

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

田中 恒平 小林 亜樹

工学院大学大学院工学研究科 工学院大学情報学部情報通信工学科

cm16029@ns.kogakuin.ac.jp aki@cc.kogakuin.ac.jp

概要 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、オートエンコーダからの出力である推薦値と正解値との誤差が大きくなるユーザがあり、これが全体の推定精度に悪影響を及ぼしている。

そこで本研究では、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する。実データを用いた実験で、除外したユーザデータが学習に悪影響を及ぼしていることを示す。

キーワード 深層学習, 情報推薦, オートエンコーダ, データ除外

1 はじめに

ECサイトを利用してアイテムを購入する場合には、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ことがある。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嗜好データから、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り、特に 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。また、MFは評価値が存在する要素のみで分解した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。情報推薦の分野では、一般的にユーザが付与可能な評価値は正の値であることから、Non-negative Matrix Factorization(NMF) も利用されている。

ニューラルネットワークでの推薦は、オートエンコーダを利用した方法が提案されている。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学習の枠組みである。本研究においても、オートエンコーダをユーザの嗜好を予測するためのアルゴリズムとして用いる。しかし、オートエンコーダはMFとは異なり、欠損値の入力を一般的に許容できないことが知られており、欠損値をなんら

かの値で補完するなどの前処理を施してから学習をさせる必要があった。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、その詳細を次章で述べる。

また、オートエンコーダで嗜好データを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ことを確認している。これに対し、本稿では、誤差が多く生じているユーザを学習の段階で除外する方法を提案する。

本稿の構成は次の通りである。第2章では、嗜好データをオートエンコーダで学習させるための既存研究を紹介する。第3章では、嗜好データをオートエンコーダで学習させるための部分次元法について詳細を述べる。第4章では、部分次元法での学習に悪影響を及ぼしていると考えられるユーザデータを除外する方法について述べる。

2 関連研究

深層学習の枠組みを利用する方法では、嗜好データをオートエンコーダで学習させる方法を用いる場合が多い。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、従来は欠損値をなんらかの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と呼んでいる、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 [1][2][3] が提案されている。さらに、[1][2][3]の方法を利用することに加え、入力となる嗜好データにユーザの年齢や性別などのメタデータを反映させる方法について検討したという報告 [4] もある。

ネットワークの構成を工夫している例も見受けられる。学習の方法は部分次元法を利用したオートエンコーダで、出力層を2つに分岐させているような構造を採用

している例 [5] がある。この研究での出力層は、入力 of 嗜好データを再構成させたデータを出力するものと、アイテムがどのジャンルに属しているかを分類するためのものがある。この2つの出力から生じた誤差を共通の重み行列に逆伝播させる方式を採用している。

また、[2] のアルゴリズムで嗜好データを学習させたモデルに対して嗜好データを入力し、隠れ層における複数ユーザの出力ベクトルそれぞれから類似度を算出し評価値推定を行う方法も提案されている [6]。

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研究や、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い。本研究では、学習させる嗜好データの取り扱い方について検討する。具体的には、入出力間の誤差行列からオートエンコーダでの学習に悪影響を及ぼしているユーザを特定し、それらユーザを除外して学習を行うアルゴリズムを提案する。部分次元法については次章で詳細を述べる。

3 部分次元法

はじめに、本研究における情報推薦の処理について述べる。図1のようなオートエンコーダに対して、ユーザ i の評価値ベクトル \mathbf{r}_i を入力する。このとき、 \mathbf{r}_i の要素は、ユーザ i がアイテム j に対して付与した評価値に相当する。従って、アイテム j が未評価である場合には、 r_{ij} は欠損値となる。図1では、 \mathbf{r}_i のうち、 r_{i2} と r_{i3} が欠損値となっている。この評価値ベクトルを推薦器であるオートエンコーダへ入力すると、出力層では欠損部分が埋められ、推定値ベクトル $\hat{\mathbf{r}}_i$ が出力される。この $\hat{\mathbf{r}}_i$ を推薦値として取り扱う。

