

מבחן קריטריון - חזון #1

חזון 1:

(a) $\text{threshold} = \frac{3}{10}$ (יש 10 איטמים).

C_i : מכלול האיטמים P באיטמים i .

L_i : האיטמים Frequent Itemsets (כאלו שיש בהם לפחות 3 איטמים).

$$C_1 = \{ \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\} \}$$

$$L_1 = \{ a, b, c, d, e \}$$

$$C_2 = \{ \{a,b\}, \{a,c\}, \{a,d\}, \{a,e\}, \{b,c\}, \{b,d\}, \{b,e\}, \{c,d\}, \{c,e\}, \{d,e\} \}$$

$$L_2 = \{ \{a,d\}, \{b,c\}, \{b,d\}, \{b,e\}, \{c,d\}, \{c,e\}, \{d,e\} \}$$

$$C_3 = \{ \{b,c,d\}, \{b,c,e\}, \{b,d,e\}, \{c,d,e\} \}$$

$$L_3 = \{ \{b,c,d\}, \{c,d,e\} \}$$

$$\Rightarrow L = \bigcup_{i=1}^3 L_i \quad \text{האיטמים Frequent Itemsets הם:}$$

(b) חוק $X \rightarrow Y$ של $\text{confidence} = 100\%$ מראה כי X הוא תמיד אמת.

המשמעות, למשל, חוק $A \rightarrow B$ של $\text{confidence} = 99\%$.

confidence מראה כי X הוא תמיד אמת כאשר Y הוא אמת.

מטרת המידה היא מדידת X של $\text{confidence} = 100\%$.

מראה כי X הוא תמיד אמת כאשר Y הוא אמת (אז X הוא תמיד אמת).

אם X הוא תמיד אמת כאשר Y הוא אמת, אז confidence הוא תמיד אמת.

אם X הוא תמיד אמת כאשר Y הוא אמת, אז confidence הוא תמיד אמת.

ואכן אם confidence של $X \rightarrow Y$ הוא תמיד אמת לא 100% .

אם X הוא תמיד אמת כאשר Y הוא אמת, אז confidence הוא תמיד אמת.

אם X הוא תמיד אמת כאשר Y הוא אמת, אז confidence הוא תמיד אמת.

(ואם Y הוא תמיד אמת כאשר X הוא אמת, אז confidence הוא תמיד אמת).

למשל $X \rightarrow Y$ הוא תמיד אמת כאשר X הוא אמת, אז confidence הוא תמיד אמת.

1. נראה כי X הוא תמיד אמת כאשר Y הוא אמת, אז confidence הוא תמיד אמת.

הוא תמיד אמת כאשר X הוא אמת, אז confidence הוא תמיד אמת.

שאלה 2:

א) נניח כי $R: X \rightarrow Y$ היא רשת חוק, $x \in X$ ו- $y \in Y$ הם נקודות חוק. נגדיר $\phi = x \cap Y$ ו- $L_z = \text{lcm}(\{z\})$.

$$L_z = \text{lcm}(\{z\})$$

• נקודת חוק, P .

$$\text{confidence}(x \rightarrow y) = \frac{\text{Support}(XY)}{\text{Support}(X)} = -$$

$$\frac{\#(\{L_{xy} \cdot i \mid 1 \leq i \leq \lfloor \frac{100}{L_{xy}} \rfloor\})}{\#(\{L_x \cdot i \mid 1 \leq i \leq \lfloor \frac{100}{L_x} \rfloor\})} = \frac{\lfloor \frac{100}{L_{xy}} \rfloor}{\lfloor \frac{100}{L_x} \rfloor}$$

• כיוון P נקודת חוק, $x \in X$ ו- $y \in Y$ הם נקודות חוק, $L_{xy} = L_x$ ו- $L_{xy} = L_x$.

• $\text{confidence}(x \rightarrow y) = 1$ נקודת חוק, $x \in X$ ו- $y \in Y$ הם נקודות חוק.

