組み合わせ依存型ランキング学習 - 競馬への応用事例

清水　豪士

富山県立大学　情報基盤工学講座
t715038@st.pu-toyama.ac.jp

October 29, 2021

TOYAMA Prefectural University

はじめに

背景

- エンティティの順位付け問題を解決するために機械学習の手法を利用することをランキング学習という.
- 競馬など領域において，情報検索の分野では利用されてきたランキング学習の手法をそのまま用いては，予測精度が芳しくないことが予想される.

目的

- 情報検索の分野では考慮されなかった，順位付けの対象となるエンティティ同士が互いに影響を及ぼしあうことでランキングが決定される領域を対象とした，Factorization Machine を利用を応用した組み合わせ依存型のランキング学習の手法の提案.

Factorization Machine：FM

Factorization Machine（FM）

- 複数の特徴量の交互作用を推定可能なモデル.
- SVM のような汎用的な予測器であり，全ての特徴量同士の交互作用による影響をモデルに落とし込むことができる.
- SVM など既存の方法ではうまくいかないような，スパースなデータでも学習できる.

FM の特徴

- 特徴量の交互作用項の推定方法.
  - 一般的なモデルは交互作用項 $x_i x_j$ に対する重みを w_{ij} として推定する.
  - FM は $\hat{w}_{ij} =< \boldsymbol{v}_i, \boldsymbol{v}_j >$ のように 2 つのベクトルの内積として推定する.
  - これは，与えられた事例において観測されない交互作用項に対する重みの推定が可能.

Factorization Machine：FM

Factorization Machine（FM）

FM では入力変数 $\boldsymbol{x}$ に含まれる d 個の要素同士の交互作用による影響を考慮することが可能で，$d = 2$ である場合のモデル式は以下の式で表される．

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \tag{1}$$

上記の式で推測されるパラメータは以下である．

$$w_0 \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^n, \mathbf{V} \in \mathbb{R}^{n \times k} \tag{2}$$

w_0：各事例において共有されるグローバルバイアス．

w_i：入力変数 i 次元目の要素による影響をモデル化したもの．

$< \boldsymbol{v}_i, \boldsymbol{v}_j >$：入力変数の i 次元目と j 次元目の要素同士の交互作用による影響の強さをモデル化したもの．

k :FM におけるハイパーパラメータ．

手法：問題設定

問題設定

E をあるクラスのエンティティ集合とし，E に属するエンティティ数を M とする．また，集合 E に属するエンティティを e_i とする．
E に属する複数のエンティティにより決定される j 番目の組み合わせを $c^{(j)}$ とし，$c^{(j)}$ に含まれるエンティティ数を $m_j = |c^{(j)}|$ と表す．
さらに，エンティティ順序 $\preceq_j$ を $c^{(j)}$ 上の全順序として定義する．

この時，$c^{(j)}$ について観測された順序集合 $O_{\preceq_j}$ から，関数 $y(\boldsymbol{x}_i^{(j)})$ を学習することが目的．
$\boldsymbol{x}_i^{(j)}$ は組み合わせ依存型ランキング学習の特徴量である．

手法：組み合わせ依存型特徴量の作成

組み合わせ依存型特徴量の作成

3 つのベクトルを定義する

1. エンティティインデックスベクトル
2. 組み合わせインデックスベクトル
3. 文脈ベクトル

エンティティインデックスベクトル

組み合わせ $c^{(j)}$ において，エンティティ e_i が順位推定の対象であることを識別するためのベクトル．
以下のように定義される．

$$\mathbf{x}_i^e = (\underbrace{0, \ldots, 0}_{i-1 \text{ 個}}, 1, \underbrace{0, \ldots, 0}_{M-i \text{ 個}}) \tag{5}$$

手法：組み合わせ依存型特徴量の作成

組み合わせインデックスベクトル

組み合わせ $c^{(j)}$ に含まれるエンティティ集合を識別するためのベクトル．以下のように定義される．

$$\mathbf{x}^{c^{(j)}} = (\underbrace{0, 1, 1, 0, \ldots, 1, 0, \ldots}_{M \text{ 個}}), \bar{z}(\mathbf{x}^{c^{(j)}}) = m_j \tag{6}$$

文脈ベクトル

モデルの利用者に応じて自由に設定されるベクトル．以下のように定義される．

$$\mathbf{x}_i^f \in \mathbb{R}^n \tag{7}$$

(5)(6)(7) を利用して，組み合わせ $c^{(j)}$ における i 番目のエンティティの特徴ベクトル $\boldsymbol{x}_i^{(j)}$ は次の式のように設定される．

$$\mathbf{x}_i^{(j)} = \mathbf{x}_i^{e} \frown \mathbf{x}^{c} \frown \mathbf{x}_i^{f} \tag{8}$$