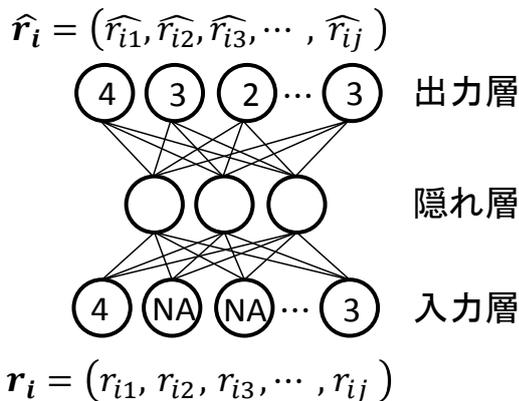


図1 情報推薦の処理

しかし、通常の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、欠損値をなんらかの値で補完するなどの方法をとる必要がある。これに対し、部分次元法は、欠損値を補完することなくオートエンコーダへの入力が可能な手法である。以下、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。

フォワードプロパゲーションでは、入力である評価値ベクトル \mathbf{r} と隠れ層の重み行列 \mathbf{W} から出力層における推定値ベクトル $\hat{\mathbf{r}}$ を得ることが目的となる。推定値ベクトル $\hat{\mathbf{r}}$ を得るために、ユーザ i の評価値ベクトル \mathbf{r}_i をオートエンコーダへ入力する。このとき、 \mathbf{r}_i の要素のうち、欠損値となっているアイテム 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で置き換えることにより、計算上は \mathbf{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り、隠れ層における出力ベクトル \mathbf{v} 及び $\hat{\mathbf{r}}$ を計算することができる。 \mathbf{v} を求めるためには、式(1)のような \mathbf{r}_i とエンコーダの重み行列 \mathbf{W} との線形和を計算する必要がある。このとき、 \mathbf{r}_i の欠損値要素を0と置き換えていたことにより、 \mathbf{v} の値はユーザが付与した評価値のみが出力に影響することになる。

$$\mathbf{v} = f(\mathbf{W}\mathbf{r} + b_1) \quad (1)$$

ここで、 f はシグモイド関数のような非線形の活性化関数であり、 b_1 はバイアスユニットである。本研究ではバイアスユニットの値は設定していないため、以降の説明では省略する。最後に、 $\hat{\mathbf{r}}$ を求めるために、式(2)のような \mathbf{v} とデコーダの重み行列 \mathbf{W}' との線形和を計算する。

$$\hat{\mathbf{r}} = f(\mathbf{W}'\mathbf{v}) \quad (2)$$

ここで、エンコーダとデコーダの重み行列は tied weight とし、 $\mathbf{W}' = \mathbf{W}^T$ となる。

バックプロパゲーションでは、 \mathbf{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2乗誤差のような損失関数の取り扱いについて述べる。

\mathbf{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、推定値である r_{ij} との差分が生じる。差分が生じるということは、対応する \mathbf{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} が欠損値を示す0が入力されていた場合には、推定値である r_{ij} にも0を代入する。これにより、入出力間の差分が無くなり、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、損失関数の出力値がより小さくなるように、重み行列 \mathbf{W} を更新する。 \mathbf{W} は、損失関数からの出力が大きい程、大きく更新される。こ

