

Многокритериальная оптимизация в задаче персональных рекомендаций

Кислинский Вадим Геннадьевич

Московский физико-технический институт

Физтех-Школа Прикладной математики и Информатики

Кафедра интеллектуальных систем

Научный руководитель: д.ф.-м.н. Воронцов К.В.

16 июня 2021

Обозначения и исходные данные:

- \mathcal{U}, \mathcal{I} — множества пользователей и объектов;
- $\mathcal{D} = \{(u_t, i_t)\}_{i=1}^T \in \mathcal{U} \times \mathcal{I}$ — транзакционные данные
- $R = (r_{ui})_{\mathcal{U} \times \mathcal{I}}$ — матрица отношений $r_{ui} \in \{0, 1\}$ или $r_{ui} \in \{1, \dots, M\}$

Требуется:

- $\hat{r}_{ui} = p_u^T q_i$ — построить модель
- $\mathcal{L}_u = \underset{i \in \mathcal{I}}{\operatorname{argtop}_k} p_u^T q_i$ — выдать пользователю отранжированный список персональных рекомендаций

Проблема:

- Что делает рекомендации персональными, полезными для пользователя?

Точность предсказания отношений:

$$MSE = \frac{1}{|\mathcal{D}|} \sum_{(u,i) \in \mathcal{D}} (r_{ui} - \hat{r}_{ui})^2 \rightarrow \min$$

Качество ранжирования списков \mathcal{L}_u :

- Без учета порядка:

$$precision@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{\sum_{i \in \mathcal{L}_u^K} r_{ui}}{K} \rightarrow \max,$$

где \mathcal{L}_u^K – первые K элементов в списке \mathcal{L}_u

- С учетом порядка:

$$map@k = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{K} \sum_{k=1}^K \frac{\sum_{i \in \mathcal{L}_u^k} r_{ui}}{k} r_{uk} \rightarrow \max$$

Качество разнообразия списков \mathcal{L}_u :

$$MDIV@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{K(K-1)} \sum_{i \in \mathcal{L}_u^K} \sum_{j \in \mathcal{L}_u^K \setminus \{i\}} dist(i, j) \rightarrow \max,$$

где $dist(i, j)$ – расстояние между объектами i, j .

Качество новизны списков \mathcal{L}_u :

$$MNOV@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{K} \sum_{i \in \mathcal{L}_u^K} -\ln \frac{\sum_{u \in \mathcal{U}} r_{ui}}{|\mathcal{U}|} \rightarrow \max$$

Качество прозорливости рекомендательной системы:

$$MSER@K = \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} \frac{1}{|\mathcal{H}_u|} \frac{1}{K} \sum_{i \in \mathcal{L}_u^K} \sum_{j \in \mathcal{H}_u} dist(i, j) r_{ui} \rightarrow \max,$$

где \mathcal{H}_u – история пользователя u .

Качество покрытия множества \mathcal{I} :

$$COV@K = \frac{|\cup_{u \in \mathcal{U}} \mathcal{L}_u^K|}{|\mathcal{I}|} \rightarrow \max$$

Матрица отношений, как произведение двух матриц:

$$P = [p_u]_{u \in \mathcal{U}}, Q = [q_i]_{i \in \mathcal{I}}$$

$$\hat{r}_{ui} = p_u^T q_i$$

$$P^*, Q^* = \underset{P, Q}{\operatorname{argmin}} L(P, Q)$$

Разложение функции потерь L :

$$L(P, Q) = \sum_{c \in \mathcal{C}} \alpha_c L_c(P, Q),$$

где \mathcal{C} – множество критериев, L_c – функция потерь, оптимизирующая критерий c , $\alpha_c > 0$ – вес критерия c .

