

Convex Factorization Machines を用いた学生モデリングの提案と有効性の検証 Verification of Student Modeling using Convex Factorization Machines

清水 大幹[†]
Daiki SHIMIZU

大枝 真一[‡]
Shinichi OEDA

1. はじめに

Educational Data Mining (EDM) と呼ばれる教育を専門としたデータマイニング分野がある。EDM は、教育システムによって収集された大量の電子データから有用な情報を発見することを目的としている。近年では、教育現場において e-Learning のような ITS (Intelligent Tutoring Systems) が普及しており、EDM の研究が盛んとなっている。

教育効果の高い ITS を提供するには、学習者のスキル状態を把握し、適切な設問を推薦する必要がある。ITS には学生モデル (Student Modeling) が取り入れられており、正確な学習者のスキル状態の推定や解答結果の予測が可能なモデルを利用することで、ITS の性能を向上させることができる。学生モデリングの目的は、試験結果などのログデータから学生のスキルを推定し、学生が設問に正答できるかどうかを予測することである。

先行研究では、学生モデリングの手法として、Knowledge Tracing や Factorization Machines[1] を用いた研究が行われてきた [2, 3]。Factorization Machines は Support Vector Machines の利点と factorization models を組み合わせたモデルであり、従来の手法より優れた結果を示している。しかし、Factorization Machines には、凸最適化ではないために局地的極小値を取得してしまう欠点がある。

本研究では、Factorization Machines を凸最適化した Convex Factorization Machines[4] を学生モデリングの手法として用いることで、予測精度の向上を図る。

2. 手法

2.1. Factorization Machines

FM (Factorization Machines) は Steffen Rendle によって 2010 年に発表された手法であり、SVM (Support Vector Machines) の長所と MF (Matrix Factorization) のような factorization models を組み合わせたモデルクラスである [1]。FM は、SVM と同様に教師あり学習を行う予測モデルであり、MF のように未知の値を推測することが可能である。

FM の次元 $d = 2$ (2-way FM) のモデル式は以下で表される。

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

このとき、 $\langle \cdot, \cdot \rangle$ は以下に示すように、 k 次元の 2 つの

ベクトルの内積を表す。

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{f=1}^k v_{i,f} \cdot v_{j,f}. \quad (2)$$

推定すべきモデルパラメータは以下となる。

$$w_0 \in \mathbb{R}, \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{V} \in \mathbb{R}^{n \times k}. \quad (3)$$

w_0 は全体のバイアス、 w_i は i 番目の変数の強度、 $\hat{w}_{i,j} := \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ は i 番目と j 番目の変数間の強度、 \mathbf{v}_i は \mathbf{V} の i 行目の k 次元ベクトルを表す。また $k \in \mathbb{N}_0^+$ は、分解の次元数を定義するハイパーパラメータである。

2-way FM では、二次の相互作用を $w_{i,j}$ という独立したパラメータではなく、 $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ というベクトルの内積に分解して表現している。これにより、SVM では対応できない非常にスパースなデータに対しても、高い精度で予測を行うことが可能となっている。また、学習時間が線形となるため、SGD (Stochastic Gradient Descent) を使用した直接最適化が実行可能になり、さまざまな損失関数に対して最適化することが可能になる。

2.2. Convex Factorization Machines

CFM (Convex Factorization Machines) では、FM の式 (1) を以下のように書き換える。

$$\hat{y}(\mathbf{x}) := \mathbf{w}^\top \mathbf{x} + \sum_{i=1}^d \sum_{j=1}^d z_{ij} x_i x_j = \mathbf{w}^\top \mathbf{x} + \langle \mathbf{Z}, \mathbf{x} \mathbf{x}^\top \rangle. \quad (4)$$

ここで、 $\mathbf{x} = \{1, \mathbf{x}\}^\top$ 、 $\mathbf{Z} = \mathbf{V} \mathbf{V}^\top \in \mathbb{S}^{d \times d}$ とした。また、 z_{ij} は \mathbf{Z} の ij 要素を表す。

以下の最適化問題を解くことで、パラメータの学習を行う。

$$\min_{\mathbf{w} \in \mathbb{R}^n, \mathbf{Z} \in \mathbb{S}^{n \times n}} \sum_{i=1}^n \ell(y_i, \hat{y}(\mathbf{x}_i)) + \frac{\alpha}{2} \|\mathbf{w}\|^2 + \beta \|\mathbf{Z}\|_*. \quad (5)$$

ここで、 ℓ は squared loss $\ell(y, \hat{y}) = \frac{1}{2}(y - \hat{y})^2$ や logistic loss $\ell(y, \hat{y}) = \log(1 + \exp(-y \cdot \hat{y}))$ などの 2 階微分可能な凸誤差関数、 $\alpha, \beta < 0$ はハイパーパラメータである。

FM は、非凸最適化問題を含むため、局所的極小値を取得してしまう。これは初期値に依存している。また、ランクハイパーパラメータ k の選択が必要であり、予測精度は k によって大きく変化してしまうという欠点がある。CFM は、 $\text{rank}(\mathbf{Z}) \ll n$ となるような低ランク行列 \mathbf{Z} を学習することを目的としており、これは nuclear norm を用いて \mathbf{Z} を正則化することにより達成することができる。さらに、行列ランクにおいて強凸下限であることが知られている。

つまり、CFM は FM を凸定式化したものであり、上記した欠点を克服したモデルとなっている。

[†]Advanced Course of Control and Information Engineering, National Institute of Technology, Kisarazu College

[‡]Department of Information and Computer Engineering, National Institute of Technology, Kisarazu College

表 1: Example of feature vector and target.

