

## 1. Бајесово персонализовано рангирање (BPR)

Као што се може уочити из елаборације о предиктивним алгоритмима, циљ оптимизације је умањивање разлике између стварног рејтинга и његове предикције. Идеја алгоритма *Бајесово персонализовано рангирање* је потпуно другачија, јер подразумева оптимизацију рангирања производа, тако да се на врху листе за препоруку нађе производ са највећом вероватноћом да ће се допасти кориснику.

Овај алгоритам се примењује у системима за препоруку који раде са **имплицитним подацима**. Након формалног излагања о алгоритму биће наведен начин на који га је могуће применити и у системима који су засновани на експлицитним подацима.

Нека је  $U$  скуп свих корисника,  $I$  скуп свих производа, а  $S$  скуп уређених парова свих корисника и свих производа. Под **позитивним** производом се подразумева сваки производ са којим је корисник на изврстан начин интераговао. **Негативни** производ се дефинише као производ са којим корисник није имао интеракцију.

|       | $i_1$    | $i_2$ | $i_3$ | $i_4$ |                |
|-------|----------|-------|-------|-------|----------------|
| $u_1$ | ?        | +     | +     | ?     | user<br>↑<br>↓ |
| $u_2$ | +        | ?     | ?     | +     |                |
| $u_3$ | +        | +     | ?     | ?     |                |
| $u_4$ | ?        | ?     | +     | +     |                |
| $u_5$ | ?        | ?     | +     | ?     |                |
|       | ← item → |       |       |       |                |

Слика 1 – Интеракције

Уколико је корисник  $u$  интераговао са производом  $i$  ( $(u, i) \in S$ ), тада се каже да корисник  $u$  више преферира производ  $i$  у односу на остале производа са којима није имао интеракција. Посматрајући слику 1 уочава се да је корисник  $u_1$  интераговао са производом  $i_2$ , али не и са производом  $i_1$ , што се формално записује на следећи начин:

$$i_2 >_{u_1} i_1 \quad (1)$$

Уопштавањем претходног записа и увођењем термина позитивни и негативни производи долази се до:

$$i >_u j \quad (2)$$

где је  $i$  позитиван производ за корисника  $u$ , а  $j$  негативан производ.

У случају да су оба производа позитивна или негативна, није могуће успоставити релацију  $>_u$  између њих.

Нека је  $\theta$  вектор параметара модела на основу кога се врши персонализовано рангирање. **У наставку је објашњен BPR алгоритам где је коришћен класичан модел заснован на факторизацији матрице, који предикцију рачуна на следећи начин:**

$$\hat{r}_{ui} = x_u \cdot y_i^T \quad (3)$$

$x_u$  – векторска репрезентација корисника  $u$ ,  $y_i$  – вектор којим је представљен производ  $i$ .

Циљ BPR оптимизације је максимизовање вероватноће дате са:

$$p(\theta | i >_u j) \propto p(i >_u j | \theta) p(\theta) \quad (4)$$

Функција  $p(i >_u j | \theta)$  представља вероватноћу да се неком кориснику више допада производ  $i$  у односу на  $j$  и дефинише се следећом формулом:

$$p(i >_u j | \theta) = \sigma(\hat{r}_{uij}(\theta)), \quad (5)$$

где је  $\sigma$  сигмоидна функција дата са  $\sigma(x) = \frac{1}{1+e^{-x}}$ .

Израз  $\hat{r}_{uij}(\theta)$  се може написати као:

$$\hat{r}_{uij} = \hat{r}_{ui} - \hat{r}_{uj} \quad (6)$$

где су  $\hat{r}_{ui}, \hat{r}_{uj}$  предикције модела на који ће начин корисник  $u$  реаговати на производе  $i$  и  $j$ , респективно. Ради једноставнијег записа, из  $\hat{r}_{uij}(\theta)$  је изостављен аргумент  $\theta$ .

