情報推薦のための深層学習の学習データとしての嗜好データ特性に関する分析

田中　恒平†　　小林　亜樹††

† 工学院大学大学院工学研究科 電気・電子工学専攻　〒 163–8677 東京都新宿区西新宿 1–24–2
†† 工学院大学情報学部情報通信工学科　〒 163–8677 東京都新宿区西新宿 1–24–2
E-mail: †cm16029@ns.kogakuin.ac.jp, ††aki@cc.kogakuin.ac.jp

あらまし　深層学習の学習の枠組みの 1 つであるオートエンコーダが，情報推薦分野の評価値推定タスクで応用が進んでいる．嗜好データをオートエンコーダで学習させた場合，推薦値と正解値との誤差が大きくなるユーザが存在する．これらを外れ値的なユーザと捉え，大衆的なユーザの推定精度に悪影響を及ぼしている可能性がある．これに対し，我々は学習過程において大きな誤差となるユーザデータを除外しつつ学習する方式を提案した．しかし，従来のユーザデータを除外する方法では，ごく少数の評価値の誤差が小さくなるに留まり，本方式の有効性に課題がある．本研究では，外れ値的なユーザを学習段階で除外するためのアルゴリズムを拡張し，評価値推定精度の向上を目指す．実データを用いて実験を行い，既存の評価値推定手法と推定精度について比較した．

キーワード　深層学習, 情報推薦, 協調フィルタリング, オートエンコーダ

1. はじめに

EC サイトを利用してアイテムを購入する際には，推薦システムからの出力をアイテム購入の判断材料とする場合がある．推薦システムは，ユーザがアイテムに付与した評価値やサイト閲覧履歴などの履歴データやユーザやアイテムそのものに付与されたメタデータなどから，評価値が付与されていないアイテムをユーザがどのくらい好むかというユーザの嗜好を予測している．ユーザの嗜好を予測するために広く利用されているアルゴリズムが協調フィルタリングである．古典的な協調フィルタリングは，ユーザが付与した評価値のベクトル同士の類似度をピアソン相関係数やコサイン類似度などで計算し，未評価アイテムについてユーザがどのくらい好むかを推定する．これは，あるユーザが好んだアイテムを類似したユーザも同様のアイテムを好むという仮定のもとユーザの嗜好を推定している．

現在は，Matrix Factorization(MF) やニューラルネットワークといった回帰モデルでユーザの嗜好を予測する例 [1–6] が多く見受けられる．MF は，複数のユーザが複数のアイテムに対して付与した評価値を要素とした評価値行列を 2 つの行列に分解し，2 つの行列の積が元の評価値行列に近似できるように最適化を行い，未評価アイテムの予測を行う．このとき，MF は評価値が存在する要素のみを利用して分解した 2 つの行列を最適化するため，一般的に欠損値が大部分を占めると言われている嗜好データでも問題なく学習を行うことが可能である．

ニューラルネットワークでの推薦は，アイテムに付与された評価値からユーザの嗜好を予測する目的では，オートエンコーダを利用した方法が多く提案されている．そのほかにも，タイムスタンプが付与された時系列データやテキストデータを取り扱うために，RNN(Recurrent neural network) を利用し推薦を行う方法などが提案されている．

オートエンコーダは，入力とネットワークからの出力が等しくなるように，隠れ層の重み行列を最適化するニューラルネットワークの学習の枠組みである．オートエンコーダは一般的にプレトレーニングに利用されていることで知られている．プレトレーニングは，一般的に過学習を抑制する働きがあることで知られている．ここで，過学習とは．学習時に生じる学習誤差のみが重み更新により小さくなり，汎化誤差は小さくならない状態を指す．これは，未知のデータが入力された場合には，予測の精度が悪化することに相当する．これに対し，プレトレーニングでは，重みの初期値にランダムな値で初期化された重み行列ではなく，あらかじめオートエンコーダで学習をし更新された重み行列とする方法であり，これによりファインチューニングの際に良い初期値で学習を行うことが可能になり過学習を抑制できる．ニューラルネットワークでの推薦は，このプレトレーニングの枠組みを利用してユーザの嗜好を予測している．

しかし，オートエンコーダは MF と異なり，欠損値を含む入力を一般的に許容できないことで知られており，嗜好データをオートエンコーダで学習させるためには欠損値を適当な値で補完するなどの前処理を施す必要がある．これに対し，欠損値を補完することなく，オートエンコーダで学習を行うことが可能な枠組みが提案されており，その詳細は 4 章で述べる．

嗜好データを回帰的なモデルで学習させた場合には，特定のユーザのみ推定値と正解値との差が多く生じる場合が存在し，学習後のモデルが過学習を引き起こす場合もある．過学習は，隠れ層の次元数などのパラメータ設定や学習に用いるデータに外れ値的なデータが存在する場合にも起こる可能性がある．これに対し，本研究では，嗜好データ中に外れ値的なユーザが存在すると考え，これらユーザの嗜好データを学習時に除外し推定精度の向上を図る．さらに，外れ値的なユーザの嗜好データについて分析をし，どのような傾向を持つユーザの嗜好データ

が過学習を引き起こしている可能性が高いかについて考察する．

本稿の構成を以下に示す．2 章では，深層学習の枠組みを利用した情報推薦の既存研究について紹介し，本稿の位置づけを示している．3 章では，本稿における推薦の処理の定義について述べ，これをオートエンコーダで実現する方法とオートエンコーダが抱えている問題について述べる．4 章では，欠損値が大部分を占める嗜好データをオートエンコーダで学習可能な部分次元法について述べている．5 章では，部分次元法の性能について，評価値推定精度の観点から評価を行っている．また，学習時に大きな誤差を生むユーザを外れ値であると判断し，これらユーザのデータを除外しつつ学習を行う方法についても述べている．6 章では，嗜好データの一部を除外して学習を行うことによる効果を検証している．7 章では本稿のまとめを述べている．