	Feature Vector \mathbf{x}										Target y			
$\mathbf{x}^{(1)}$	1	0	0	...	1	0	0	...	14	0	0	0	...	1
$\mathbf{x}^{(2)}$	1	0	0	...	0	1	0	...	15	1	0	0	...	1
$\mathbf{x}^{(3)}$	0	1	0	...	0	1	0	...	18	0	0	0	...	0
$\mathbf{x}^{(4)}$	0	0	1	...	0	1	0	...	10	0	0	0	...	1
$\mathbf{x}^{(5)}$	0	0	1	...	0	0	1	...	17	0	1	0	...	0
	Student				Item				Hour	Last Item				

表 2: Details of datasets.

	Records	Students	Items	Skills
Synthetic	10,000,000	5,000	10,000	1,000
Algebra	8,918,054	3,310	781,620	1,070

3. 先行研究

先行研究 [5] では、学習者、設問、設問を解いた時間、その設問を解く前に解いた設問を特徴ベクトルに採用し、その設問が解けたかどうかを教師データとした FM による予測を行っている。ここで、設問を解いた時間は、生物学的観点より、午後の方が午前中より疲れているために正答率は下がるという仮定の元で用いられている。表 1 に、先行研究で用いられた特徴ベクトルと教師データの例を示す。

4. 提案手法

本研究では、先行研究により提案されている FM を用いた学習者の解答結果予測に対し、精度向上を目指す。具体的には、FM を凸最適化した CFM を用いる。また、特徴量による精度の比較を行い、有用な特徴量を探索する。

5. 実験

5.1. 概要

本実験では、次節に示す 2 種類のデータセットを用いて FM と CFM の予測精度の比較実験を行う。このとき、各データセットで特徴量の組み合わせを変え、精度の比較を行う。

精度の評価尺度として、ACC (Accuracy), AUC (Area Under the Curve) と RMSE (Root Mean Squared Error) の 3 つを用いる。

5.2. データセット

本研究では、人工データと Algebra I 2008-2009[6] の 2 種類のデータに対して予測実験を行う。各データの概要を表 2 に示す。

5.2.1. 人工データ

人工データの生成には IRT (項目反応理論) を用いた。学習者が設問を 1 回解くと、その設問を解くために必要なスキルに対する学習者の能力値が上昇する。また、忘却を加えるために、学習者がある設問を最後に

解答した時間からの経過時間を算出し、経過時間が長いほどその設問を解くために必要なスキルに対する学習者の能力値が減少する。さらに、各設問を解くためには 1 つ以上の複数スキルが必要になるように設定した。

5.2.2. Algebra I 2008-2009

実際の e-Learning システムから取得されたデータとして、KDD Cup 2010 Educational Datamining Challenge で提供されたデータセットである Algebra I 2008-2009 を使用する。出題された設問には「四角形の面積を求めよ」などの数学の問題がある。

謝辞

本研究は、JSPS 科研費 19H01728 の助成を受けたものです。

参考文献

- [1] Steffen Rendle. “factorization machines”. In *Proceedings of the 2010 IEEE International Conference on Data Mining, ICDM '10*, pp. 995–1000, Washington, DC, USA, 2010. IEEE Computer Society.
- [2] Shinichi Oeda and Kouta Asai. “student modeling method integrating knowledge tracing and irt with decay effect”. In *EKM@EKAW*, 2016.
- [3] Jill-Jênn Vie and Hisashi Kashima. “knowledge tracing machines: Factorization machines for knowledge tracing”. *CoRR*, Vol. abs/1811.03388, , 2018.
- [4] Makoto Yamada, Amit Goyal, and Yi Chang. “convex factorization machine for regression”. *ArXiv*, Vol. abs/1507.01073, , 2015.
- [5] Nguyen Thai-Nghe, Lucas Drumond, Tomas Horvath, and Lars Schmidt-Thieme. “using factorization machines for student modeling”. Vol. 872, , 01 2012.
- [6] J. Stamper, A.Niculescu-Mizil, S.Ritter, G.J.Gordon, and K.R.Koedinger. Algebra I 2008-2009. <http://pslclatashop.web.cmu.edu/KDDCup/downloads.jsp>.