

Factorization Machines を用いた教育データの解析

山下 滉†

大枝 真一‡

木更津工業高等専門学校 制御・情報システム工学専攻†

木更津工業高等専門学校 情報工学科‡

1. はじめに

近年、教育現場において e-Learning システムが広く普及してきており、システムを利用した学習者のログが大量に蓄積されるようになってきている。また、大量の教育データに対してデータマイニングの手法を活用し、有益な知識を抽出しようとする試みが注目を集めている。国外では Educational Data Mining (EDM) と呼ばれるコミュニティが誕生し、教育データの活用方法について活発に議論がなされている [1]。

EDM の目標の 1 つに、学習者のモデル化がある。学習者のモデル化とは、過去の解答結果などのログデータから学習者のスキル状態を推定し、その推定結果から設問に正答するか誤答するかの予測を行うことの総称である。学習者のモデル化は学生モデリング (Student Modeling) と呼ばれており、Knowledge Tracing [2] という手法が有名である。

Intelligent Tutoring System (ITS) と呼ばれる e-Learning システムでは、学習者のスキルに見合った設問を提示することで、効率の良い学習環境を提供する。ITS には学生モデルが取り入れられており、より高性能な ITS を作るにはより正確に学習者のスキル状態の推定や解答結果の予測が可能な学生モデルが必要となる。

本研究では、Support Vector Machine (SVM) と Matrix Factorization (MF) を組み合わせた手法である Factorization Machines [3] とその拡張手法を用いて教育データ、特に e-Learning システムのログデータを解析し、学生モデリングを行う。

2. Factorization Machines

Factorization Machines (FM) は、SVM と MF のアドバンテージを組み合わせたモデルである [3]。

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 の様に未知の値を推測することが可能である。また、FM は SVM では扱うことの難しい非常にスパースな入力データに対しても高い精度で予測を行うことが可能なモデルとなっている。

式 (1) に、FM の次元 $d = 2$ のときのモデルの式を示す。

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

また、パラメータは以下である。

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

ここで、 $\langle \cdot, \cdot \rangle$ は k 次元の 2 つのベクトルの内積を表す。

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \sum_{f=1}^k v_{i,f} \cdot v_{j,f} \quad (3)$$

式 (1) の計算量は、 $\mathcal{O}(kn^2)$ であるが、式 (4) による式変換を行うことにより、 $\mathcal{O}(kn)$ に削減することができる。

$$\begin{aligned} & \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j - \frac{1}{2} \sum_{i=1}^n \langle \mathbf{v}_i, \mathbf{v}_i \rangle x_i x_i \\ &= \frac{1}{2} \left(\sum_{i=1}^n \sum_{j=1}^n \sum_{f=1}^k v_{i,f} v_{j,f} x_i x_j \right. \\ & \quad \left. - \sum_{i=1}^n \sum_{f=1}^k v_{i,f} v_{i,f} x_i x_i \right) \\ &= \frac{1}{2} \sum_{f=1}^k \left(\left(\sum_{i=1}^n v_{i,f} x_i \right) \left(\sum_{j=1}^n v_{j,f} x_j \right) \right. \\ & \quad \left. - \sum_{i=1}^n v_{i,f}^2 x_i^2 \right) \\ &= \frac{1}{2} \sum_{f=1}^k \left(\left(\sum_{i=1}^n v_{i,f} x_i \right)^2 - \sum_{i=1}^n v_{i,f}^2 x_i^2 \right) \quad (4) \end{aligned}$$

Analysis of Educational Data using Factorization Machines

†Hikaru YAMASHITA · Advanced Course of Control and Information Engineering, National Institute of Technology, Kisarazu College

‡Shinichi OEDA · Department of Information and Computer Engineering, National Institute of Technology, Kisarazu College

表 1: Example of feature vector and target.

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

FM のパラメータは、確率的勾配降下法をはじめとした勾配法により求められる。

SVM と FM の大きな違いは、パラメータの成り立ちにある。FM は、 $\langle v_i, v_j \rangle$ と $\langle v_i, v_l \rangle$ のように 2 つのベクトルの内積で相互作用を表すパラメータを表現する。これにより、 $\langle v_i, v_j \rangle$ と $\langle v_i, v_l \rangle$ の 2 つのパラメータは同じベクトル v_i を共有しているため、互いに依存関係がある。そのため、SVM では正しく扱うことが難しい非常な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になる。

3. 先行研究

先行研究 [4] では、特徴ベクトルに学習者、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採用し、その設問が解けたかどうかを教師データとしている。ここで、設問を解いた時間は、生物学的観点から言うと午前中より疲れているであろう午後の方が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 [4] での特徴ベクトルと教師データの例を示す。

先行研究 [4] では、実データに対して Knowledge Tracing や MF など従来の手法と FM を用いた手法による学習者の正答予測の比較を行っている。比較の結果、FM を用いた手法が他のどの手法よりも予測精度が高いと記している。

また、先行研究 [5] では、特徴ベクトルに学習者、設問、設問の階層構造、設問の系列、必要とされているスキルなどを用いている。

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の予測をより精度の高

いものに改良していく。具体的には、特徴ベクトルの要素をより最適なものに変更することや、FM の派生モデルを用いることを考えている。

5. まとめ

本研究では、Factorization Machines を用いて学習者の解答結果の精度の高い予測と、学習者のスキル状態の推定を行うことを提案した。

今後は、人工データ及び実データで実験を行い、提案手法の有効性を検証していく。

謝辞 本研究は JSPS 科研費 16K01095 の助成を受けたものである。

参考文献

- [1] Toon Calders and Mykola Pechenizkiy. “introduction to the special section on educational data mining”. *SIGKDD Explor. Newsl.*, Vol. 13, No. 2, pp. 3–6, May 2012.
- [2] Albert T. Corbett and John R. Anderson. “knowledge tracing: Modelling the acquisition of procedural knowledge”. *User Model. User-Adapt. Interact.*, Vol. 4, No. 4, pp. 253–278, 1995.
- [3] 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.
- [4] Nguyen Thai-Nghe, Lucas Drumond, Tomas Horvath, and Lars Schmidt-Thieme. “using factorization machines for student modeling”. Vol. 872, , 01 2012.
- [5] Steffen Rendle. “factorization machines”. The Machine Learning Conference 2014, November 2014.