Введем:

$$\mathcal{D}_S = \{(u, i, j) : i, j \in \mathcal{I}, u \in \mathcal{U}, (u, i) \in \mathcal{D}\}$$

Пусть $L(W)$ – функция потерь:

$$L(W) = \sum_{(u, i, j) \in \mathcal{D}_S} h(\hat{r}_{ui}(W) - \hat{r}_{uj}(W))$$

Тогда:

- $\frac{\partial h}{\partial w_k} = \frac{\partial h}{\partial \hat{r}_{ui}} \left(\frac{\partial \hat{r}_{ui}}{\partial w_k} - \frac{\partial \hat{r}_{uj}}{\partial w_k} \right)$ – оптимизирует парные критерии ранжирования
- $\frac{\partial h}{\partial w_k} = \frac{\partial h}{\partial \hat{r}_{ui}} \left(\frac{\partial \hat{r}_{ui}}{\partial w_k} - \frac{\partial \hat{r}_{uj}}{\partial w_k} \right) \Delta_{ij}$ – оптимизирует списочные критерии ранжирования, учитывающие порядок, Δ_{ij} – изменение списочного критерия при перестановке элементов i, j

Christopher J. C. Burges. "From RankNet to LambdaRank to LambdaMART: An Overview". в: (2010)

\mathcal{I}_u – список объектов, прикрепленных к пользователю u при обучении, отранжированный по $p_u^T q_i$,

c – критерий,

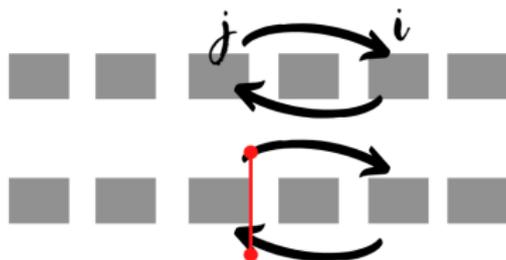
$\mathcal{I}_u(i, j)$ – список \mathcal{I}_u с перестановкой i, j

- для критериев, учитывающих порядок:

$$\Delta_{ij} = c(\mathcal{I}_u) - c(\mathcal{I}_u(i, j))$$

- для критериев, не учитывающих порядок

$$\Delta_{ij} = c(\mathcal{I}_u^{\min(i, j)}) - c(\mathcal{I}_u^{\min(i, j)}(i, j))$$



\mathcal{I}_u

Функция потерь:

$$L(P, Q) = \sum_{(u,i) \in \mathcal{D}} \sum_{\alpha} \alpha \ell_{\alpha} \alpha (p_u^T q_i, r_{ui}) + \\ \sum_{(u,i,j) \in \mathcal{D}_S} h(p_u, q_i, q_j, \beta_1, \beta_2, \dots) + \\ \lambda_P R_P(P) + \lambda_Q R_Q(Q),$$

где ℓ — поточечная функция потерь, h — функция потерь, оптимизирующая списочные критерии, R_P, R_Q — регуляризаторы

$$P^*, Q^* = \underset{P, Q}{\operatorname{argmin}} L(P, Q)$$

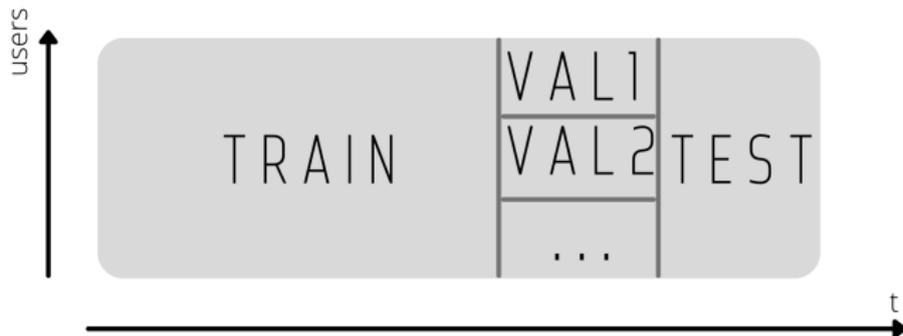
- Множество транзакций

$\mathcal{D} = \{(u, i, r, t) : u \in \mathcal{U}, i \in \mathcal{I}, r \in \overline{1, 5}, t \in \mathcal{T}\}$, \mathcal{U} – пользователи, \mathcal{I} – фильмы, \mathcal{T} – временные метки