2. 関 連 研 究

深層学習技術を利用して評価値の推定を行う場合には，オートエンコーダを利用する場合が多い．オートエンコーダは欠損値を含むデータの入力を許容しないため，従来は欠損値を適当な値で補完するなどの処理を行う必要があった．これに対し，我々が部分次元法という名前で呼んでいる欠損値を欠損値であるという情報を持たせたままオートエンコーダへの入力を可能とする方法 [7–9] が提案されている．以降のオートエンコーダを利用した評価値推定では，部分次元法を利用して嗜好データの学習を行う例が多い [10–12]．本研究においても部分次元法を適用したオートエンコーダを嗜好データの学習方法として採用する．

既存研究では，学習に用いる嗜好データにメタデータを反映させる方法や，ネットワークの構造に工夫を入れて推定精度を上げようと試みる研究が多く見受けられる．具体的には，ネットワークの隠れ層の層数を増やした積層オートエンコーダや，ガウシアンノイズなどで破損させたデータを入力とし，破損させる前のデータを教師信号とするデノイジングオートエンコーダが採用される場合が多い．

しかし，オートエンコーダで学習させるデータそのものである嗜好データの性質についての議論がなされている例は見受けられず，データセットに存在するすべての嗜好データを用いて学習を行う場合がほとんどである．これは，MF を利用して評価値の推定を行う既存研究についても同様である．

本研究では，一部外れ値的なユーザの嗜好データが存在すると考え，これらユーザを学習時に除外してオートエンコーダで学習を行った場合の評価値推定精度について検証する．

3. オートエンコーダを用いた情報推薦

3.1 推 薦 処 理

本節では，本研究における情報推薦の処理について述べる．本研究における情報推薦の処理は，図 1 のように欠損要素を含む n 次元のベクトルを推薦器へ入力すると，欠損要素がなんらかの値で埋められたベクトルが出力されることに相当する．ここで，推薦器への入力となる n 次元のベクトルは，ユーザ i が

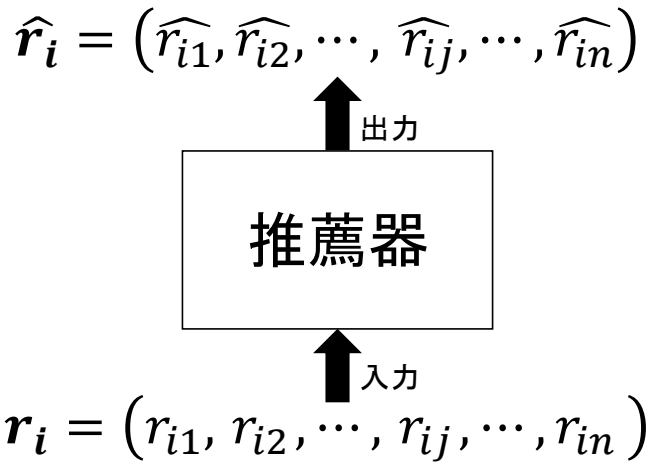

図 1 推 薦 処 理

n 個のアイテムに対して付与した評価値を要素にもつベクトルであり，これを評価値ベクトル $\boldsymbol{r_i}$ と呼ぶ．このとき，ユーザ i がアイテム j に評価値を付与していない場合が存在すると考えられ，この場合には $\boldsymbol{r_i}$ の要素である r_{ij} が欠損値となる．この欠損値を含む評価値ベクトル $\boldsymbol{r_i}$ を推薦器へ入力する．

推薦器は，複数ユーザの嗜好データの特徴が反映されるように学習を行ったモデルに相当する．具体的には，MF やオートエンコーダのような回帰的なモデルの重みに嗜好データの特徴が反映され，重みを用いて推薦器からの出力を決定する．

推薦器からの出力は，入力の $\boldsymbol{r_i}$ の欠損要素がなんらかの値で埋められたベクトルであり，これを推定値ベクトル $\hat{\boldsymbol{r_i}}$ と呼ぶ．

本研究では，推薦器にオートエンコーダを利用する．次節で嗜好データを学習するための枠組みとして利用しているオートエンコーダについて説明する．

3.2 オートエンコーダ

オートエンコーダとは，ニューラルネットワークの学習の枠組みの 1 つである．ニューラルネットワークの学習は，出力と教師信号との誤差を 2 乗誤差といった損失関数で定義し，損失関数の出力を最小化するように重み行列を更新することである．ニューラルネットワークの中でもオートエンコーダは，入力となったデータそのものを教師信号とする学習手法である．これは入力と出力が等しくなるように重み行列を更新することに相当する．

ニューラルネットワークの学習は，ネットワークへ入力されたデータから出力を得るフォワードプロパゲーションと，出力と教師信号との誤差から隠れ層の重み行列を更新するバックプロパゲーションの 2 つの段階からなる．以下に，情報推薦分野にオートエンコーダを利用した場合においてのフォワードプロパゲーション，バックプロパゲーションそれぞれの段階の処理について述べる．

3.2.1 フォワードプロパゲーション

フォワードプロパゲーションでは，評価値ベクトル $\boldsymbol{r_i}$ をオートエンコーダへ入力し，出力層の出力である推定値ベクトル $\hat{\boldsymbol{r_i}}$ を得ることが目的となる．オートエンコーダへユーザ i の評価値ベクトル $\boldsymbol{r_i}$ を入力したとき，出力となる推定値ベクトル $\hat{\boldsymbol{r_i}}$

は式 (1) のように定義される．

$$\hat{\boldsymbol{r}}_{\boldsymbol{i}} = g(\boldsymbol{W'} f(\boldsymbol{W}\boldsymbol{r}_{\boldsymbol{i}} + b_1) + b_2) \tag{1}$$