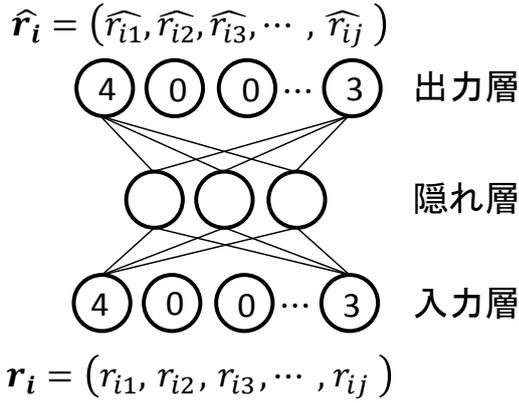


図2 部分次元法

のとき、部分次元法を用いた場合には、損失関数の出力値が小さく算出される。一般的によく用いられる損失関数である式 (3) のような 2 乗誤差で説明する。

$$L = \frac{1}{n} \sum (r_i - \hat{r}_i)^2 \quad (3)$$

ここで、 n は r_i の要素数であり、これはアイテムの数と等しい。部分次元法は、 r_i のある要素 r_{ij} が欠損値であった場合には 0 を入力し、さらに対応する出力 \hat{r}_{ij} の値も 0 とす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 (3) の n の値が r_i の要素数であると、入出力の差分が必ず 0 となる欠損値要素も n に含まれてしまい、結果として損失関数からの出力が小さく算出されてしまう。損失関数からの出力が小さいと、学習速度が低下し良い精度が得られない可能性がある。

これに対し、2 乗誤差和を r_i の要素数 n で除するのではなく、 r_i の評価値が存在する要素数 n_i で除することで、部分次元法による学習を行うことができる。

$$L = \frac{1}{n_i} \sum (r_i - \hat{r}_i)^2 \quad (4)$$

4 ユーザデータ除外法

すべての嗜好データを用いて学習を行った場合には、特定のユーザのみネットワークからの出力である推定値と正解値との差分が大きい傾向がある。本手法は、推定値と正解値との差分が大きなユーザを除外しつつ学習を行う方法である。

はじめに、学習用データから悪影響を及ぼすユーザデータを除外するための関数 `exclude` を定義する。

関数 `exclude(R, i)`

関数 `exclude` に行列 R と整数値 i を与えると、 R の i 行目成分をすべて R から削除し、インデックスを詰める。その後、インデックスを詰めた後の行列を返す。

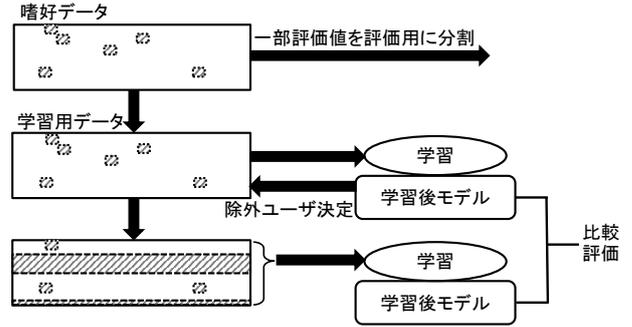


図3 データ処理図

図3を用いてデータの処理手順について述べる。

まず、嗜好データの評価値の一部をモデル評価用として分割し、残りを学習用データとする。次に、学習用データを部分次元法を適用したオートエンコーダへ入力し学習を行い、学習後モデルを作成する。さらに、学習後モデルに対し再び学習用データを入力し欠損部分を推定させ、その結果から除外するユーザを決定する。具体的には、全ユーザの推定値ベクトル \hat{r}_i と正解値から 2 乗誤差行列を算出し、その後 2 乗誤差行列の i 行目に相当するユーザ i の誤差ベクトル e_i から誤差平均 \bar{e}_i を求める。

$$\bar{e}_i = \frac{1}{n_i} \sum (e_i) \quad (5)$$

ただし、 n_i はユーザ i の観測値数である。同様の手順を全ユーザに対して行い、すべてのユーザの誤差平均を算出し、図4のような誤差頻度分布グラフを作成する。

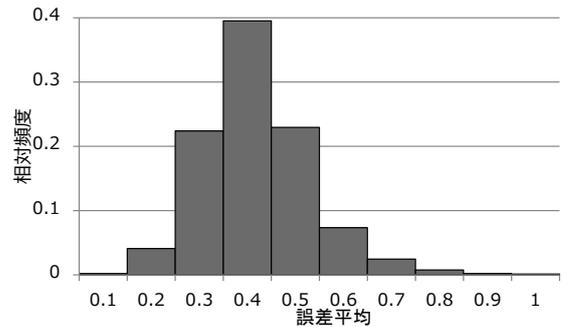


図4 誤差頻度分布図

次に図4の誤差頻度分布グラフからどのように学習に悪影響を及ぼしていると考えられるユーザを除外していくかについて述べる。まず、全てのユーザの \bar{e} から、平均値 μ と標準偏差 σ を求める。

