

Mixture Factorization Machines Considering the Latent Characteristics Behind Target Data

SUGISAKI Tomoya

1. 研究背景・目的

近年、情報技術の発展により、蓄積された大量のデータを活用した様々な分析モデルが構築されている。このうち、説明変数から目的変数を予測する最も基本的な手法として線形回帰モデルが知られている。しかしながら、例えば、ユーザによる評価値データの中には、他のユーザに比べて高い評価傾向にあるユーザ群や低い評価傾向にあるユーザ群など、特徴の異なる潜在的なグループが混在していることがある。このような潜在的特徴が異なるグループ間では、説明変数と目的変数の間の関係性も異なり、単独の回帰式ではこれらの関係性を適切に表現できない可能性がある。このような場合に適用可能な手法として、データの背景に隠れた潜在的特徴を仮定し、潜在クラス毎に異なる回帰式を混合する混合回帰モデル [1] が知られている。このモデルは、各潜在クラス毎に異なる回帰式を仮定し、その混合をするため、異なる関数関係が混合している問題に対して、精度の高い推定を可能とする。

一方、比較的少ないパラメータ数で交互作用を考慮可能な回帰モデルの 1 つとして Factorization Machines [2] (以下、FM) が注目されている。FM では得られた交互作用から、特徴量が共起した場合の目的変数に対する影響力を定量的に捉えることができる。また、一般的な多項式回帰モデルと比べて、少ないパラメータ数で同様の表現能力を有するため、低自由度で過学習しづらく、汎化能力の高いモデルが得られる。しかし、データの中には特徴の異なる潜在的なグループが混在している場合があり、潜在的特徴が異なるグループ間では、交互作用効果も異なる。そのため、単独の FM ではこのような場合、説明変数と目的変数の関係性を適切に表現できない可能性がある。

そこで本研究では、混合回帰モデルを FM に援用することで、データの潜在的特徴を考慮し、FM を混合した Mixture FM を提案する。それぞれの潜在クラス毎に FM を仮定し、混合することで、データの潜在的特徴を表現することが可能となり、単一な FM よりも高い予測精度が期待される。また、潜在クラス毎に交互作用が得られるため、単一の FM と比較して詳細な分析が可能となる。さらに、新規データの潜在クラスへの所属確率を計算することで、新規データに対する予測も可能となる。最後に、実データに対して提案モデルを適用することで、モデルの評価を行い、推定されたモデルを分析することで、有用な知見が得られることを示す。

2. 準備

2.1. 基本情報を考慮した混合回帰モデル

永森ら [3] は、説明変数と目的変数の関係性を混合回帰モデルをもとに、基本情報を考慮した予測モデルを提案してい

る。いま、潜在クラス集合を $\mathcal{Z} = \{z_k : 1 \leq k \leq K\}$ とすると、混合回帰モデルは各潜在クラスにおける確率密度関数 $P_k(y_n | \mathbf{x}_n)$ の線形結合によりモデル化される。ここで、説明変数ベクトルは $\mathbf{x}_n = (x_{n0}, x_{n1}, \dots, x_{nI})^T$ 、目的変数は y_n である。また、潜在クラス z_k における $I + 1$ 個の回帰係数を $\boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k}, \dots, \beta_{Ik})^T$ としたとき、 y_n の確率密度関数は分散を σ_k^2 とすると次式で表される。

$$P_k(y_n | \mathbf{x}_n) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(y_n - h_k(\mathbf{x}_n))^2}{2\sigma_k^2}\right) \quad (1)$$

$$h_k(\mathbf{x}_n) = \boldsymbol{\beta}_k^T \mathbf{x}_n \quad (2)$$

さらに入力データに関する j 番目の補助情報 ($1 \leq j \leq J$) の要素集合を $\mathcal{D}^j = \{d_{m_j}^j : 1 \leq m_j \leq M_j\}$ とする。ただし、 $d_{m_j}^j$ は j 番目の補助情報の m_j 番目の要素、 M_j は j 番目の補助情報の要素数を表す。また、補助情報を表す変数ベクトルを $\mathbf{d}_n = (d_{n1}, \dots, d_{nj}, \dots, d_{nJ})^T$ 、 d_{nj} を n 番目のデータの j 番目の補助情報の要素とする。このとき基本情報を考慮した混合回帰モデルは式 (3) で表される。

$$P(y_n, \mathbf{x}_n, \mathbf{d}_n) = \sum_{k=1}^K P(z_k) P_k(y_n | \mathbf{x}_n) \gamma_{nk} \quad (3)$$