ただし，$\boldsymbol{W}$，b_1，f はそれぞれ入力層-隠れ層間 (エンコーダ) の重み行列，バイアスユニット，活性化関数を示し，$\boldsymbol{W'}$，b_2，g はそれぞれ隠れ層-出力層間 (デコーダ) の重み行列，バイアスユニット，活性化関数を示す．ここで，エンコーダとデコーダの重み行列は tied weight とし，$\boldsymbol{W'} = \boldsymbol{W}^{\mathbf{T}}$ とする．また本研究においては，エンコーダ，デコーダどちらのバイアスユニットにも値は設定していないため，以降の説明では省略する．活性化関数は，シグモイド関数や ReLU，恒等関数などがこれに該当する．本研究では，エンコーダの活性化関数 f をシグモイド関数，デコーダの活性化関数 g を恒等関数とした．

得られた $\hat{\boldsymbol{r}}_{\boldsymbol{i}}$ と教師信号である入力の評価値ベクトル $\boldsymbol{r}_{\boldsymbol{i}}$ との誤差を最小化するためのバックプロパゲーションについて次節で述べる．

3.2.2 バックプロパゲーション

バックプロパゲーションでは，フォワードプロパゲーションにより得られた出力と教師信号との誤差 E を最小化するように隠れ層の重み行列を更新することが目的となる．このとき，オートエンコーダは入力と出力が等しくなるように重み行列を更新する学習の枠組みであるため，教師信号は入力の評価値ベクトルとなる．E は出力と教師信号との誤差を損失関数で定義したものである．本研究における損失関数は，入力である $\boldsymbol{r}_{\boldsymbol{i}}$ と出力 $\hat{\boldsymbol{r}}_{\boldsymbol{i}}$ の平均二乗誤差としており，式 (2) のように定義される．

$$E = \frac{1}{n}\sum(\boldsymbol{r}_{\boldsymbol{i}} - \hat{\boldsymbol{r}}_{\boldsymbol{i}})^2 \tag{2}$$

ただし，n は入出力の次元数に相当する．この E を最小化するために隠れ層の重み行列を更新する必要があり，これには E を偏微分することによって得られる勾配ベクトル $\boldsymbol{g}$ が必要になる．勾配ベクトル $\boldsymbol{g}$ は，E を各次元ごとに偏微分をすることによって得ることができる．

$$\boldsymbol{g} = \sum(\frac{\partial E}{\partial \boldsymbol{W'}}) \tag{3}$$

$$\boldsymbol{g} = \boldsymbol{r}_{\boldsymbol{i}} - \hat{\boldsymbol{r}}_{\boldsymbol{i}} \tag{4}$$

その後，得られた勾配ベクトル $\boldsymbol{g}$ と隠れ層の出力ベクトル $\boldsymbol{v}$ からデコーダにおける重み更新量 $\Delta W'$ を求め，$\boldsymbol{v}$ と入力となる評価値ベクトル $\boldsymbol{r}_{\boldsymbol{i}}$ からエンコーダにおける重み更新量 ΔW を求める．式 (5)，式 (6) はデコーダにおける重み更新量と更新式であり，式 (7)，式 (8) はエンコーダにおける重み更新量と更新式である．ただし η は学習率である．

$$\Delta \boldsymbol{W'} = \boldsymbol{g}\boldsymbol{v} \tag{5}$$

$$\boldsymbol{W'} = \boldsymbol{W'} - \frac{\eta}{n}\Delta \boldsymbol{W'} \tag{6}$$

$$\Delta \boldsymbol{W} = \boldsymbol{v}\boldsymbol{r}_{\boldsymbol{i}} \tag{7}$$

$$\boldsymbol{W} = \boldsymbol{W} - \frac{\eta}{n}\Delta \boldsymbol{W} \tag{8}$$

3.3 オートエンコーダによる推薦

深層学習の枠組みを利用したオートエンコーダを情報推薦の処理に用いる．オートエンコーダにより推薦を行うためには，複数ユーザの嗜好データを入力として学習を行う必要がある．このとき，学習とは，重み行列 $\boldsymbol{W}$ および $\boldsymbol{W'}$ に入力となった嗜好データの特徴が反映されることに相当し，これは出力の推定値ベクトル $\hat{\boldsymbol{r}}_{\boldsymbol{i}}$ と教師信号である入力の評価値ベクトル $\boldsymbol{r}_{\boldsymbol{i}}$ を引数にとった損失関数の出力を最小化することに等しい．嗜好データの特徴が反映され学習が進んだオートエンコーダに対し欠損要素を含む評価値ベクトル $\boldsymbol{r}_{\boldsymbol{i}}$ を入力すると，出力層では入力の欠損要素がなんらかの値で埋められた推定値ベクトル $\hat{\boldsymbol{r}}_{\boldsymbol{i}}$ が出力される．この $\hat{\boldsymbol{r}}_{\boldsymbol{i}}$ を推薦値として取り扱う．図 2 の例で

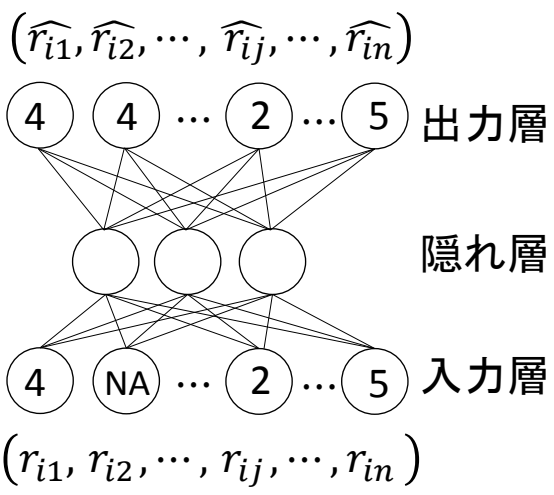

図 2　オートエンコーダによる推薦

は $\boldsymbol{r}_{\boldsymbol{i}}$ のうち，r_{i2} が欠損値となっており，これら未評価要素の評価値をオートエンコーダからの出力で推定することが推薦の目的となる．

しかし，オートエンコーダは欠損を含む入力を許容しない問題が存在し，単純には嗜好データをオートエンコーダへの入力とすることはできない．オートエンコーダが情報推薦の分野において抱えている問題について次節で述べる．

