Bayesian Personalized Ranking (BPR)

井上 晴幾

May 28, 2019

Contents

1 Background

ランキング最適化問題。「アクションのあったものは無かったものより好まれる」という前提のもと、ユーザーに対して recommend すべきアイテムのランキング問題を解く。recommend の事後確率を最大化するように最適化する。

2 Fomalization

U は全ユーザー、I は全アイテムとし、アクションのあった組み合わせ S を

$$S \subseteq U \times I \tag{1}$$

と定義する。この時、personalized total ranking $>_u \subset I^2$ を決定する問題を考える。$>_u$ は以下の全順序性（totality, antisymmetry, transitivity）を満たす必要がある。

$$\forall i, j \in I : i \neq j \Rightarrow i >_u j \vee j >_u i \tag{2}$$
$$\forall i, j \in I : i >_u j \wedge j >_u i \Rightarrow i = j \tag{3}$$
$$\forall i, j \in I : i >_u j \wedge j >_u k \Rightarrow i >_u k \tag{4}$$

また、便宜的に以下を定義する。

$$I_u^+ := \{i \in I : (u, i) \in S\} \tag{5}$$
$$U_u^+ := \{u \in U : (u, i) \in S\} \tag{6}$$

以上を元に、データのペアを「アクション有り/無し」で比較する。アクション有りのものは他のものより好まれているとし、以下のようにトレーニングデータセットを作成する。両方ともアクション有り/無しの場合は学習できないのでトレーニングデータからは除外する。

$$D_s := \left\{(u, i, j) | i \in I_u^+ \wedge j \in I \setminus I_u^+\right\} \tag{7}$$

なお、D_s に含まれない欠損値に関しては将来的にはランキングされるべきもの。テストデータとして用いるので D_s とテストデータが独立であることが保証される。

3 Bayesian Personalized Ranking (BPR)

3.1 BPR Optimization Criterion

モデルの事後確率最大化を考える。

$$\text{Max} \quad p(\theta | >_u) \propto p(>_u | \theta) p(\theta) \tag{8}$$

θ はモデルのパラメータ、$>_u$ はユーザー u に関する順序構造。全ユーザーを考慮に入れた尤度関数 $p(>_u | \theta)$ は以下のように書き直せて

$$\prod_{u \in U} p(>_u | \theta) = \prod_{(u,i,j) \in U \times I \times I} p(i >_u j | \theta)^{\delta((u,i,j) \in D_s)} \cdot \left(1 - p(i >_u j | \theta)^{\delta((u,i,j) \notin D_s)}\right) \tag{9}$$

$$\delta(b) = \begin{cases} 1 & \text{if } b \text{ is true} \\ 0 & \text{else} \end{cases} \tag{10}$$

完全性と反対称律から

$$\prod_{u \in U} p(>_u | \theta) = \prod_{(u,i,j) \in D_s} p(i >_u j | \theta) \tag{11}$$

と書ける。また、

$$p(i >_u j|\theta) := \sigma\left(\hat{x}_{uij}\left(\theta\right)\right) \tag{12}$$

$$\sigma\left(x\right) := \frac{1}{1+e^{-x}} \tag{13}$$

と定義する。$\hat{x}_{uij}\left(\theta\right)$ の推定を MF や kNN に投げる。この枠組みなら D_s についてのみ尤度が計算されるので、「0」を学習することによる過学習も起きない。

3.2 BPR-OPT の導出

事前分布は以下のように設定する。

$$p(\theta) \sim \mathcal{N}\left(0, \Sigma_\theta\right) \tag{14}$$

$$\Sigma_\theta = \lambda_\theta I \Rightarrow \Sigma_\theta^{-1} = \frac{1}{\lambda_\theta} I \tag{15}$$