ただし、 γ_{nk} は次式で与えられる項であり、 $\delta(a, b)$ は a と b が一致していれば 1、さもなければ 0 をとる指示関数である。

$$\gamma_{nk} = \prod_{j=1}^J \prod_{m_j=1}^{M_j} P(d_{m_j}^j | z_k) \delta(d_{m_j}^j, d_{nj}) \quad (4)$$

2.2. Factorization Machines

一般に、入力データの特徴量数を I としたとき、多項式回帰モデルによって 2 次の交互作用を表現するためのパラメータ数は I^2 に比例する。そのため、特徴量数が増えると、推定に必要なパラメータ数が指数的に増加してしまい、その結果、過学習などの問題が生じてしまう。FM ではこのような問題を防ぐため、交互作用行列と呼ばれる低次元行列を用い、行ごとに内積を計算することで比較的少ないパラメータ数で特徴量間の交互作用を表現している。これにより、有限の学習データを用いて推定する場合、多項式回帰モデルと比べて、高い予測精度を得ることができる。

いま、目的変数を $y \in \mathbb{R}$ 、説明変数ベクトルを $\mathbf{x} = (x_1, x_2, \dots, x_I)^T$ ($x_i \in \{0, 1\}, i = 1, 2, \dots, I$) としたデータが与えられているとし、 w_0 をバイアス項、 $\mathbf{w} = (w_1, w_2, \dots, w_I)^T$ を重みベクトルとする。特徴量間の交互

作用をその内積により表現することができる交互作用行列を $V = [\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_I^T]^T \in \mathbb{R}^{I \times K}$ とし、その要素を交互作用ベクトル $\mathbf{v}_i = (v_{i1}, v_{i2}, \dots, v_{iK})^T$ と定義する。このとき、 \mathbf{x} が与えられたときの FM は次の式 (5) で表される。

$$f(\mathbf{x}) = w_0 + \sum_{i=1}^I w_i x_i + \sum_{i=1}^{I-1} \sum_{j=i+1}^I \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \quad (5)$$

ただし、 $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ は \mathbf{v}_i と \mathbf{v}_j の内積を表す。FM は重回帰モデルに対して式 (5) の第 3 項を加えることで、特徴量間の交互作用を表現している。そのため、FM は交互作用が存在する問題に対して重回帰モデルよりも予測精度の向上が期待できる。また、予測問題を対象とした場合、損失関数として二乗誤差を用いることが多く、パラメータは確率的勾配降下法 (SGD) などを用いて推定される。

さらに、Rendle らは予測問題に対する損失関数として二乗誤差を用いた場合、FM の誤差は回帰モデルと同様に正規分布に従うことを仮定している。このときの誤差の分散を σ^2 とすると、FM の確率密度関数は次の式 (6) で表される。

$$P(y|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - f(\mathbf{x}))^2}{2\sigma^2}\right) \quad (6)$$

3. 提案モデル

データの中には特徴の異なる潜在的なグループが混在していることがあるため、潜在的特徴が異なるグループ間では、FM の交互作用効果も異なる。そのため、単独の FM ではこのような場合、説明変数と目的変数の関係性を適切に表現できない可能性がある。そこで本研究では、基本情報を考慮した混合回帰モデルを援用し、データの背景に隠れた潜在的特徴を考慮した潜在クラスモデルと FM を組み合わせ、潜在クラスごとに異なる FM を混合する手法として Mixture FM を提案する。このモデルは FM と入力データの補助情報との共起を考えており、補助情報からそのデータの潜在クラスへの所属確率を推定することができる。そのため、学習データ中に存在しない新規データに対しても、目的変数の予測を可能とする。さらに、潜在クラスごとの FM から得られるパラメータを分析することで、説明変数が目的変数に与える影響力を潜在クラスごとに定量的に捉えることができる。

