי: מיכאל גרינבאום (211747639), גיא לוי (211744636)

19 ביוני 2021

1 Product Search

נתאר את הפונקציה:

- מוצאים את כל ה-reviews השונים עם דירוג ה-language model הכי גבוהים שמתאימים לשאלתא הנתונה, יכול להיות שיש חזרות על pid ולכן מהדירוג הגבוה לנמוך מחפשים את ה-k ה-pid הראשונים שמופיעים בדירוג זה.
- הוספנו פרמטר נוסף אותו בחרנו בעצמינו לתוך הפונקציה $\lambda \in [0, 1]$. נעבור על k ה-pid שמצאנו בשלב הקודם ונדרג את המוצר p על-ידי הפונקציה

$$\frac{1}{5 \cdot \# [\text{rev of } p]} \cdot \left[\sum_{\text{rev in } \{\text{rev's of } p\}} \left(\lambda \cdot \text{helpfulness}_{\text{rev}} + (1 - \lambda) \cdot \frac{1}{\# [\text{rev of } p]} \right) \cdot \text{score}_{\text{rev}} \right]$$

עכשיו נרצה להסביר את הרציונל מאחורי בחירת הפונקציה הנ"ל:

- בחירת השלב הראשון: דירוג ההתאמה של משוב לשאלתא בדרך כלל יעיד גם על הקשר בין המוצר לשאלתא ולכן נסתכל על k המוצרים עם הדירוג הכי גבוה ונמצא יחס סדר ביניהם (השלב השני).
- בחירת השלב השני: נתאר תכונות רצויות לפונקציה יחס הסדר ונסביר כיצד היא נכנסת בביטוי:

- רצינו לייצג את התוחלת ולכן כדי להימנע מ-bias של כמות המשובים נרמלנו בפקטור $\# [\text{rev of } p]$
- ערך ה-helpfulness יכול להיות 0 ולא רצינו לתת ניקוד כזה לכל משוב (זה יכול לנבוע מכך שאף אחד לא ניקד עד כמה שימושי המשוב) ולכן נתנו ערך דיפולטיבי של $\frac{1}{\# [\text{rev of } p]}$ ועשינו את המשקול $\lambda \cdot \text{helpfulness}_{\text{rev}} + (1 - \lambda) \cdot \frac{1}{\# [\text{rev of } p]}$ לכל משוב עבור $\lambda \in [0, 1]$ שנבחר על ידנו (בקוד בחרנו $\lambda = 0.5$).
- תכונות אינטואיטיביות שמתקיימות:

- ככל שה-helpfulness של משוב גדול יותר, הוא מגדיל את השפעת ה-score של המשוב על דירוג המוצר
- ככל שה-score של משוב גדול יותר ככה הוא מגדיל את ניקוד הדירוג של המוצר.
- ככל שיש יותר משובים, משקל הניקוד של כל משוב קטן (זה אינטואיטיבי כי מעבר ל-helpfulness אין לנו ידע על משוב בעל דעה אבסולוטית שמכריעה עבור כל הצרכנים ולכן יש אינטרס להקשיב יותר לדעת הרוב) (זה בא לידי ביטוי בביטוי $(1 - \lambda) \cdot \frac{1}{\# [\text{rev of } p]}$ שנהיה קטן יותר).