Циљна функција алгорита BPR се назива **BPR..Opt** и дата је са:

$$\begin{aligned}
 BPR..Opt &\stackrel{\text{def}}{=} \prod_{u,i,j} p(i >_u j | \theta) p(\theta) \stackrel{\text{def}}{=} \ln \left( \prod_{u,i,j} p(i >_u j | \theta) p(\theta) \right) \\
 &= \sum_{u,i,j} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) + \ln p(\theta) = \sum_{u,i,j} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) - \frac{\lambda_\theta}{2} \|\theta\|^2 \\
 &= \sum_{u,i,j} \ln \sigma(\hat{r}_{ui} - \hat{r}_{uj}) - \frac{\lambda_\theta}{2} \|x_u\|^2 - \frac{\lambda_\theta}{2} \|y_i\|^2 - \frac{\lambda_\theta}{2} \|y_j\|^2 \\
 &= \sum_{u,i,j} \ln \sigma(x_u \cdot y_i^T - x_u \cdot y_j^T) - \frac{\lambda_\theta}{2} \|x_u\|^2 - \frac{\lambda_\theta}{2} \|y_i\|^2 - \frac{\lambda_\theta}{2} \|y_j\|^2 \\
 &= \sum_{u,i,j} \ln \frac{1}{1 + e^{-(x_u \cdot y_i^T - x_u \cdot y_j^T)}} - \frac{\lambda_\theta}{2} \|x_u\|^2 - \frac{\lambda_\theta}{2} \|y_i\|^2 - \frac{\lambda_\theta}{2} \|y_j\|^2
 \end{aligned} \tag{7}$$

Будући да је производ вероватноћа вредност са великим бројем децимала, ради лакшег меморисања, а и због математичких особина природног логаритма, BPR-Opt се представља и својим логаритмованим обликом. Функција  $p(\theta)$  представља функцију густине нормалне расподеле са математичким очекивањем 0 за све параметре<sup>1</sup> модела и матрицом варијанси  $\Sigma = \lambda_\theta I$ . Дакле, за сваки параметар модела се претпоставља да има нормалну дистрибуцију  $\mathcal{N}$  са математичким очекивањем 0 ( $\mu = 0$ ) и јединичном варијансом ( $\Sigma = I$ , занемаримо  $\lambda_\theta$  за сада). У складу са наведеним, формула за мултиваријантну нормалну расподелу променљиве  $x$  гласи:

$$p(\theta) = \mathcal{N}(x | \mu, \Sigma) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} \cdot e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)} = \frac{1}{(2\pi)^{d/2}} \cdot e^{\left(-\frac{1}{2}\theta^T \theta\right)} \tag{8}$$

Једини израз који зависи од  $\theta$  је  $e^{\left(-\frac{1}{2}\theta^T \theta\right)}$ , што значи да се приликом оптимизације може изоставити константа  $\frac{1}{(2\pi)^{d/2}}$ . Ако се претходна једнакост логаритмује добија се:

$$\ln p(\theta) = \ln e^{\left(-\frac{1}{2}\theta^T \theta\right)} = -\frac{1}{2} \theta^T \theta = -\frac{1}{2} \|\theta\|^2 \tag{9}$$

Уврштавањем изостављеног елемента  $\lambda_\theta$  у претходну формулу добија се:

$$\ln p(\theta) = -\frac{\lambda_\theta}{2} \|\theta\|^2 \tag{10}$$

где је  $\lambda_\theta$  параметар регуларизације.

<sup>1</sup> Параметри овог модела су: вектор којим је представљен корисник, вектор којим је представљен први производ (*префериран* у односу на други) и векторска репрезентација другог производа. Сваки од наведених вектора представља једну колону матрице  $\theta$ . Будући да су вектори вишедимензионалне случајне променљиве могуће је говорити о њиховој расподели.