3.3.1 オートエンコーダの問題点

オートエンコーダは，一般的に欠損値を含む入力を許容することができない．情報推薦の分野で取り扱う嗜好データは欠損値が大部分を占めており，生データのままではオートエンコーダへの入力とすることができないという問題が存在する．この問題に対しての 1 つの解は，欠損値を入力データから取り除くことであり，これは欠損値を含むデータを削除するリストワイズ法の適用や，欠損値を適当な値で補完する方法で実現できる．しかし，上述したように嗜好データは欠損値が大部分を占めるという理由からリストワイズ法の適用はほぼ不可能である．そこで，筆者らは欠損値をユーザが付与した評価値の平均値や評価値定義域の中央値で補完するなどの簡易的な手法でオートエンコーダへの入力を可能にし，嗜好データの欠損値要素の評価値を推定した [13]．その結果，欠損値を補完し嗜好データの学習を行った場合には，未評価要素には補完した値に近い値が推

定されており，これがユーザの嗜好を反映し評価値を推定しているとは言い難い．さらに，推定精度においても MF といった既存の評価値推定手法に劣る結果が得られた．

これに対し，筆者らが部分次元法と呼んでいる欠損値を欠損値であるという情報を残したままオートエンコーダへ入力することが可能な手法 [7–9] が提案されている．部分次元法について次章で説明する．

4. 部分次元法による推薦

オートエンコーダは，一般的に欠損値を含む入力を許容することができない．そのため，大部分が欠損値である嗜好データをオートエンコーダへの入力とするためには，欠損値を適当な値で補完するなどの方法をとる必要があった．これに対し部分次元法は，欠損値を適当な値で補完することなく嗜好データをオートエンコーダへ入力することが可能である．以下に，部分次元法におけるフォワードプロパゲーション，バックプロパゲーションそれぞれの処理について説明する．

4.1 フォワードプロパゲーション

フォワードプロパゲーションでは，推定値ベクトル $\hat{\boldsymbol{r}}$ を得るためにユーザ i の評価値ベクトル $\boldsymbol{r_i}$ をオートエンコーダへ入力するが，$\boldsymbol{r_i}$ に欠損値が存在する場合にはオートエンコーダへ入力することができない．これに対し，部分次元法では，$\boldsymbol{r_i}$ の要素のうち，欠損値となっている要素 j が存在するとき，r_{ij} の値は 0 と置き換える．欠損値を 0 と置き換えることにより，$\boldsymbol{r_i}$ に欠損値が存在しなくなるため，オートエンコーダへの入力が可能になる．さらに，$\boldsymbol{r_i}$ の欠損値要素を 0 と置き換えていたことにより，ユーザが付与した評価値のみが出力に寄与するようになる．

4.2 バックプロパゲーション

バックプロパゲーションでは，$\boldsymbol{r_i}$ に欠損値が存在した場合の処理と，最小化する対象である損失関数の取り扱いについて述べる．

$\boldsymbol{r_i}$ のうち r_{ij} が欠損値を示す 0 が入力されていた場合においても，推定値である $\hat{r_{ij}}$ との誤差が生じる．誤差が生じるということは，対応する $\boldsymbol{W}$ の要素を更新するということになり，これは存在しないデータを用いて学習を進めていることと同義であり好ましくない．この問題に対処するために部分次元法では，r_{ij} に欠損値を示す 0 が入力されていた場合には，推定値である $\hat{r_{ij}}$ も 0 と置き換える．これにより，入出力間の誤差が無くなり，ユーザが付与した評価値のみで学習を行うことが可能になる．

ニューラルネットワークでは，損失関数の出力がより小さくなるように重み行列 $\boldsymbol{W}$ を更新する．$\boldsymbol{W}$ は，損失関数の出力が大きい程，大きく更新される．このとき，部分次元法を用いた場合には，採用する損失関数によっては出力値が小さく算出されてしまう問題が存在する．本研究で用いている損失関数である，式 (9) で定義される平均二乗誤差で説明する．

$$L = \frac{1}{n}\sum(\boldsymbol{r_i} - \hat{\boldsymbol{r_i}})^2 \tag{9}$$

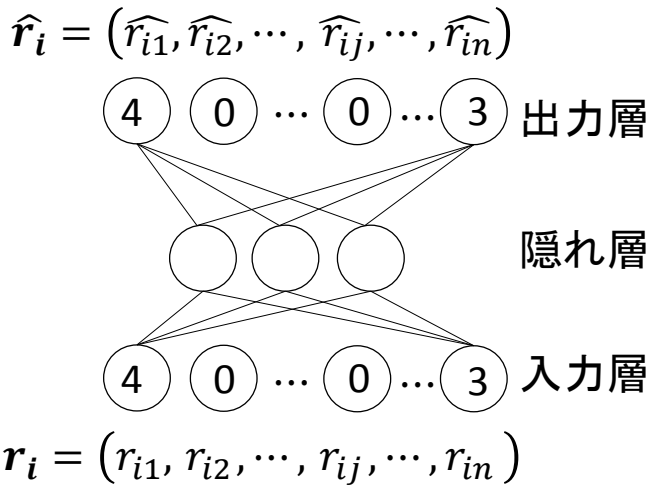

図 3 部分次元法

ただし，n は $\boldsymbol{r_i}$ の入出力の次元数である．ここで，部分次元法は，$\boldsymbol{r_i}$ のある要素 r_{ij} が欠損値であった場合には 0 と置き換え，さらに対応する出力 $\hat{r_{ij}}$ の値も 0 と置き換えることで入出力の差分を無くし，学習を行う方法であった．このとき，式 (9) の n の値が入出力の次元数である場合には，学習速度の低下などの問題を引き起こす．これは，部分次元法の学習において，入出力の誤差を 0 とする次元，すなわち欠損値を示す 0 が入力となった次元も n の値に含まれており，これが損失関数の出力が小さく算出する原因になり得る．