$$\mu = \frac{1}{N} \sum \bar{e}_i \quad (6)$$

$$\sigma = \sqrt{\frac{1}{N} \sum (\mu - \bar{e}_i)^2} \quad (7)$$

ただし、 N は全ユーザ数である。次に、ユーザ i の \bar{e}_i の値が閾値 $\mu + \alpha\sigma$ 以上である場合に、ユーザ i が学習に悪影響を及ぼしていると考え、関数 `exclude` に学習用データと i を与え、学習用データからユーザ i のデータを除外する。このとき、 α は除外するユーザの量を決定するためのパラメータである。

ここまでの手順を終えると、学習に悪影響を及ぼしていると考えられるユーザデータが除外された学習用データが作成できる。この除外法適用後の学習用データを、部分次元法を適用したオートエンコーダへの入力として学習を行い、学習後モデルを作成する。ただし、ここで作成した学習後モデルは、除外法適用前の学習用データで作成したモデルとは異なるものである。最後に、2つの学習後モデルの推定精度を評価して比較する。

5 実験

5.1 目的

提案した除外法にて、除外したデータのみで学習させた場合の推定精度について検証する。

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 6040 人、アイテム数が 3952 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。

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 100, 200 の 2 パターンとした。バッチサイズは 5 とした。損失関数である入出力間の 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た。学習回数は 500 と設定した。

除外するユーザの割合を決定するためのパラメータである α の値は、2 と 3 の 2 パターンとした。このとき、 α の値によって除外されるユーザ数は、 α が 2 のときに 232 人、3 のときに 64 人である。

5.3 評価

評価値の推定精度は RMSE で評価する。

$$\text{RMSE} = \sqrt{\frac{1}{k} \sum (r_i - \hat{r}_i)^2} \quad (8)$$

ただし、 k は評価値行列に存在する全評価値数である。RMSE は値が小さい程良い結果である。

5.4 実験結果

図 5 は、学習用データに対して除外法を適用した際に生じる除外されたデータのみで作成したモデルの評価値

推定精度を示している。

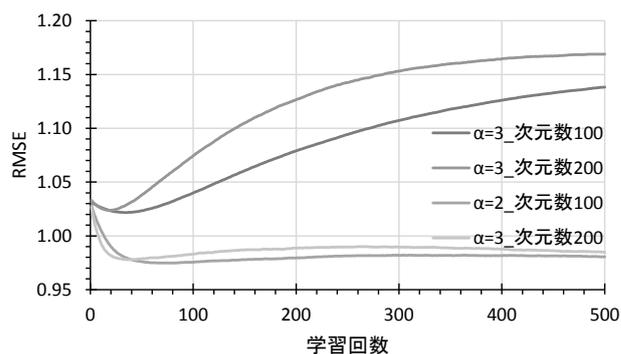


図 5 除外されたデータで作成したモデルの推定精度

5.5 考察

図 5 の実験結果から、どのパラメータにおいても過学習を引き起こしている可能性が高いということが見て取れる。特に、ユーザ数が少ない $\alpha = 3$ の場合には、推定精度の悪化が顕著であることから、データ数が少ないことが過学習の原因であると考えられる。また、推定精度が悪化していることから、これらユーザデータが学習に悪影響を及ぼす可能性がある。

6 おわりに

本稿では、学習に悪影響を及ぼすと考えられるユーザデータを除外する方法を提案した。除外法適用後の学習用データから作成したモデルの評価は今後の課題となる。

参考文献

- [1] 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.
- [2] Florian Strub, Jérémie Mary, "Collaborative Filtering with Stacked Denoising AutoEncoders and Sparse Inputs" NIPS Workshop on Machine Learning for eCommerce, Dec 2015, Montreal, Canada.
- [3] 田中恒平, 小林重樹, "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.
- [4] 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.
- [5] 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.
- [6] 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.