組合せ依存型の Factorization Machines

$$\hat{y}(x_i^{(j)}) = w_0 + \sum_{h=1}^{2M+n} w_h x_h + \sum_{h=1}^{2M+n} \sum_{l=h+1}^{2M+n} \langle \mathbf{v}_h, \mathbf{v}_l \rangle x_h x_l \tag{1}$$

2 つのエンティティの関係

k 番目のエンティティに対応する組合せインデックスベクトルの添え字は $M+k$ で表現されている．ここでは，エンティティインデックスベクトルの i 次元目と組合せインデックスベクトルの k 次元目の交互作用項の推定値である $\langle \mathbf{v}_i, \mathbf{v}_k \rangle$ により，次に示すような関係性のいずれも考慮可能であることを示す．

j 番目の組合せに含まれる 2 つのエンティティ $e_i^{(j)}, e_k^{(k)}$ の関係について，次の 4 つを考えることができる．

1. $e_i^{(j)}, e_k^{(k)}$ は同じ組合せの中で，共にパフォーマンスを向上させ合う．
2. $e_i^{(j)}, e_k^{(j)}$ は同じ組合せの中で，共にパフォーマンスを低下させ合う．
3. $e_i^{(j)}$ は $e_k^{(j)}$ に対して『得手』であり，同じ組合せの中で $e_i^{(j)}$ のパフォーマンスは向上し，$e_k^{(j)}$ のパフォーマンスは低下する．
4. $e_i^{(j)}$ は $e_k^{(j)}$ に対して『不得手』であり，同じ組合せの中で $e_i^{(j)}$ のパフォーマンスは低下し，$e_k^{(j)}$ のパフォーマンスは向上する．

いま，エンティティ $e_i^{(j)}$ についてのモデルの特徴量である $x_i^{(j)}$ について，$\hat{y}(x_i^{(j)})$ の値が小さいほど，そのエンティティはより良いパフォーマンスをすると推定されると考える．この時，1 の関係性について $x_i^{(j)} x_{M+k}^{(j)} > 0$ かつ $x_k^{(j)} x_{M+i}^{(j)} > 0$ であるから，式 (1) より $\langle \mathbf{v}_i, \mathbf{x}_{M+k} \rangle$ と $\langle \mathbf{v}_k, \mathbf{x}_{M+i} \rangle$ が負の値を取れば，お互いのパフォーマンスを向上させ合うことを表現できる．また，$\langle \mathbf{v}_i, \mathbf{v}_{M+k} \rangle < 0$ かつ $\langle \mathbf{v}_k, \mathbf{v}_{M+i} \rangle > 0$ であれば，3 の関係性を表現していると言える．このような因子ベクトルの組合せは実際に実現可能であり，因子ベクトルの次元数が 2 である場合は，$\mathbf{v}_i = (1,1), \mathbf{v}_{M+k} = (-1,-1)$ や $\mathbf{v}_k = (-1,1), \mathbf{v}_{M+k} = (-1,1)$ がその具体例である．$2,4$ の関係性については，それぞれ $1,3$ の場合の議論と同様である．

因子ベクトルの次元数

一般に，n 個のエンティティについて考えられるエンティティ対の 1 から 4 の関係性を考慮するために必要な次元数は n である．

目的関数

データセット D に対して設定される損失関数 J は以下の (2) で表現される．

$$J(D) = \frac{1}{2}\sum_{D}\left(y - \hat{y}\right)^2 + \frac{1}{2}C\left\{\sum_{i=1}^{M}\sum_{j=i+1}^{M}\left(\langle \mathbf{v}_i, \mathbf{v}_j \rangle\right)^2 + \sum_{i=1}^{M}\sum_{j=i+1}^{M}\left(\langle \mathbf{v}_i, \mathbf{v}_j \rangle\right)^2\right\} \tag{2}$$

最適化

$$\frac{\partial}{\partial \theta} J = \begin{cases} y - \hat{y} & \text{if } \theta = w_0, \\ (y - \hat{y}) x_i & \text{if } \theta = w_i, \\ x_i \sum_{j=1}^{2M+n} v_{j,f} x_j - v_{i,f} x_i^2 + C \left(\sum_{j=i+1}^{M} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \right) v_{i,f} & \text{if } \theta = v_{i,f} \wedge i \in [1, M], \\ x_i \sum_{j=1}^{2M+n} v_{j,f} x_j - v_{i,f} x_i^2 + C \left(\sum_{j=i+1}^{2M} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \right) v_{i,f} & \text{if } \theta = v_{i,f} \wedge i \in [M+1, 2M], \\ x_i \sum_{j=1}^{2M+n} v_{j,f} x_j - v_{i,f} x_i^2 & \text{otherwise;} \end{cases} \tag{3}$$