これに対し，n の値を入出力の次元数とするのではなく，ユーザ i の評価値ベクトル $\boldsymbol{r_i}$ のうち，評価値が存在する要素数 n_i とすることで，部分次元法による学習をミニバッチ学習も含め行うことができる．

$$L = \frac{1}{n_i}\sum(\boldsymbol{r_i} - \hat{\boldsymbol{r_i}})^2 \tag{10}$$

本研究では，この部分次元法を利用して評価値の推定を行う．

5. 実験：部分次元法の性能検証

5.1 目　　的

部分次元法を適用したオートエンコーダの評価値推定精度について検証する．また，学習に用いた嗜好データについて，学習誤差が大きくなるユーザの嗜好データがどのような傾向を持っているかについて考察する．

5.2 条　　件

使用したデータセットは MovieLens1M である．MovieLens1M は，ユーザが映画に対して 5 段階評価を行った記録を収集したデータセットである．ユーザ数は 6,040 人，アイテム数が 3,952 であり評価値の数は 1,000,209 存在する．評価値以外のメタデータに，ユーザの性別や年齢，映画のジャンルなど存在するが，本研究ではこれらのメタデータを用いていない．交差検定のため，データセットに存在する評価値の 8 割を学習用データ，残りの 2 割をテスト用データに分割する．データの分割は，ユーザ単位で行うのではなく，付与された評価値 1 つ 1 つを学習用，テスト用に分割する．これにより，テスト用

データは学習用データの評価値を隠したものに相当する．実験では，テスト用データを学習後のモデルへ入力し隠した評価値を推定させる．

また，嗜好データをオートエンコーダへ入力する前に，ユーザごとの評価値の偏りを緩和するための事前処理を行う．具体的には，ユーザ i の評価値ベクトル $\boldsymbol{r_i}$ の欠損値でない要素から，$\boldsymbol{r_i}$ の欠損要素を除いた平均値 $\overline{r_i}$ をあらかじめ引くという処理を行う．この処理が行われたベクトル $\widetilde{\boldsymbol{r_i}}$ がオートエンコーダへの実際の入力となる．

$$\widetilde{\boldsymbol{r_i}} = \boldsymbol{r_i} - \overline{r_i} \tag{11}$$

パラメータ設定について説明する．部分次元法を適用した隠れ層が 1 層のみの 3 層オートエンコーダの隠れ層の次元数は 200 とした．バッチサイズは 10 とした．損失関数である入出力間の平均 2 乗誤差を最小化するための最適化アルゴリズムは確率的勾配降下法とし，学習率は 0.1 とした．学習回数は 1000 とした．

5.3 評　　価

評価値の推定精度は RMSE で評価する．

$$\mathrm{RMSE} = \sqrt{\frac{1}{k}\sum(\boldsymbol{T} - \hat{\boldsymbol{R}})^2} \tag{12}$$

ここで，テスト用のデータは学習用データの評価値を隠したデータであり，これを推定することが目的であるため，k は学習用データの評価値数となり，$\hat{\boldsymbol{R}}$ は推定評価値行列であり，$\boldsymbol{T}$ は正解値行列である．RMSE は値が小さいほど良い結果である．

5.4 実 験 結 果

図 4 に 5.2 節のパラメータ設定で学習させたモデルの学習曲線を示す．グラフの横軸は学習回数を示し，縦軸は RMSE の値を示している．RMSE は値が小さい程良い結果であるため，実験結果から学習回数が増加するにしたがって推定精度が良くなっていることがわかる．

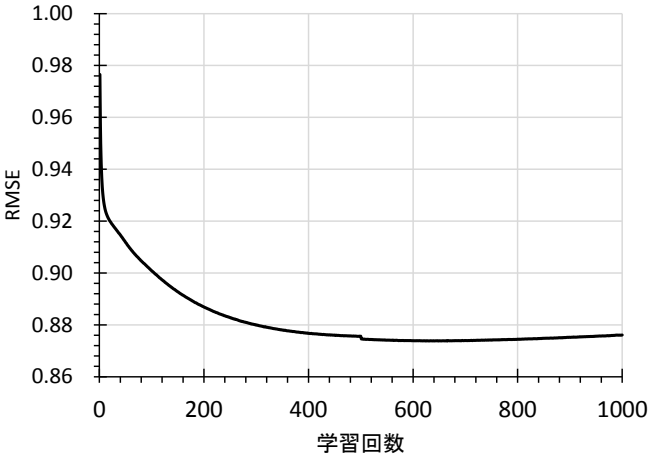

図 4　部分次元法の推定精度

5.5 考　　察

図 4 の学習曲線は，学習回数が 625 のときに最小の RMSE0.873 をとる．このとき，学習が進んでいるユーザとそうでないユーザとが存在することが予想される．ここで，ユーザごとの学習誤差を得るために，学習回数が 625 のモデルにおけるユーザ i の推定評価値ベクトル $\hat{\boldsymbol{r}}_{\boldsymbol{i}}$ と正解値ベクトル $\boldsymbol{t_i}$ との二乗平方根誤差の平均値 $\overline{e_i}$ を算出する．

$$\overline{e_i} = \frac{1}{n_i}\sqrt{\sum(\boldsymbol{t_i} - \hat{\boldsymbol{r}}_{\boldsymbol{i}})^2} \tag{13}$$

n_i はユーザ i の評価値ベクトル $\boldsymbol{r_i}$ の欠損値でない要素の数を示している．$\overline{e_i}$ はすべてのユーザについて算出し，その分布を図 5 に示す．

