Matrix factorization models

Each item i is associated with a vector q_i, and each user u is associated with a vector p_u. For a given item i, the elements of q_i measure the extent to which the item possesses those factors, positive or negative. For a given user u, the elements of p_u measure the extent of interest the user has in items that are high on the corresponding factors (again, these may be positive or negative). The resulting dot product, $q_i^T p_u$, captures the interaction between user u and item i—i.e., the overall interest of the user in characteristics of the item. The final rating is created by also adding in the aforementioned baseline predictors that depend only on the user or item. Thus, a rating is predicted by the rule

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T p_u$$

In order to learn the model parameters, we minimize the regularized squared error

$$\min_{b^*,q^*,p^*} \sum \left(r_{ui} - \mu - b_i - b_u - q_i^T p_u\right)^2 + \lambda_4 \left(b_i^2 + b_u^2 + \|q_i\|^2 + \|p_u\|^2\right)$$

The constant λ_4 which controls the extent of regularization, is usually determined by cross validation. Minimization is typically performed by either stochastic gradient descent or alternating least squares.

Several types of implicit feedback can be simultaneously introduced into the model by using extra sets of item factors. For example, if a user u has a certain kind of implicit preference to the items in $N_1(u)$ (e.g., she rented them), and a different type of implicit feedback to the items in $N_2(u)$ (e.g., she browsed them), we could use the model

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T \left(p_u + |N_1(u)|^{-\frac{1}{2}} \sum y_j^{(1)} + |N_2(u)|^{-\frac{1}{2}} \sum y_j^{(2)}\right)$$

with the factor vectors $y_j^{(k)}$ which represents the perspective of implicit feedback. Since the y_j are centered around zero (by the regularization), the sum is normalized by $|R(u)|^{-\frac{1}{2}}$, in order to stabilize its variance across the range of observed values of $|R(u)|$.