3.1. 定式化

いま、入力データに関する m 番目の補助情報 ($1 \leq m \leq M$) の要素集合を $\mathcal{D}^m = \{d_{o_m}^m : 1 \leq o_m \leq O_m\}$ とする。ただし、 $d_{o_m}^m$ は m 番目の補助情報の o_m 番目の要素、 O_m は m 番目の補助情報の要素数を表す。また、 n 番目のデータの説明変数と目的変数を \mathbf{x}_n と y_n 、補助情報を表す変数ベクトルを $\mathbf{d}_n = (d_{n1}, \dots, d_{nm}, \dots, d_{nM})^T$ 、 d_{nm} を n 番目の入力データの m 番目の補助情報の要素とする。さらに、潜在クラス集合を $\mathcal{Z} = \{z_l : 1 \leq l \leq L\}$ としたとき、提案する確率モデルは次の式 (7) で表される。

$$P(y_n, \mathbf{x}_n, \mathbf{d}_n) = \sum_{l=1}^L P(z_l) P_l(y_n|\mathbf{x}_n) \gamma_{nl} \quad (7)$$

ただし、 γ_{nl} は次式で与えられる項である。

$$\gamma_{nl} = \prod_{m=1}^M \prod_{o_m=1}^{O_m} P(d_{o_m}^m | z_l)^{\delta(d_{o_m}^m, d_{nm})} \quad (8)$$

また、潜在クラス z_l ごとに仮定される FM は w_{l0} をバイアス項、 $\mathbf{w}_l = (w_{l1}, w_{l2}, \dots, w_{lI})^T$ を重みベクトル、交互作用ベクトルを $\mathbf{v}_{li} = (v_{li1}, v_{li2}, \dots, v_{liK})^T$ とすると式 (9) のように表すことができる。さらに、分散を σ_l^2 としたとき、確率密度関数は次の式 (10) で表される。

$$f_l(\mathbf{x}_n) = w_{l0} + \sum_{i=1}^I w_{li} x_{ni} + \sum_{i=1}^{I-1} \sum_{j=i+1}^I \langle \mathbf{v}_{li}, \mathbf{v}_{lj} \rangle x_{ni} x_{nj} \quad (9)$$

$$P_l(y_n|\mathbf{x}_n) = \frac{1}{\sqrt{2\pi\sigma_l^2}} \exp\left(-\frac{(y_n - f_l(\mathbf{x}_n))^2}{2\sigma_l^2}\right) \quad (10)$$

さらに、 l に関しての和が 1 となる n 番目のデータが与えられたときの各潜在クラスの事後確率を g_{nl} とすると、Mixture FM のモデル式は式 (11) で表される。また、提案モデルのグラフィカルモデルは次の図 1 で表される。

$$P(y_n|\mathbf{x}_n) = \sum_{l=1}^L g_{nl} P_l(y_n|\mathbf{x}_n) \quad (11)$$

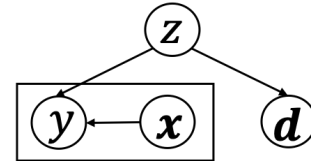


図 1: 提案モデルのグラフィカルモデル

3.2. パラメータ推定

提案モデルのパラメータは EM アルゴリズムを用いて推定することができる。学習データに対する対数尤度関数 LL は次の式 (12) で表される。

$$LL = \sum_{n=1}^N \log P(y_n, \mathbf{x}_n, \mathbf{d}_n) \quad (12)$$

EM アルゴリズムでは式 (12) で表される対数尤度を局所最大化するように E-step と M-step を収束するまで繰り返し計算することでパラメータ推定を行う。