* נניח $X = \emptyset$ ו- $Y = \emptyset$ ו- $X \cup Y = \emptyset$.

• $\text{threshold} = \frac{5}{100}$ (ע. 100 סופים) (ב)

- frequent Items: $X = \{x\}$ (נקודת חוק, $x \in [100]$)

• P היא נקודת חוק

$$\text{Support}(X) = \text{Support}(\{x\}) = \lfloor \frac{100}{L_x} \rfloor = \lfloor \frac{100}{x} \rfloor \geq 5$$

• $x \leq \frac{100}{5} = 20$ ו- $x \in [20]$ היא נקודת חוק.

- frequent Pairs: $X = \{x, y\}$ (נקודת חוק, $x, y \in [100]$)

• P היא נקודת חוק

$$\text{Support}(X) = \text{Support}(\{x, y\}) = \lfloor \frac{100}{L_{xy}} \rfloor \geq 5$$

• $x \leq 20$ ו- $y \leq 20$ ו- $\text{lcm}(x, y) \leq 20$ היא נקודת חוק.

• $y \leq 20$ ו- $y \in [\frac{20}{x}]$ היא נקודת חוק.

• נקודת חוק / נקודת חוק

©

$$1) \text{Conf}(\{5,7\} \rightarrow 2) = \frac{\text{Support}(\{5,7,2\})}{\text{Support}(\{5,7\})} = \frac{\lfloor \frac{100}{70} \rfloor}{\lfloor \frac{100}{35} \rfloor} = \frac{1}{2}$$

$$2) \text{Conf}(\{2,3,4\} \rightarrow 7) = \frac{\text{Support}(\{2,3,4,7\})}{\text{Support}(\{2,3,4\})} = \frac{\lfloor \frac{100}{84} \rfloor}{\lfloor \frac{100}{12} \rfloor} = \frac{1}{8}$$

3 שלב

נניח שפרטים מחולקים: כלים בולים, ומהחומר
 (ל מופיע בול i בול j, כל $\frac{1}{2}$ בול i בול j)
 : פרט R: $X \rightarrow Y$

$$\begin{aligned} \text{Interest}(X \rightarrow Y) &= \text{Conf}(X \rightarrow Y) - \text{Support}(Y) = \\ &= \frac{\text{Support}(XY)}{\text{Support}(X)} - \text{Support}(Y) = \star \end{aligned}$$

(במקום שבו פרטים מופיע בול i היה $\frac{1}{2}$ (הפרט))
 ולכן $\text{Support}(Y) = \frac{1}{2}$ כיוון ש, כל בול X הופיע $x \in X$ (פרט)
 . $\text{Support}(X) = \text{Support}(\{x\}_{x \in X}) = \prod_{x \in X} \frac{1}{x}$: כל הכלים וכל הכלים
 כיוון ש, כל בול i בול j, X וכל Y $\text{Support}(XY) = \text{Support}(X) \text{Support}(Y)$ ולכן

$$\star = \frac{(\prod_{x \in X} \frac{1}{x}) (\prod_{y \in Y} \frac{1}{y})}{\prod_{x \in X} \frac{1}{x}} - \prod_{y \in Y} \frac{1}{y} =$$

$$= \prod_{y \in Y} \frac{1}{y} - \prod_{y \in Y} \frac{1}{y} = 0$$

פרט R: $X \rightarrow Y$ כל הכלים, פרט R: $X \rightarrow Y$ כל הכלים
 : פרט $\text{Interest}(X \rightarrow Y) = 0$

☒

: 4 תשובות

Jaccard similarity.