Циљ оптимизације се може написати и на следећи начин:

$$\operatorname{argmin}_{x_u, y_i, y_j} \sum_{u, i, j} -\ln \frac{1}{1 + e^{-(x_u \cdot y_i^T - x_u \cdot y_j^T)}} + \frac{\lambda_\theta}{2} \|x_u\|^2 + \frac{\lambda_\theta}{2} \|y_i\|^2 + \frac{\lambda_\theta}{2} \|y_j\|^2 \quad (11)$$

Због обимности тренинг скупа (све могуће тројке  $(u, i, j)$ ) оптимизација се спроводи стохастичким градијентним спутом. У свакој итерацији се насумично бира одређени број корисника који чине један пакет (*batch*). Наведена варијанта градијентног спуста у свакој пакетној итерацији за корисника  $u$  бира само једну уређену тројку  $(u, i, j)$  на основу које врши промену параметара.

$$\begin{aligned} \frac{\partial}{\partial x_u} &= \frac{1}{1 + e^{(x_u \cdot y_i^T - x_u \cdot y_j^T)}} \cdot (y_j - y_i) + \lambda x_u \\ \frac{\partial}{\partial y_i} &= \frac{1}{1 + e^{(x_u \cdot y_i^T - x_u \cdot y_j^T)}} \cdot -x_u + \lambda y_i \\ \frac{\partial}{\partial y_j} &= \frac{1}{1 + e^{(x_u \cdot y_i^T - x_u \cdot y_j^T)}} \cdot x_u + \lambda y_j \end{aligned} \quad (12)$$

Како би се овај алгоритам имплементирао у системе за препоруку који су засновани на експлицитним подацима, неопходно је те податке претворити у неку имплицитну форму. Први начин јесте да се у матрици корисности сваки рејтинг замени јединицом или нулом у зависности од тога да ли је већи од неке унапред задате вредности<sup>2</sup>. Под позитивним производом за корисника сматраће се сваки производ коме је он доделио оцену већу од унапред задате вредности. Сви остали производи спадају у категорију за њега негативних производа. Овај приступ је искључујући, јер није спроводљив за оне кориснике који ниједан производ нису оценили рејтингом већим од унапред задате границе која раздваја позитивне и негативне производе. Како су сви производи негативни за такве кориснике, следи да за њих и није могуће успоставити релацију  $i >_u j$ .

Други начин се огледа у „нормализацији“ матрице корисности заснованој на кориснику. Наиме, од сваког рејтинга у матрици корисности потребно је одузети просечну вредност оцена коју тај корисник оставља. У следећем кораку све вредности које су веће или једнаке нули заменити јединицом, а у преостала поља уписати нулу. На овај начин се за сваког корисника обавља „персонализовање“ раздвајања позитивних од негативних производа. Наиме, сваки производ се за корисника сматра позитивним уколико га је оценио бар нивоу свог просека. Сви остали, који не задовољавају наведени критеријум, се сматрају негативним производима.

Једна од имплементација BPR алгоритма је приказана на веб страници коју можете пронаћи на линку: [http://ethen8181.github.io/machine-learning/recsys/4\\_bpr.html](http://ethen8181.github.io/machine-learning/recsys/4_bpr.html).

<sup>2</sup> Ова вредност је иста за сваког корисника. Ако је скала од 1 до 5, сви производи који су оцењени рејтингом већим од 3 могу се сматрати, условно речено, позитивним за корисника.

### Додатна литература:

- Оригинални рад: [BPR: Bayesian Personalized Ranking from Implicit Feedback](#)
- Блог: <https://towardsdatascience.com/recommender-system-using-bayesian-personalized-ranking-d30e98bba0b9>

## 2. Завршни преглед градива и даље смернице за рад

### 2.1. Кластеризација у системима за препоруку

**Књига:** Charu C. Aggrawal, *Recommender systems: The Textbook*. Springer, 2016.

#### 2.4 Clustering and Neighborhood-based Methods

**Co-clustering у Surprise библиотеци:**

[https://surprise.readthedocs.io/en/stable/co\\_clustering.html#surprise.prediction\\_algorithms.co\\_clustering.CoClustering](https://surprise.readthedocs.io/en/stable/co_clustering.html#surprise.prediction_algorithms.co_clustering.CoClustering)

