

Übungsblatt 1 Recommendersysteme

Markus Bilz – Matr. Nr. 2327197

11. Juni 2020

- a. Die naive Prognose entspricht dem arithmetischen Mittel der abgegebenen Bewertungen des Nutzers (Geyer-Schulz & Sonnenbichler, 2019). Für Lina gilt damit:

$$\tilde{x}_L = \frac{1}{n} \sum s_k = \frac{4 + 4 + 1 + 5 + 2}{5} = \frac{16}{5},$$

wobei s_k das k -te Rating der betrachteten Nutzerin ist.

- b. Um die Abhängigkeit zwischen zwei Benutzern zu messen, kann der Pearson'sche Korrelationskoeffizient verwendet werden.
- c. Zunächst sind die Mittel der gemeinsamen Profile zu bestimmen:

$$\bar{x}_F = \frac{2 + 2 + 4 + 1 + 5}{5} = \frac{14}{5}$$

und

$$\bar{x}_J = \frac{5 + 4 + 2 + 1}{4} = 3,$$

woraus später dann der Korrelationskoeffizient bestimmt werden kann. Weiterhin sind die lokalen Mittelwerte für Lina notwendig. Diese sind:

$$\bar{x}_L = \frac{4 + 4 + 1 + 5 + 2}{5} = \frac{16}{5},$$

und

$$\bar{x}_L = \frac{4 + 4 + 1 + 5}{4} = \frac{7}{2}.$$

Nach Geyer-Schulz und Sonnenbichler (2019) gilt für den Korrelationskoeffizient:

$$r_{kl} = \frac{\text{cov}(s_k, s_l)}{\sigma_k \sigma_l} = \frac{\sum_i (x_{i,k} - \bar{x}_k)(x_{i,l} - \bar{x}_l)}{\sqrt{\sum_i (x_{i,k} - \bar{x}_k)^2} \sqrt{\sum_i (x_{i,l} - \bar{x}_l)^2}}.$$

Damit folgt:

$$\begin{aligned} r_{L,F} &= \frac{2(4 - \frac{16}{5})(2 - \frac{14}{5}) + (1 - \frac{16}{5})(4 - \frac{14}{5}) + (5 - \frac{16}{5})(1 - \frac{14}{5}) + (2 - \frac{16}{5})(5 - \frac{14}{5})}{\sqrt{2(4 - \frac{16}{5})^2 + (1 - \frac{16}{5})^2 + (5 - \frac{16}{5})^2 + (2 - \frac{16}{5})^2} \sqrt{2(2 - \frac{14}{5})^2 + (4 - \frac{14}{5})^2 + (1 - \frac{14}{5})^2 + (5 - \frac{14}{5})^2}} \\ &= -\frac{49}{54}, \end{aligned}$$

$$\begin{aligned} r_{L,J} &= \frac{(4 - \frac{7}{2})(5 - 3) + (4 - \frac{7}{2})(4 - 3) + (1 - \frac{7}{2})(2 - 3) + (5 - \frac{7}{2})(1 - 3)}{\sqrt{2(4 - \frac{7}{2})^2 + (1 - \frac{7}{2})^2 + (5 - \frac{7}{2})^2} \sqrt{(5 - 3)^2 + (4 - 3)^2 + (2 - 3)^2 + (1 - 3)^2}} \\ &= \frac{1}{3\sqrt{10}}. \end{aligned}$$

Damit entspricht die Korrelation zwischen Lina und Felix $-\frac{49}{54}$ und zwischen Lina und Johannes $\frac{1}{3\sqrt{10}}$.

- d. Der Korrelationskoeffizient nimmt einen Wert der zwischen -1 und 1 an, wobei 1 eine vollständige positive Korrelation indiziert. Die hohe negative Korrelation zwischen Lina und Felix lässt darauf schließen, dass die beiden häufig gegensätzlich bewerten. Umso besser Lina bewertet, umso schlechter bewertet Johannes und vice versa.

Da der Korrelationskoeffizient für Lina und Johannes nahe Null liegt, lässt sich schließen, dass kein bzw. sehr schwacher Zusammenhang zwischen den Bewertungen von Lina und Johannes besteht.

- e. Zunächst ist der Korrelationskoeffizient zwischen Hannah und Lina zu bestimmen.

Der lokale Mittelwert für Hannah ist:

$$\bar{x}_H = \frac{1 + 4 + 2 + 4}{4} = \frac{11}{4}.$$

Der lokale Mittelwert für Lina ist:

$$\bar{x}_L = \frac{4 + 1 + 5 + 2}{4} = 3.$$

Der Korrelationskoeffizient zwischen beiden ist:

$$\begin{aligned} r_{L,H} &= \frac{(4-3)\left(1-\frac{11}{4}\right) + (1-3)\left(4-\frac{11}{4}\right) + (5-3)\left(2-\frac{11}{4}\right) + (2-3)\left(4-\frac{11}{4}\right)}{\sqrt{(4-3)^2 + (1-3)^2 + (5-3)^2 + (2-3)^2} \sqrt{\left(1-\frac{11}{4}\right)^2 + 2\left(4-\frac{11}{4}\right)^2 + \left(2-\frac{11}{4}\right)^2}} \\ &= -\frac{7\sqrt{30}}{45}. \end{aligned}$$

Es gilt nach Geyer-Schulz und Sonnenbichler (2019):

$$x_{1,L} = \tilde{x}_L + \frac{(x_{1,H} - \bar{x}_H) r_{L,H} + (x_{1,F} - \bar{x}_F) r_{L,F} + (x_{1,J} - \bar{x}_J) r_{L,J}}{|r_{L,H}| + |r_{L,F}| + |r_{L,J}|}.$$

Damit folgt eine verbesserte Prognose unter Berücksichtigung aller Bewerter von:

$$x_{1,L} = \frac{16}{5} + \frac{\left(2-\frac{11}{4}\right)\left(-\frac{7\sqrt{30}}{45}\right) + \left(3-\frac{14}{5}\right)\left(-\frac{49}{54}\right) + (5-3)\frac{1}{3\sqrt{10}}}{\left|-\frac{7\sqrt{30}}{45}\right| + \left|-\frac{49}{54}\right| + \left|\frac{1}{3\sqrt{10}}\right|} = 3.558 \approx 4.$$

Lina würde den Roman Rabenfrauen wahrscheinlich mit 4 Punkten bewerten.

Literatur

Geyer-Schulz, A. & Sonnenbichler, A. (2019). Recommendersysteme.