(a)

$$J(A,B) = \frac{4}{8}$$

$$J(B,C) = \frac{4}{8}$$

$$J(A,C) = \frac{4}{8}$$

$$J(B,D) = \frac{4}{8}$$

$$J(A,D) = \frac{6}{8}$$

$$J(C,D) = \frac{4}{8}$$

Cosine similarity

(b)

$$(*) \|A\| = \sqrt{80}, \|B\| = \sqrt{55}, \|C\| = \sqrt{44}, \|D\| = \sqrt{85}$$

$$\cos(A,B) = \frac{42}{\sqrt{80 \cdot 55}} \approx 0.633$$

$$\cos(A,C) = \frac{39}{\sqrt{80 \cdot 44}} \approx 0.657$$

$$\cos(A,D) = \frac{81}{\sqrt{80 \cdot 85}} \approx 0.982$$

$$\cos(B,C) = \frac{30}{\sqrt{55 \cdot 44}} \approx 0.609$$

$$\cos(B,D) = \frac{41}{\sqrt{55 \cdot 85}} \approx 0.6$$

$$\cos(C,D) = \frac{46}{\sqrt{44 \cdot 85}} \approx 0.752$$

Centered Cosine Similarity

(c)

$$\cos(A,B) \approx 0.658$$

$$\cos(B,C) \approx 0.013$$

$$\cos(A,C) \approx -0.808$$

$$\cos(B,D) \approx 0.602$$

$$\cos(A,D) \approx 0.888$$

$$\cos(C,D) \approx -0.058$$

יש להיזהר, כי A, B, C, D אינם נמצאים באותו מרחב וקטורי (d)

(הערות/הבהרות: A, B הם וקטורים במרחב 2D, C, D הם וקטורים במרחב 3D)

במרחב 2D: A, B הם וקטורים במרחב 2D, C, D הם וקטורים במרחב 3D

אם A, B הם וקטורים במרחב 2D, C, D הם וקטורים במרחב 3D

$$A \cdot B = \sum_{i=1}^{n+1} A_i B_i = \sum_{i=1}^n A_i B_i + \underbrace{A_{n+1} B_{n+1}}_0 = \sum_{i=1}^n A_i B_i$$

$$\|A\|^2 = \sum_{i=1}^{n+1} A_i^2 = \sum_{i=1}^n A_i^2 + \underbrace{A_{n+1}^2}_0 = \sum_{i=1}^n A_i^2$$

לכן $\cos(A,B)$ הוא זהה לזה במרחב 2D, A, B הם וקטורים במרחב 2D

כמו כן ρ Centered cosine similarity לא נחשב:
 הסיבה היא שישנם מקרים בהם המכנה
 מתאפס. למשל, אם המכנה הוא 0, אז
 המכונה ρ הסימילריות הסימטרית, שכן
 המכונה ρ הסימילריות הסימטרית, שכן
 המכונה ρ הסימילריות הסימטרית, שכן
 המכונה ρ הסימילריות הסימטרית, שכן

e) Prediction of user A's rating of item f
 (with user-user collaborative filtering)
 (INI=2, predicted score weighted by similarity)

$$P_{A,f} = \text{Sim}(x,y) \cdot S_{x,y}$$

הסימילריות הסימטרית, שכן

הסימילריות הסימטרית, שכן

$$\textcircled{*} r_{x_f} = \frac{S_{AB} \cdot r_{B_f} + S_{AC} \cdot r_{C_f}}{S_{AB} + S_{AC}} = \frac{0.658 \times 2 - 0.208}{0.658 - 0.208} = 2.462$$

f) הסימילריות הסימטרית: $(x_1, y_1), \dots, (x_n, y_n)$ ויש להם ρ הסימילריות הסימטרית, שכן

$$R := \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} = \star$$

הסימילריות הסימטרית, שכן

הסימילריות הסימטרית, שכן

הסימילריות הסימטרית, שכן

הסימילריות הסימטרית, שכן

$$\text{Centered Cosine Similarity}(x,y) = \frac{(x-\bar{x})(y-\bar{y})}{\|x-\bar{x}\| \|y-\bar{y}\|} =$$

$$= \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$