Predicting Ratings with Neighborhood-Based Methods

Florian Orpelière

June 27, 2017

1 Formules de bases

Posons les éléments suivants :

- La moyenne des notes d'un utilisateur u est μ_u. I_u représente l'ensemble des items que l'utilsateur u a noté :

  $$\mu_u = \frac{\sum_{k \in I_u} r_{uk}}{|I_u|}$$

- La similarit de Pearson (général):

  $$Sim(x, y) = Pearson(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

  Où $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ est la moyenne des x.
  On pourrait en utiliser d'autres. Celle-ci la proprit d'avoir des valeurs comprises entre -1 et 1.

  $$s_{uj} = r_{uj} - \mu_u$$

  Où r_{u_j} est la note d'un utilisateur u sur l'item j.

2 Neighborhood Models basé sur les utilisateurs

On commence par calculé la similarité entre deux utilisateurs u et v.

$$Sim(u,v)$$
$$= Pearson(u,v) = \frac{\sum_{k\in I_u\cap I_v}(r_{uk}-\mu_u).(r_{vk}-\mu_v)}{\sqrt{\sum_{k\in I_u\cap I_v}(r_{uk}-\mu_u)^2}\sqrt{\sum_{k\in I_u\cap I_v}(r_{vk}-\mu_v)^2}}$$
$$= \frac{\sum_{k\in I_u\cap I_v} s_{uk}.s_{vk}}{\sqrt{\sum_{k\in I_u\cap I_v} s_{uk}^2}\sqrt{\sum_{k\in I_u\cap I_v} s_{vk}^2}}$$

Avec $\sum_{k\in I_u\cap I_v}(r_{uk}-\mu_u).(r_{vk}-\mu_v)$ qui est le produit scalaire des vecteurs r_u et r_v minoré par leur moyenne μ_u et μ_v respectivement. Donc lle produit scalaire des notes d'un utilisateurs u et v centrées sur la moyenne.

Puis, $\sqrt{\sum_{k\in I_u\cap I_v}(r_{v_u}-\mu_u)^2}$ et $\sqrt{\sum_{k\in I_u\cap I_v}(r_{v_k}-\mu_v)^2}$ qui est la norme des vecteurs r_u et r_v minoré par leur moyenne μ_u et μ_v.

Nous pouvons prédire la note d'un utilisateur u sur l'item j :

$$r_{uj} = \mu_u + \frac{\sum_{v\in P_u(j)} Sim(u,v).s_{vj}}{\sum_{v\in P_u(j)} |Sim(u,v)|}$$