手法：組み合わせ依存型特徴量の作成

例

$M = 5$ であり，一番目の組み合わせ $c^{(1)}$ においてエンティティ e_1, e_3, e_4 を順位付けているとする．このとき，エンティティインデックスベクトルはそれぞれ

$$\mathbf{x}_1^e = (1,0,0,0,0) \quad \mathbf{x}_3^e = (0,0,1,0,0) \quad \mathbf{x}_4^e = (0,0,0,1,0)$$

であり，組み合わせインデックスベクトルは

$$\mathbf{x}^{c^{(1)}} = (1,0,1,1,0)$$

となる．それぞれのエンティティの文脈ベクトルが

$$\mathbf{x}_1^f = (x_{11}^f, x_{12}^f) \quad \mathbf{x}_3^f = (x_{31}^f, x_{32}^f) \quad \mathbf{x}_4^f = (x_{41}^f, x_{42}^f)$$

のように 2 次元で表現されているとすれば，以下のベクトルがモデルの入力特徴量となる．

$$\mathbf{x}_1^{(1)} = (1,0,0,0,0,1,0,1,1,0,x_{11}^f, x_{12}^f)$$
$$\mathbf{x}_3^{(1)} = (0,0,1,0,0,1,0,1,1,0,x_{31}^f, x_{32}^f)$$
$$\mathbf{x}_4^{(1)} = (0,0,0,1,0,1,0,1,1,0,x_{41}^f, x_{42}^f)$$

手法：組み合わせ依存型ランキング学習

組み合わせ依存型ランキング学習

特徴量 $\boldsymbol{x}_i^{(j)}$ を入力とし，組み合わせ $c^{(j)}$ における i 番目のエンティティの順位を推定する回帰問題を FM を用いて解いた結果，推定されるパラメータは以下のものになる．

$$w_0 \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^{2M+n}, \mathbf{V} \in \mathbb{R}^{(2M+n)\times k} \tag{9}$$

(9) における交互作用項に対する重みの因子ベクトル $\boldsymbol{v}_i, i \in [1, (2M+n) \times k]$ について考えると，1 行目から M 行目までの因子ベクトルは i 番目のエンティティ固有の性質を表す因子ベクトルと捉えることができる．
また，$M \times 1$ 行目から $2M$ 行目までの因子ベクトルは，組み合わせ i 番目のエンティティの性質を表す因子ベクトルとして解釈することができる．

これら因子ベクトルを利用して，同じ組み合わせに含まれるエンティティが互いに影響を与え合うことを表現可能である．

FM と組み合わせ依存型の特徴量を利用することにより，組み合わせ $c^{(j)}$ が決まることにより変動するエンティティのパフォーマンスをモデルに落とし込むことができ，その結果として変動する順位を推定することができる．

実験

実験：データセット

- 2010 年から 2016 年の間に JRA の 10 の競馬場で開催されたレースを対象としたデータの収集.
- 競走馬については競馬新聞などに公開されている情報から抽出.
- 騎手については netkeiba から情報を抽出.

表 1　全競馬場，および，各競馬場ごとのレース数と競走馬数

競馬場名	レース数	出走数	平均出走馬数
札幌競馬場	1008	12783	12.68
函館競馬場	1248	15472	12.40
福島競馬場	1502	22330	14.87
新潟競馬場	2072	30788	14.86
東京競馬場	3517	52392	14.90
中山競馬場	3153	46968	14.90
中京競馬場	1575	24326	15,46
京都競馬場	3769	53598	14.12
阪神競馬場	3272	47918	14.64
小倉競馬場	1982	29449	14.86
全競馬場	23098	38631	14.51

実験：提案手法の競馬への応用

特徴量の作成

- M をデーセットに含まれるユニークな競走馬の総数とする.
- 特徴量抽出の対象となるデータに含まれる競走馬名から各競走馬に対して識別子を割り振り，各競走馬の識別子を M 次元のベクトルの各次元に 1 対 1 に対応させる.
- 着順の推定の対象となる競走馬に対応する次元が 1 であり，その他の要素が 0 であるような one-hot ベクトルをエンティティインデックスベクトル.
- 競争相手として同じレースに出走する競走馬に対応する次元が 1 であり，その他の要素が 0 であるようなベクトルを組み合わせインデックスベクトル.
- 特徴量を各競走馬の過去の出走成績から抽出したものを文脈ベクトル.
- これらの 3 つのベクトルを結合し，モデルの入力とする.