[E-step]

$$g_{nl} = \frac{P(z_l) P_l(y_n|\mathbf{x}_n) \gamma_{nl}}{\sum_{l=1}^L P(z_l) P_l(y_n|\mathbf{x}_n) \gamma_{nl}} \quad (13)$$

[M-step]

M-step では E-step で得られた g_{nl} を固定したもとの、各

パラメータを更新する。まず、混合比 $P(z_l)$ と潜在クラス z_l における分散 σ_l^2 は次の式 (14) および (15) で更新する。

$$P(z_l) = \frac{\sum_{n=1}^N g_{nl}}{N} \quad (14)$$

$$\sigma_l^2 = \frac{\sum_{n=1}^N g_{nl}(y_n - f_l(\mathbf{x}_n))^2}{\sum_{n=1}^N g_{nl}} \quad (15)$$

これらの更新式により各データを L 個の潜在クラスに確率的に所属させることで、その重みを用いて各潜在クラスごとに FM を構築することを考える。潜在クラス z_l の FM におけるパラメータ λ_l は式 (16) を SGD などを用いて更新する。

$$\lambda_l = \arg \min_{\lambda_l} \sum_{n=1}^N g_{nl}(y_n - f_l(\mathbf{x}_n))^2 \quad (16)$$

最後にデータの補助情報の各潜在クラスにおける出現確率は次の式 (17) を用いて更新する。

$$P(d_{om}^m | z_l) = \frac{\sum_{n=1}^N g_{nl} \delta(d_{om}^m, d_{nm})}{\sum_{n=1}^N g_{nl}} \quad (17)$$

3.3. 新規データに対する予測

本提案モデルは補助情報を考慮した混合回帰モデルを援用しているため、新規データの予測を行うことが可能である。データの特徴はそのデータの補助情報で表現されるとし、 t 番目の予測対象データ $\mathbf{x}'_t = (x_{t1}, \dots, x_{ti}, \dots, x_{tI})^T$ の補助情報を表すベクトルを $\mathbf{d}'_t = (d'_{t1}, \dots, d'_{tm}, \dots, d'_{tM})^T$ 、予測値を \hat{y}_t とする。このとき、補助情報 d'_t から各潜在クラスへの所属確率を計算することができ、それぞれの潜在クラスごとに構築した FM を混合することによって新規データに対する予測を行う。ここで、 t 番目の予測対象データの潜在クラス z_l への所属確率の予測値を \hat{g}_{tl} としたとき、 \hat{g}_{tl} および、予測値 \hat{y}_t は次の式 (18) および (19) で求められる。

$$\hat{g}_{tl} = \frac{P(z_l) \gamma'_l}{\sum_{l=1}^L P(z_l) \gamma'_l} \quad (18)$$

$$\hat{y}_t = \sum_{l=1}^L \hat{g}_{tl} f_l(\mathbf{x}_t) \quad (19)$$

ただし、 γ'_l は次式で与えられる項である。

$$\gamma'_l = \prod_{m=1}^M \prod_{o_m=1}^{O_m} P(d_{om}^m | z_l)^{\delta(d_{om}^m, d'_{tm})} \quad (20)$$

4. 提案モデルの実データへの適用

提案手法の有効性を検証するため、ユーザーレビューを対象とした被レビュー数の予測を行う。さらに得られたパラメータを用いた分析を行う。

4.1. 実験条件

実験データとして、2010年1月から2012年12月までに楽天市場に登録された商品のレビューデータ [5] を用いる。新規商品を掲載する際、良い評価をどれほど得られるかを予

測するために、平均評価値が3以上の商品を対象とした。また、外れ値を削除するために価格が101円以上10,000円以下の商品を対象とするものとし、500円毎に価格を区切り、それぞれをカテゴリ値として扱う。同様に、外れ値を削除するために、被レビュー数が5件以上1,000件以下の商品を用いるものとした。説明変数には商品の小ジャンル、写真掲載の有無と価格を用い ($I = 1, 635$)、被レビュー数を目的変数とする。また、補助変数は商品の13種類の大ジャンルを用い ($M = 1, O_1 = 13$)、ジャンルごとのデータ数を揃えるために各ジャンルから8,000件のデータをランダムに抽出し、104,000件のデータを用いた。ここで表1に商品情報の例を示す。