この時、対数尤度関数 BPR-OPT を以下のように定義する。

$$\mathrm{BPR-OPT} := \ln p\left(\theta| >_u\right) \tag{16}$$

$$\propto \ln p\left(>_u |\theta\right) p(\theta) \tag{17}$$

$$= \ln \prod_{(u,i,j)\in D_s} \sigma\left(\hat{x}_{uij}\left(\theta\right)\right) p\left(\theta\right) \tag{18}$$

$$= \Sigma_{(u,i,j)\in D_s} \ln \sigma\left(\hat{x}_{uij}\left(\theta\right)\right) + \ln p\left(\theta\right) \tag{19}$$

$$\propto \Sigma_{(u,i,j)\in D_s} \ln \sigma\left(\hat{x}_{uij}\left(\theta\right)\right) - \frac{1}{2\lambda_\theta}\|\theta\| \tag{20}$$

3.3 Analogies to AUC optimization

ユーザーごとの AUC を以下のように定義する。

$$AUC(u) := \frac{1}{|I_u^+||I \setminus I_u^+|} \Sigma_{i\in I_u^+} \Sigma_{j\in I\setminus I_u^+} \delta(\hat{x}_{uij} > 0) \tag{21}$$

これより、AUC の平均は

$$AUC_{ave} := \frac{1}{|U|} \Sigma_{u\in U} AUC(u) \tag{22}$$

$$= \Sigma_{(u,i,j)\in D_s} z_u \delta(\hat{x}_{uij} > 0) \tag{23}$$

$$z_u = \frac{1}{|U||I_u^+||I \setminus I_u^+|} \tag{24}$$

$$\delta(x > 0) = H(x) = 1 \text{ if } x > 0, \text{ else } 0 \tag{25}$$

$H(x)$ はヘヴィサイド関数（step 関数）の意味。微分不可能なので sigmoid 関数でよく置き換えられる。

3.4 BPR Learning Algorithm

対数尤度関数 BPR-OPT の極値を求める。

$$\frac{\partial \mathrm{BPR\text{-}OPT}}{\partial \theta} = \Sigma_{(u,i,j)\in D_s} \frac{\partial}{\partial \theta} \ln \sigma\left(\hat{x}_{uij}(\theta)\right) - \frac{1}{2\lambda_\theta} \frac{\partial}{\partial \theta} \|\theta\|^2 \tag{26}$$

$$= \Sigma_{(u,i,j)\in D_s} \left(1 + e^{-\hat{x}_{uij}(\theta)}\right) \left(1 + e^{-\hat{x}_{uij}(\theta)}\right)^{-2} \cdot \left(-e^{-\hat{x}_{uij}(\theta)}\right) \frac{\partial}{\partial \theta} \hat{x}_{uij} - \frac{1}{\lambda_\theta}\theta \tag{27}$$

$$\propto \Sigma_{(u,i,j)\in D_s} \frac{-e^{-\hat{x}_{uij}(\theta)}}{1 + e^{-\hat{x}_{uij}(\theta)}} \frac{\partial}{\partial \theta} \hat{x}_{uij}\left(\theta\right) - \frac{1}{\lambda_\theta}\theta \tag{28}$$

これを用いて SGD を行い、BPR-OPT の極値を求める。

3.5 Learning models with BPR

MF と kNN は user-item ペア (u, l) に対して実数 $\hat{x}_{ul}$ を求める問題。$(u, i, j) \in D_s$ は 3 要素なので $\hat{x}_{uij}$ を以下のように分解・定義する。

$$\hat{x}_{uij} := \hat{x}_{ui} - \hat{x}_{uj} \tag{29}$$

$\hat{x}_{ul}$ だけを回帰しようとする他の方法と異なり、$\hat{x}_{ui} - \hat{x}_{uj}$ を分類する問題（triplet loss）となっている。アルゴリズムは論文参照（ただの SGD だけど）。

3.6 Matrix Factorization

以下のような $X : U \times I$ を推定する問題 (Fig. 1)。

$$\hat{X} := W H^T \tag{30}$$
$$W : |U| \times k, H : |I| \times k \tag{31}$$
$$k : dimensinality/rank \tag{32}$$