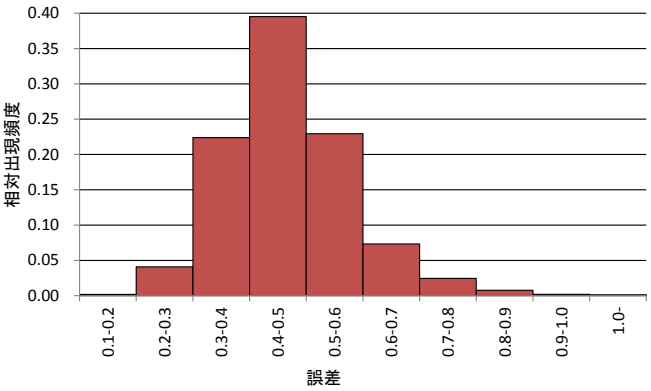

図 5　学習誤差分布

本研究では，すべてのユーザの誤差平均 $\overline{e}$ よりも，大きな誤差平均 $\overline{e_i}$ となったユーザ (以下，外れ値ユーザと呼ぶ) を学習時においての誤差が大きいと判断し，これらユーザについて分析する．

はじめに，オートエンコーダへの入力となる評価値ベクトル $\boldsymbol{r_i}$ の分散値をすべてのユーザについて算出し，分散の分布を図 6 に示す．このグラフは，青色がすべてのユーザに対しての

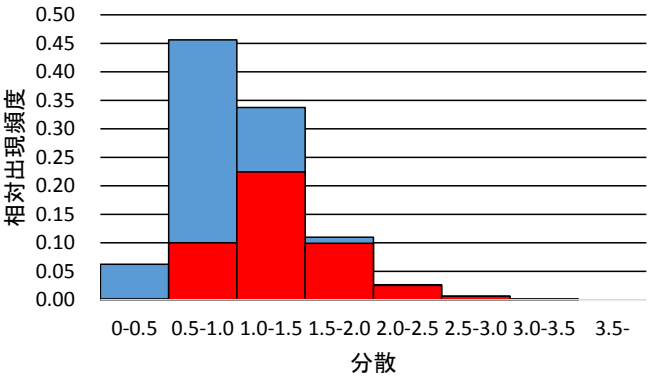

図 6　分 散 分 布

分散の出現頻度を示しており，赤色は外れ値ユーザの分散の出現頻度を示している．また，すべてのユーザの分散の平均値は 1.05 ほどであった．図 6 から，外れ値ユーザの評価値ベクトルは，1 割のユーザを除いて分散が平均値よりも大きく，分散値が増加するほど誤差が大きなユーザである割合が増加していることが見て取れる．これは，5 段階評価を例にすると，1 や 5 といった極端な評価値を多く付与するユーザは分散値が大きく，誤差も大きいという傾向を示していると考えられる．

これら外れ値ユーザが，誤差が小さなユーザ (以下，大衆ユー

ザと呼ぶ) の学習を阻害している可能性があると考え，分散が平均値よりも大きなユーザを学習用データから除外して学習を行う方法について実験を行う．

6. 実験：データ除外の効果

6.1 目　　的

評価値ベクトルの分散が平均値よりも大きなユーザの評価値ベクトルを学習用のデータから除外して学習を行ったモデルについて，既存の評価値推定の枠組みである Non-negative Matrix Factorization(NMF) と推定精度及び誤差の分布について比較する．

6.2 条　　件

使用したデータセットは 5.2 節と同様に MovieLens1M である．部分次元法による学習では，式 (11) で定義された評価値の偏りを緩和するための事前処理を行うが，NMF は非負値のみで学習を行うため嗜好データの事前処理は行わない．本実験においてもデータセットを分割することによる交差検定を行う．分割の割合は 5.2 節と同様に評価値の 8 割を学習用データ，残りの 2 割をテスト用データとした．その後，学習用データのみ嗜好データの分散に基づいてデータの分割を行う．具体的には，評価値ベクトルの分散が小さな大衆ユーザのみで構成されたデータセットと分散が大きな外れ値ユーザのみで構成されたデータセットの 2 つに分割する．このとき，大衆ユーザと外れ値ユーザのユーザ数はそれぞれ 3275 人，2765 人となった．実験では，学習用データのすべてを用いて学習を行う方法 (baseline 法)，大衆ユーザのみのデータセットで学習を行う方法，外れ値ユーザのみで学習を行う方法，そして既存の推薦の枠組みである NMF を用いた方法の 4 手法について推定精度を比較した．

パラメータ設定について説明する．baseline 法は 5.2 節と同様に隠れ層の次元数は 200，バッチサイズが 10，学習率を 0.1 とした．これに対しデータの除外し学習を行う方法では，隠れ層の次元数を 100，バッチサイズを 20 とし，学習率は 0.15 とした．NMF のパラメータ設定は，潜在変数の数を 5 とし，最小化すべき損失関数は KL-divergence とした．実験結果では，最も RMSE が小さくなった学習回数 85 の推定精度を示してある．

6.3 評　　価

推定精度の評価は，5.3 節と同様に RMSE で行う．データセットから除外などの処理を行っていないテスト用データを用いて，各手法で作成したモデルの評価を行う．このとき，テスト用データは学習用データの評価値を隠したものに相当し，実験では隠した評価値を推定させる．そのため，学習用データからデータの除外を行っているかに関わらず，実験にて推定すべき評価値は隠された学習用データの評価値となる．

6.4 実 験 結 果

図 7 は，データを除外せずに学習を行った結果 (図 4 の実験結果と同様)，大衆ユーザのみで学習を行った結果，外れ値ユーザのみで学習を行った結果，そして NMF で嗜好データを学習し評価値推定した結果を示している．図 8 は他の手法で嗜好データの除外を行った場合の実験結果 [14] である．具体的には，評価値の分散から除外するユーザを決定するのではなく，あらかじめオートエンコーダによる学習を行い，その際に生じる学習誤差から除外するユーザを決定する．除外するユーザの決定の方法は，図 5 のような誤差分布において，あるユーザ i の誤差平均 $\overline{e_i}$ が $\mu + \alpha\sigma$ の値を超えていた場合，ユーザ i の嗜好データを学習時に除外する．このとき，μ は全てのユーザの誤差平均の平均値であり，σ は全てのユーザの誤差平均の標準偏差である．

$$\mu = \frac{1}{N}\sum \overline{e_i} \tag{14}$$

$$\sigma = \sqrt{\frac{1}{N}\sum(\mu - \overline{e_i})^2} \tag{15}$$

α は除外するユーザ数を決定するためのパラメータであり，図 8 には $\alpha = 2$ と $\alpha = 3$ の条件でユーザを除外した結果を載せている．N はデータセットに存在するユーザの数である．除外ユーザ数は $\alpha = 2$ 場合には 230 人ほど，$\alpha = 3$ の場合には 60 人ほどとなった．