表 1: 商品情報例

商品名	フレーバー紅茶【白桃】 三角ティーバッグ 20 個入り
小ジャンル	ティーバッグ
大ジャンル (補助情報)	食品
写真掲載	有
価格 (円)	101~500

比較モデルは線形回帰モデルと単一の FM を用いる。FM の交互作用ベクトルの次元数は、事前実験による探索的方法で40と設定した。本実験ではそれぞれのモデルでテストデータの商品に対して被レビュー数の予測を行い、算出された予測値と実測値の平均絶対誤差を評価することにより、予測精度を評価した。またデータセットを5分割し、そのうちの1つをテストデータ、残りを学習データとしたときの平均を用いて評価を行う。

4.2. 結果と考察

実験結果を表2に示す。なお、太字は最も精度が高いことを表す。

表 2: 平均絶対誤差の比較結果

L	提案モデル		FM		線形回帰	
	学習	テスト	学習	テスト	学習	テスト
2	22.16	30.81				
3	22.16	30.79				
4	20.76	30.74				
5	20.73	30.50				
6	20.40	30.62	32.06	32.59	32.17	32.98
7	19.82	30.78				
8	19.55	30.91				
9	19.72	31.04				
10	19.85	32.07				

表2から複数の FM を混合することにより、モデルの推定精度が向上していることがわかる。この結果から、商品の基本情報と被レビュー数の関係性の構造は複数存在し、潜在クラスを仮定し潜在的特徴を考慮した Mixture FM モデルがより適していることがわかる。また、潜在クラス数が5のとき、テストデータに対する予測精度が最も高く、さらにそこからクラス数が増加すると、学習データに対する予測精度が高くなる一方、テストデータに対する予測精度が低下していることがわかる。これは、潜在クラス数の増加に伴い、必要なパラメータ数が増加するため、過学習が起きている可能性がある。そのため、学習データ数とパラメータ数に応じて適切な潜在クラス数が存在し、本提案モデルを適用する場合

は使用するデータに応じて適切な潜在クラス数を決定する必要があると考えられる。

4.3. 提案モデルを用いた分析

以下では提案モデルにおいて最も予測精度が高かった潜在クラス数 $L = 5$ のときに推定されたパラメータを用いて分析を行う。潜在クラスごとに商品の特徴が現れているかを確認し、各クラスの交互作用に異なる特徴が現れているかどうかを分析する。分析データは前節と同様のデータを用いている。

4.3.1. 潜在クラスごとの商品分析

各潜在クラスに所属している商品の補助情報を分析するために、その生起確率 $P(d_{om}^m | z_l)$ をもとに、生起確率が高い補助情報をまとめた表を次に示す。

表 3: 各潜在クラスに所属する商品の特徴

	特徴
z_1	インテリア, 宝石
z_2	PC・スマホ・電子機器, スポーツ, 健康グッズ
z_3	DIY 系, キッチン, ベット, 乗り物
z_4	子供用品, 生活用品, 食べ物
z_5	本, 生活用品

表 3 からクラス 1 とクラス 2 には比較的価格帯の高い商品ジャンルの所属確率が高いことがわかる。また、クラス 4 とクラス 5 には日常的に必要なものや比較的価格帯の低い商品ジャンルの所属確率が高いことがわかる。このように、各潜在クラスに異なる特徴を持った商品が所属していることがわかる。

4.3.2. 潜在クラスごとの交互作用分析

本節では各潜在クラスごとに交互作用の違いが表現できているかを確認するため、商品出品時の写真掲載と出品価格との間の交互作用に着目した分析を行う。以下に各潜在クラスごとの商品出品時の写真掲載と各出品価格帯との交互作用の値の大きい上位 3 件を示す。