### 2.2. Правила асоцијације (Association rules)

**Књига:** Charu C. Aggrawal, *Recommender systems: The Textbook*. Springer, 2016.

#### 3.3 Rule-Based Collaborative Filtering

*Априори алгоритам...*

### 2.3. Хибридни системи за препоруку

| Метода хибридизације | Опис                                                                                                                                                                                                                                                                              |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Weighted             | Неколико система за препоруку се комбинују у један систем.                                                                                                                                                                                                                        |
| Switching            | Замена система за препоруку у зависности од ситуације.                                                                                                                                                                                                                            |
| Mixed-interleaving   | Препоруке из више система за препоруку се приказују у исто време.                                                                                                                                                                                                                 |
| Cascade              | Препоруке једног система пролазе кроз нови систем за препоруку као провера.                                                                                                                                                                                                       |
| Feature Combination  | There exist two very different recommendation components for this hybrid, contributing and actual recommender. The actual recommender works with data modified by the contributing one. The contributing one injects features of one source to the source of the other component. |
| Feature Augmentation | This is similar to the feature combination hybrids but different in that the contributor generates new features. It is more flexible and adds smaller dimension than feature combination method.                                                                                  |

|            |                                                                                                                                                                                                                                                   |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Meta-level | <p>Излаз једног система за препоруку се користи као улаз у други систем за препоруку.</p> <p>For this one, contributing and actual recommenders exist but the former one completely replaces the data for the latter one, not just part of it</p> |
|------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

## 2.4. Предности и мане техника сарадничког филтрирања

**TABLE 1**  
OVERVIEW OF COLLABORATIVE FILTERING TECHNIQUES

| CF categories   | Main advantages                                                                                                                                                                                                                                   | Main shortcomings                                                                                                                                                                                                                              |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Memory-based CF | <ul style="list-style-type: none"> <li>*easy implementation</li> <li>*new data can be added easily and Incrementally</li> <li>*need not consider the content of the items being recommended</li> <li>*scale well with correlated items</li> </ul> | <ul style="list-style-type: none"> <li>*are dependent on human ratings</li> <li>*performance decrease when data are sparse</li> <li>*cannot recommend for new users and items</li> <li>*have limited scalability for large datasets</li> </ul> |
| Model-based CF  | <ul style="list-style-type: none"> <li>*better address the sparsity, scalability and other problems</li> <li>*improve prediction performance</li> <li>*give an intuitive rationale for the recommendations</li> </ul>                             | <ul style="list-style-type: none"> <li>* expensive model-building</li> <li>* have a trade-off between prediction performance and scalability</li> <li>* lose useful information for dimensionality reduction techniques</li> </ul>             |

## 2.5. Алгоритми за факторизацију матрица

- Stochastic Gradient Descent (SGD)
- Bias Stochastic Gradient Descent (B-SGD)
- Alternating Least Squares (ALS); пример за имплицитне податке: [Collaborative Filtering for Implicit Feedback Datasets](#)
- Weighted Alternating Least Squares (W-ALS)

Поређење брзина ових алгоритама можете видети [овде](#).

Пример имплементације за ALS и SGD: <https://blog.insightdatascience.com/explicit-matrix-factorization-als-sgd-and-all-that-jazz-b00e4d9b21ea>

## 2.6. Проблеми система за препоруку

Приликом имплементације система за препоруке обавезно узети у разматрање реалне изазове са којима се сусрећу системи за препоруку (**погледати предавања са трећег термина**).

Изазови код система за препоруку:

- Хладан старт (Cold start)
- Листа „неадекватних“ производа (Stoplist)
- „Бочица филтера“ (Filter bubble)

**Препоруке:**

- Никако не користити имплицитне стрим податке (кликове на рекламе). Такви имплицитни подаци доводе до експлицитних проблема 😊
- Водити рачуна о међународним законима
- Ефекат времена је доста битан у системима за препоруку. Сезоналост (Божих, лето, ...), рејтинг данас VS рејтинг од пре годину дана, ...
- Пратити и пословне циљеве компаније (value-aware recommendations)



### 3. Пројектни задатак

1. R4 скуп података
2. Kaggle takmičenje
3. Bank data
4. Đole mi je dao lepu ideju :D Da vidimo da li je raspoložen i da malo pomogne.  
<https://webscope.sandbox.yahoo.com/catalog.php?datatype=r>