また，図 9 から図 11 は，RMSE が最も小さな値をとった学習回数のモデルにおいて，推定した評価値と正解値との誤差についての出現頻度を示しており，図 9 は大衆ユーザのみで学習した方法の誤差出現頻度からデータを除外せずに学習を行った場合の誤差出現頻度を引いた結果を示している．同様に図 10，図 11 はそれぞれ，外れ値ユーザのみで学習した方法の誤差出現頻度から大衆ユーザのみで学習した方法の誤差出現頻度を引いた結果と，外れ値ユーザのみで学習した方法の誤差出現頻度から NMF での誤差出現頻度を引いた結果を示している．

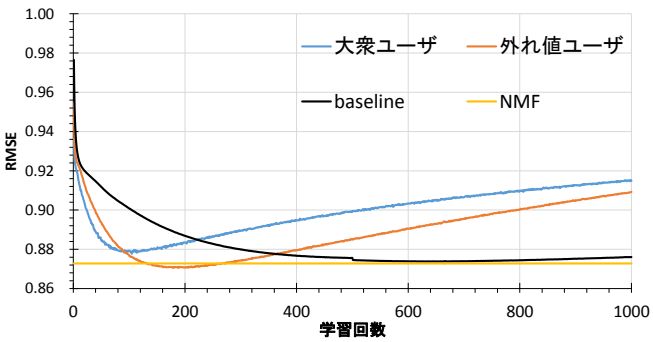

図 7　4 手法の学習曲線

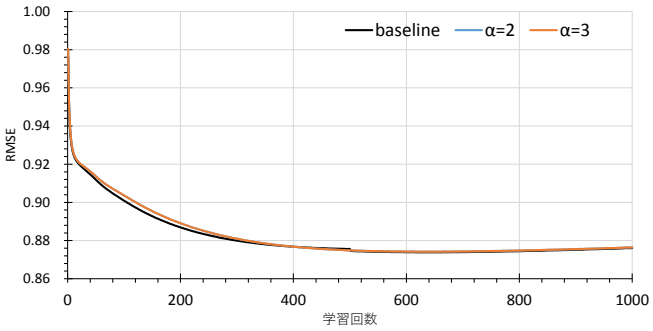

図 8　他の除外法での結果

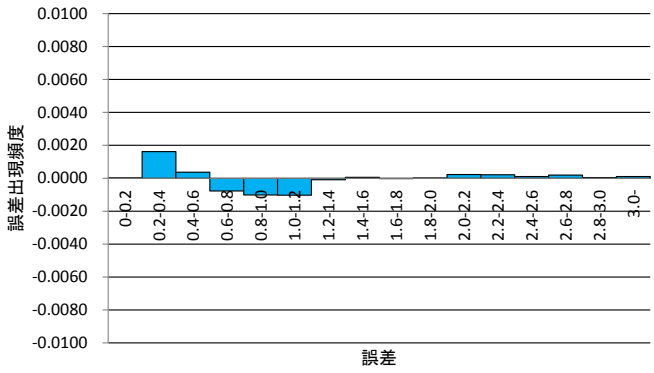

図 9　大衆ユーザ-baseline

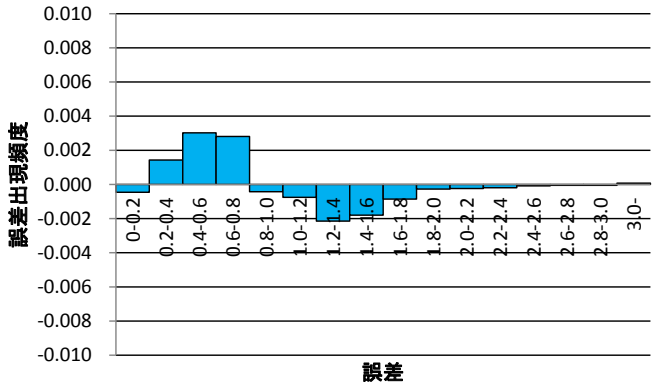

図 10　外れ値ユーザ-大衆ユーザ

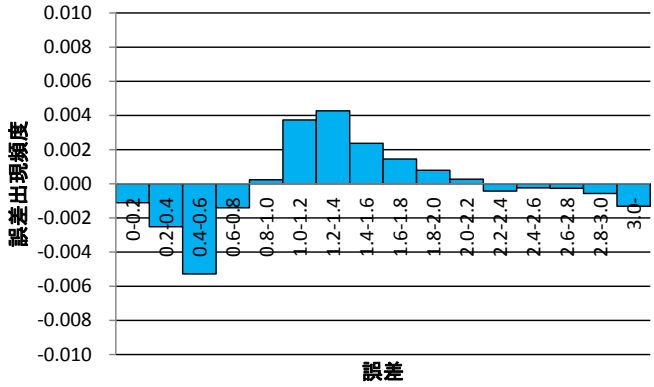

図 11　外れ値ユーザ-NMF

6.5 考　　察

図 7 の実験結果は，それぞれの条件で学習率や学習に用いるユーザ数が異なるため，推定精度についてのみ比較する．実験結果から，最も評価値の推定精度が高い手法であるのは外れ値ユーザのみで学習を行った場合である．すべてのユーザの嗜好データを用いて学習を行う場合よりも推定精度が良いことから，学習に有益なユーザとそうでないユーザとが存在し，学習に有益でないユーザが占める割合によっては学習がうまく進まないのではないかと考える．これにより，嗜好データをオートエンコーダで学習させる場合には，データの取捨選択について考える必要があるといえる．