表 4: 写真掲載と各価格帯との交互作用

z_1	特徴量	交互作用
1	3,001 円~3,500 円	23.59
2	4,001 円~4,500 円	17.02
3	1,501 円~2,000 円	16.03
z_2	特徴量	交互作用
1	101 円~500 円	24.00
2	9,501 円~10,000 円	15.15
3	4,501 円~5,000 円	11.73
z_3	特徴量	交互作用
1	1,001 円~1,500 円	28.05
2	5,501 円~6,000 円	25.63
3	1,501 円~2,000 円	25.34
z_4	特徴量	交互作用
1	3,501 円~4,000 円	11.00
2	7,501 円~8,000 円	10.55
3	8,501 円~9,000 円	8.485
z_5	特徴量	交互作用
1	2,501 円~3,000 円	20.51
2	7,501 円~8,000 円	8.773
3	2,001 円~2,500 円	6.910

表 4 から、潜在クラス毎に異なる交互作用が得られていることがわかる。表 3 と合わせて分析することで、詳細な分析が可能となる。例えば、他に比べて比較的高価格な商品が多いインテリアや宝石が多く所属する潜在クラス 1 に着目すると、3,000 円~4,500 円の価格帯に対して写真掲載の交互作用が高いことがわかる。すなわち、これらの商品は価格が中程度の場合には写真掲載をした方が良い評価を受けやすい傾向にある可能性がある。また、子供用品や生活用品、食べ物が多く所属する潜在クラス 4 に着目すると、比較的高価格の商品に対して写真掲載の交互作用が高いことがわかる。すなわち、これらの商品は価格が高い場合には写真掲載をした方が良い評価を受けやすい傾向にある可能性がある。

5. 考察

本研究のモデルは、テストデータに対して、補助情報から各潜在クラスへの所属確率を計算し、予測を行っている。そのため、各データの各潜在クラスへの所属確率は、補助情報に大きく依存しており、予測精度にも大きく影響を与えていると考えられる。本研究の実験では、補助情報として商品の大ジャンルのみを用いてモデルの構築を行った。しかし、本研究のモデルは補助情報の共起を考慮したモデルとなっている。そのため、例えば、男性向け商品や女性向け商品などのターゲット情報や、昔ながらの商品や最近流行している商品などのトレンド情報といった、他の補助情報をモデルに取り込むことで、さらなる予測精度の向上が期待できる。

6. まとめと今後の課題

本研究では、データの潜在的特徴を考慮し、各潜在クラス毎に FM を仮定し混合した Mixture FM を提案した。提案モデルの有効性を示すために実データを用いて実験を行い、得られたパラメータを分析することで、有用な知見が得られることを示した。今後の課題として、予測精度の向上および適切な潜在クラス数の決定方法の検討などが挙げられる。

謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより楽天株式会社から提供を受けた「楽天データセット」を利用した。

参考文献

- [1] S.Faria, and G.Soromenho, "Fitting Mixtures of Linear Regressions." *Journal of Statistical Computation and Simulation*, Vol.80, No.2, pp.201-225, 2010.
- [2] S.Rendle, "Factorization Machines," *IEEE International Conference on Data Mining*, pp.995-1000, 2010.
- [3] 永森誠矢, 山下遥, 荻原大陸, 後藤正幸, "混合回帰に基づく就職ポータルサイトの被エントリ数分析モデルに関する一考察", *情報処理学会論文誌*, Vol.59, No.4, pp.1273-1285, 2018 年 4 月.
- [4] B.Grun, and F.Leisch, "FlexMix Version 2: Finite Mixtures with Concomitant Variables and Varying and Constant Parameters." *Journal of Statistical Software*, Vol. 28, Issue 4, pp. 1-35, 2008.
- [5] 楽天株式会社 (2014): 楽天市場データ. 国立情報学研究所情報学研究データリポジトリ. (データセット). <https://doi.org/10.32130/idr.2.1>