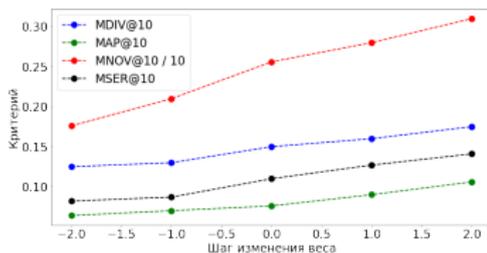
- Описания фильмов: категория, год выпуска

$$A_{ij} = \begin{cases} 1, & \text{фильм имеет категорию или год выпуска } i \\ 0, & \text{иначе} \end{cases}$$

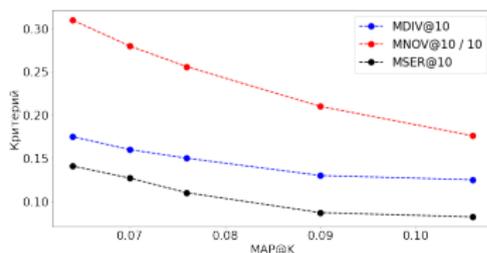
Стратегия валидации и оценки качества



Зависимость критериев от веса в общей функции потерь



Связь релевантности ($\text{map}@10$) с другими критериями



- предложенный подход позволяет повышать качество по более важным критериям
- необходимо искать оптимальный баланс между различными критериями качества, так как оптимизация отдельных приводит к понижению других

Weighted Matrix Factorization:

$$L = \sum_{(u,i) \in \mathcal{D}} C_{ui} (r_{ui} - p_u^T q_i)^2 + \lambda \|P\|_2^2 + \lambda \|Q\|_2^2$$

Bayesian Personalized Ranking:

$$L = \sum_{(u,i,j) \in \mathcal{D}_s} -\ln(\sigma(p_u^T q_i - p_u^T q_j)) + \lambda \|P\|_2^2 + \lambda \|Q\|_2^2$$

Proposed method:

$$L = \alpha \sum_{(u,i)} C_{ui} (r_{ui} - p_u^T q_i)^2 + \sum_{(u,i,j)} h(p_u, q_i, q_j, \beta_1, \beta_2, \dots) + \lambda \|P\|_2^2 + \lambda \|Q\|_2^2,$$

$$\frac{\partial h}{\partial w_k} = - \frac{\left(\frac{\partial \hat{r}_{ui}}{\partial w_k} - \frac{\partial \hat{r}_{uj}}{\partial w_k} \right)}{1 + e^{\hat{r}_{ui} - \hat{r}_{uj}}} (\beta_1 |\Delta_{ij}^{ndcg}| + \beta_2 |\Delta_{ij}^{div}| + \beta_3 |\Delta_{ij}^{nov}| + \beta_4 |\Delta_{ij}^{ser}|)$$

Результаты экспериментов на датасете Movielens-1M, $K = 10$

Алгоритм	<i>MAP</i>	<i>MDIV</i>	<i>MNOV</i>	<i>MSER</i>	<i>COV</i>
WMF	0.097	0.12	2.12	0.10	0.45
BPR	0.15	0.13	1.67	0.087	0.20
Proposed method	0.076	0.15	2.56	0.11	0.52

- предложенный подход показывает более высокие результаты по критериям разнообразия, новизны, прозорливости, покрытия, но более низкие по ранжированию
- использование различных критериев более полно раскрывает качество работы рекомендательной системы

- Предложен подход к совместной оптимизации нескольких критериев в задаче персональных рекомендаций
- Обобщен подход к оптимизации критериев ранжирования на списочные критерии, не учитывающие порядок