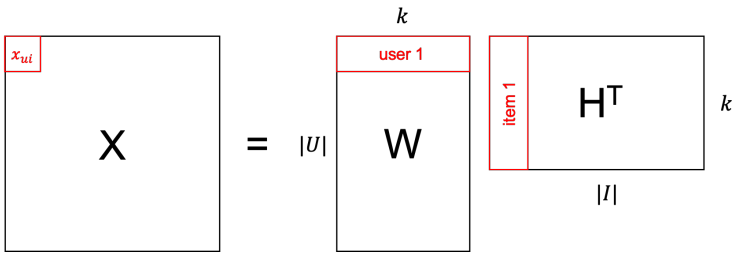

Figure 1: Matrix Factorization

また、予測式は以下のように書き直せる。

$$\hat{x}_{ui} = \langle w_u, h_i \rangle = \Sigma_{f=1}^{k} w_{uf} h_{if} \tag{33}$$
$$\theta = (W, H) \tag{34}$$

$\hat{x}_{ui}$ には linear kernel$\langle w_u, h_i \rangle$ を用いたが、他のカーネルも使える（RBF カーネルなど）。BPR は gradient descent を使っているので $\partial \hat{x}_{uij} / \partial \theta$ を計算しておく必要がある。

$$\hat{x}_{uij} = \hat{x}_{ui} - \hat{x}_{uj} = \langle w_u, h_i \rangle - \langle w_u, h_j \rangle \tag{35}$$

より、

$$\frac{\partial \hat{x}_{uij}}{\partial \theta} = \begin{cases} h_{if} - h_{jf} & \text{if } \theta = w_{uf} \\ w_u & \text{if } \theta = h_{if} \\ -w_u & \text{if } \theta = h_{jf} \\ 0 & \text{else} \end{cases} \tag{36}$$

これを、SGD の更新式

$$\theta \leftarrow \theta + \alpha \left(\Sigma_{(u,i,j) \in D_s} \frac{e^{-\hat{x}_{uij}(\theta)}}{1 + e^{-\hat{x}_{uij}(\theta)}} \frac{\partial}{\partial \theta} \hat{x}_{uij}(\theta) + \frac{1}{\lambda_\theta} \theta \right) \tag{37}$$

で用いる。λ_θ は更新するパラメータごとに設定される（MF では合計 3 つ）

3.7 Adaptive k-Nearest Neighbor

1. item-based と user-based の 2 通りがある
2. 類似度の選び方に依存する

ユーザーにが過去に見た全アイテムと、特定のアイテム i との類似度を以下のように計算する。

$$\hat{x}_{ui} = \Sigma_{l \in I_u^+ \wedge l \neq i} c_{il}, \quad c \in C, C : I \times I \tag{38}$$

ここで、C の計算はヒューリスティックに行われ、例えば

$$c_{ij}^{cosine} := \frac{|U_i^+ \cap U_j^+|}{\sqrt{|U_i^+||U_j^+|}} \tag{39}$$

である。「metric learning してしまった方が良い」と論文では述べている。HH^T のように MF してしまうという手もある。以上より、

$$\hat{x}_{uij} = \hat{x}_{ui} - \hat{x}_{uj} = \Sigma_{l \in I_u^+ \wedge l \neq i} c_{il} - \Sigma_{l \in I_u^+ \wedge l \neq j} c_{jl} \tag{40}$$

$$\frac{\partial \hat{x}_{uij}}{\partial \theta} = \begin{cases} +1 & \text{if } \theta \in \{c_{il}, c_{li}\} \wedge l \in I_u^+ \wedge l \neq i \\ -1 & \text{if } \theta = \in \{c_{jl}, c_{lj}\} \wedge l \in I_u^+ \wedge l \neq j \\ 0 & \text{else} \end{cases} \tag{41}$$

このときの正則化パラメータ λ_θ は 2 つである。