モデルの学習

- モデルの出力 $\hat{y}$ は各競走馬のレースでの着順であり，学習の段階では次に示すような正規化を行う．

$$y_i^{(j)} = \frac{FP_i^{(j)} - FP_{min}^{(j)}}{FP_{max}^{(j)} - FP_{min}^{(j)}} \tag{10}$$

- この変換により，各競走馬の変換後の着順は最小値が 0，最大値が 1 となるため，出走馬数が異なるレースに対しても同一のモデルで学習を行うことが可能になる．
- 特徴量と着順のデータを利用し，変換した実際の着順とその予測値の 2 乗誤差の総和が，与えられたトレーニングデータに対して最小となるようにモデルの学習を行う．

$FP_i^{(j)}$ ：j 番目のレースでの i 番目の競走馬の着順
$FP_{min}^{(j)}, FP_{max}^{(j)}$ ：レースでの競走馬の着順の最小値と最大値
$y_i^{(j)}$ ：$FP_i^{(j)}$ に対応した，変換後の値

2 段階条件付きロジスティック回帰モデル

- 競馬の勝ち馬予測のタスクにおいて，ベースラインとして利用される手法は条件付きロジスティック回帰を利用した 2 段階の予測モデル.
- 出力する形は各競走馬の勝率を出力する.

モデルの特徴

- 勝ち馬予測を 2 段階に分け，各競走馬の過去の成績から得られる競走馬固定の競争能力と，オッズの 2 つの情報から包括的に予測を行うという点.

この研究において

- 競走馬の特徴同士が互いに影響を与えて順位を変動させることを確かめるために，オッズを特徴量として利用しない.
- よって，比較対象となる 2 段階条件付きロジスティック回帰モデルでもオッズを利用しない第 1 段階の出力を利用する.

実験結果とモデルの比較

- 提案手法とベースラインの手法である条件付きロジスティック回帰モデルの異なる両手法を比較するために，提案手法については着順の予測値を昇順，ベースラインの手法については勝率の推定値を降順に並び替えることで各モデルの出力と 1 対 1 に対応するランキングを作成する．
- ランキングを導出して，nDCG@3 と nDCG@5 を評価指標として，各手法のパフォーマンスの比較を行う．

コンテクスト依存型情報フィルタリング

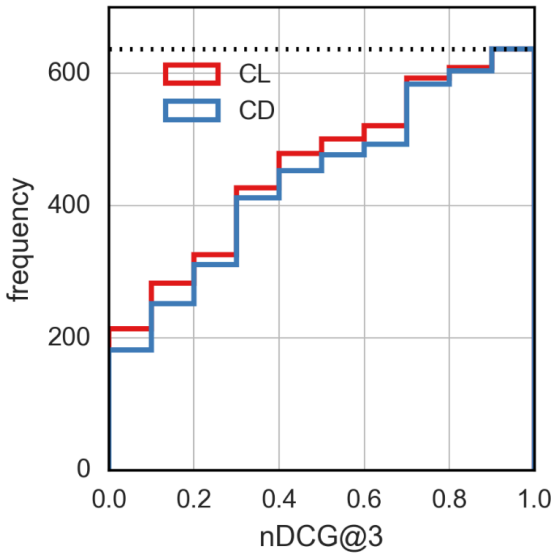

図 1: nDCG@3 の累積度数分布

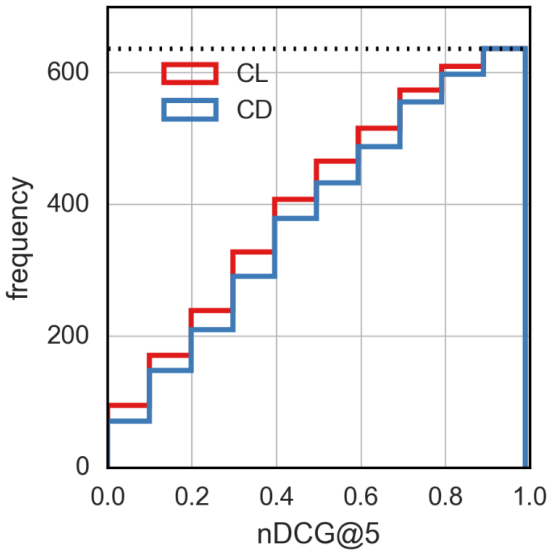

図 2: nDCG@5 の累積度数分布

表 2　各手法の評価指標の値

	ベースライン手法	提案手法
nDCG@3 の平均値	0.314	0.345
nDCG@5 の平均値	0.412	0.446

- 提案手法はベースラインと比較して全体的に評価指標が高い値を出している．
- 表 2 より，提案手法はベースラインよりも評価指標の平均値が高く，良いパフォーマンスを実現している．

まとめ

まとめ

- エンティティの組み合わせに応じて各エンティティの特徴量同士が互いに与える影響を考慮したエンティティの順位付けの問題に取り組んだ.
- 組み合わせ依存型ランキング学習を提案し，競馬を具体的な応用事例として実験を行い，既存の勝ち馬予測の手法を上回るパフォーマンスが得られることを明らかにした.

課題

- ランキングにおいて特に重要となる上位 k 件（競馬だと k=3）の推定を重視したモデルの学習.
- 学習されるパラメータの 1 つである因子行列 V の値を利用したエンティティの性質や説明のクラスタリングの手法の提案