図 8 は，学習時に誤差が大きくなったユーザの嗜好データを除外し，再度学習を行う方法についての実験結果である．実験結果から，除外前後で推定精度はほとんど変化していないことが見て取れる．これは，除外したユーザ数が全体の 3%程であり，少数のユーザしか除外しなかったことによるものだと考える．あるいは，そもそもこれら誤差が大きなユーザの嗜好データが，学習に寄与していないことも考えられる．これはミニバッチ学習により隠れ層の重みを更新することによる効果であると考えられる．今回の実験では，除外ユーザ数が全体の 3%ほどであり，バッチのユーザのほとんどが大衆ユーザであることから，外れ値ユーザの嗜好データを除外せずとも大衆ユーザに向けた最適化が行われるのではないかと考える．

図 9 は大衆ユーザのみの嗜好データを用いて学習したモデルと嗜好データを除外せずに学習したモデルの誤差頻度の差を示している．結果として，推定精度はデータ除外を行わない方法の方が良い一方で，大衆ユーザのみで学習を行った場合の方が誤差が小さな要素が増加していることが見て取れる．

図 10 では，外れ値ユーザのみで学習を行った場合において，大衆ユーザのみで学習を行う場合と比較して誤差が小さな要素が増加していることが見て取れる．

図 11 から NMF は誤差が小さな要素が，除外するか否かに関わらずオートエンコーダで学習した場合と比較して多いが，誤差が大きな要素は外れ値ユーザのみで学習させた方法よりも多いことが見て取れる．これにより，NMF は評価値が正確に推定できる場合とそうでない場合とが存在する可能性がある．

図 9 から図 11 では，全ての嗜好データを用いて学習を行うよりも，分散に基づいて一定数の嗜好データを除外し学習を行う方が，誤差が小さな要素が増加する傾向にある．これらの結果から，付与された評価値の分散が大きな外れ値ユーザの嗜好データは，オートエンコーダの学習に有用であり精度向上をもたらしているといえる．

7. おわりに

本稿では，学習段階において誤差が大きくなるユーザの嗜好データについて分析した．その結果，誤差が大きくなるユーザは，オートエンコーダへの入力となる評価値ベクトルの要素の分散が大きいという傾向が存在することを示した．本稿では，この傾向を利用し，評価値ベクトルの要素の分散が大きなユーザを除外し学習を行うことにより推定精度の向上が図れると考え，実験を行った．その結果，外れ値と呼んでいた分散が大きな嗜好データのみで学習を行う場合の方がより推定精度が高いという結果が得られた．現状では，なぜ分散が大きな嗜好データのみで学習を行うほうが，全ての嗜好データを用いる場合と比較して推定精度が良くなるのか不明である．

今後の課題は，分散が大きな嗜好データのみで学習を行うことによる効果について検証することである．さらに，全てのデータを用いた場合には，現在利用している単一の隠れ層のみが存在するネットワークで学習を行うことが困難であることが示唆されたため，ネットワークの構造を変えるなど学習のモデルそのものを変更する必要がある．

文　　献

[1] Sheng Zhang, Weihong Wang, James C Ford, and Fillia Makedon, “Learning from Incomplete Ratings Using Non-

negative Matrix Factorization", Proceedings of the Sixth SIAM International Conference on Data Mining, April 20-22, 2006, Bethesda, MD, USA

[2] R. Salakhutdinov and A. Mnih, "Probabilistic Matrix Factorization" Advances in Neural Information Processing Systems 20 (NIPS 07), ACM Press, pp. 1257-1264, 2008.

[3] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems" IEEE Computer 42 (8), pp.4249, 2009.

[4] X. Luo, M. Zhou, Y. Xia, and Q. Zhu, "An Efficient Non-Negative Matrix-Factorization-Based Approach to Collaborative Filtering for Recommender Systems" IEEE Transactions on Industrial Informatics Vol.10 No.2 pp.1273-1284, 2014.

[5] H. Wang, N. Wang, and D. Y. Yeung, "Collaborative Deep Learning for Recommender Systems" In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD' 15, pp.12351244, 2015.

[6] S. Li, J. Kawale, Y. Fu, "Deep Collaborative Filtering via Marginalized Denoising Auto-encoder" In Proceedings of CIKM' 15, pp.811820. ACM, 2015.

[7] S. Sedhain, A. K. Menon, S. Sanner, and L. Xie, "Autorec: Autoencoders meet collaborative filtering" Proceedings of the 24th International Conference on World Wide Web, pp. 111-112, 2015.

[8] Florian Strub, Jérémie Mary, "Collaborative Filtering with Stacked Denoising AutoEncoders and Sparse Inputs" NIPS Workshop on Machine Learning for eCommerce, Dec 2015, Montreal, Canada.

[9] 田中恒平, 小林亜樹, "評価値推定タスクにおける推定評価値の評価に関する検討", DEIM2017, B1-1, 2017.

[10] Florian Strub, Romaric Gaudel, Jérémie Mary, "Hybrid Recommender System based on Autoencoders" In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. ACM, pp. 11-16, 2016.

[11] B. Yi, X. Shen, Z. Zhang, J. Shu and H. Liu, "Expanded autoencoder recommendation framework and its application in movie recommendation" 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA), pp. 298-303, 2016.

[12] Yosuke Suzuki, Tomonobu Ozaki, "Stacked Denoising Autoencoder-Based Deep Collaborative Filtering Using the Change of Similarity" In Advanced Information Networking and Applications Workshops (WAINA), 2017 31st International Conference on. IEEE, pp. 498-502.

[13] 田中恒平, 小林亜樹, "深層学習を用いた情報推薦のための欠損値補完手法", DEIM2016, C7-4, 2016.

[14] 田中恒平, 小林亜樹, "嗜好データによる深層学習を用いた情報推薦におけるデータ選別と正確性に関する検討" 第 11 回 Web インテリジェンスとインタラクション研究会予稿集, 2017-11-16.