

学生モデリングにおける説明可能性向上のための Deep Factorization Machines DeepFM for Improving Explainability in Student Modeling

柿崎 透真[†]
Toma Kakizaki

大枝 真一[‡]
Shinichi Oeda

1. はじめに

教育現場において ITS(Intelligent Tutoring System) を効果的に活用するためには、学習者のスキル状態を把握し、それに見合った設問を推薦する必要がある。そのため学習者のスキル状態の推定が可能な学生モデリングを用いることで、ITS の性能を向上させる研究が行われている。学生モデリング手法は Knowledge Tracing が主流となっており、Deep Knowledge Tracing や Self-Attentive Knowledge Tracing などのディープラーニングアプローチが盛んに研究されている。しかし、これらのモデルは学生が解いた問題の番号と、その問題に対する回答の正誤のみを入力としており、その他の特徴量を用いることを想定していない。

本研究では、特徴量の相互作用を考慮できる手法の FM(Factorization Machines) と DNN(Deep Neural Network) を組み合わせたモデルである DeepFM と FiBiNET を用いて学生モデリングを行い、精度の比較を行う。また、特徴量重要度を SHAP などの手法で可視化することで、実世界への機械学習手法の応用を目指す。

2. 先行研究

2.1. Knowledge Tracing

学生モデリングのデファクトスタンダードな手法として知られているのが、Knowledge Tracing である。近年は DNN(Deep Neural Network) を組み込んだ手法が盛んに研究されており、代表的なものとして、DKT[1]、DKVMN[2]、SAKT[3] などが挙げられる。上記の DNN ベースのアプローチでは、問題番号と対する正答結果が与えられていれば学習が可能であり、学生が保有している潜在スキル関係を把握できる。

2.2. Factorization Machines

FM(Factorization Machines)[4] は、教師あり学習と行列分解モデルを組み合わせたモデルであり、推薦システムでの CTR 予測を中心に盛んに研究が行われている。FM(次元 $d=2$) の計算式を式 (1) に示す。

$$y_{FM} = \langle w, x \rangle + \sum_{i=1}^n \sum_{j=i+1}^n \langle v_i, v_j \rangle x_i \cdot x_j \quad (1)$$

$\langle \cdot, \cdot \rangle$ は 2つのベクトルの内積を示しており、 $w_{1:n} \in \mathbb{R}^n$, $v_{1:k} \in \mathbb{R}^k$ は学習パラメータである (k はハイパーパラメータ)。式 (1) の第 1 項で特徴量の 1 次での重要度を、第 2 項で特徴量の 2 次の相互作用を計算している。2次

の相互作用を w_{ij} のような独立したパラメータではなく、ベクトルの内積に分解して表現することで、推薦システムや教育データでよく見られるスパースなデータに対しても高い精度で予測できる。

2.3. Deep Factorization Machines

DeepFM(Deep Factorization Machines)[5] は DNN(Deep Neural Network) と FM(Factorization Machines) を組み合わせたモデルである。DeepFM の計算式を式 (2) に示す。

$$\hat{y} = \text{sigmoid}(y_{FM} + y_{DNN}) \quad (2)$$

入力ベクトルに対して、式 (1) で示した次元 $d=2$ の FM を用いる FM コンポーネントと、FFN(Feed Forward Neural Network) を用いる Deep コンポーネントに分けて計算を行い、sigmoid 関数で確率を算出する前に両者を足し合わせている。DeepFM では、FM による特徴量の低次の組み合わせに加えて、DNN の導入による特徴量の高次の組み合わせも考慮できる。また、FM コンポーネントと Deep コンポーネントで入力を共有しているため、特別な特徴量エンジニアリングが必要なく、効率的な計算が可能である。

2.4. FiBiNET

FiBiNET[6] は、画像処理分野で提案された SENet を用いたモデルであり、CTR 予測を主なタスクとしている。SENet Layer は、特徴量を 1 次元に圧縮する Squeeze ステップ、圧縮された特徴量をもとに重みを学習する Excitation ステップ、重みをもとに重要な特徴量を抽出する Re-Weight ステップからなる。この機構により、重要な特徴量に対する重みは大きく、それ以外の特徴量に対する重みは小さくなるように学習される。埋め込みベクトル $E = [e_1, e_2, \dots, e_n]$ を、Squeeze, Excitation, Re-Weight に対応する関数に通すことで、新たな埋め込みベクトル $V = [v_1, v_2, \dots, v_n]$ が出力される。

また、FiBiNET では、特徴量の相互作用を内積やアダマール積で計算するのではなく、Bilinear-Interaction という形で計算をしており、内積とアダマール積の間に学習可能な重み W を入れることで、より複雑な相互作用を捉えることを試みている。

3. 実験概要

本研究では、先行研究により提案されている DNN ベースの KT 手法 (DKT, DKVMN, SAKT) と、推薦システムでの CTR 予測手法として知られている DeepFM, FiBiNET を用いて学生モデリングを行い、両者の比較を行う。

3.1. 実データセットによる実験

本実験は、3つの実データセット (ASSISTment2009, ASSISTment2015, EdNet) を用いて行った。それぞれ

[†]木更津工業高等専門学校 制御・情報システム工学専攻
Advanced Course of Control and Information Engineering, National Institute of Technology, Kisarazu College

[‡]木更津工業高等専門学校 情報工学科
Department of Information and Computer Engineering, National Institute of Technology, Kisarazu College

表 1: 実データセットの概要/学生モデリングの実験結果

Name	Information					AUC				
	User	Skill_tag	Record	Unique	Density	DKT	DKVMN	SAKT	DeepFM	FiBiNET
ASSIST2009	4,217	124	401,462	44,020	0.0842	0.811	0.7969	0.8007	0.8334	0.8274
ASSIST2015	19,840	100	683,801	124,449	0.0627	0.7297	0.7209	0.7247	0.6833	0.6658
EdNet	2,000	×(10,770 問)	157,881	96,643	0.00449	0.5972	0.5519	0.6420	0.7546	0.7502

の実データセットの概要を表 1 に示す。表 1 で、User はユーザ数、Skill_tag は問題に属するスキルの種類の総数、Record はデータセットの総レコード数を表している。また、Unique はユーザ名とスキル情報が重複しているレコードを除いたレコード数であり、Density はデータセットの密度である。ここで、Density は Unique / (User × Skill_tag) と計算した。この値が小さいほど、スパースなデータであるといえる。また、EdNet には Skill_tag が与えられていないため、問題の総数 10,770 を用いて Unique なレコード数を計算し、Density は Unique / (User × 10,770) と計算した。

ASSISTment は小学校の算数テストの結果を収集したデータセットである。ASSISTment2009 は列数が 30 列あり、様々な特徴量の使用が考えられるデータセットであるのに対し、ASSISTment2015 は列数が 4 列のみであるため、用いることができる特徴量が限られている。EdNet は TOEIC 対策プラットフォームのデータを収集したものであり、リスニング音声を再生した回数や対策動画を視聴した回数など、学習行動に関する情報が多く保管されている。

3.2. 評価指標

本実験では、評価指標として AUC (Area Under the Curve) を用いる。AUC は 2 群の分類問題における精度評価指標として使われており、False Positive Rate (モデルが正と判断して不正解だった割合) を横軸に、True Positive Rate (モデルが正と判断して正解だった割合) を縦軸にとった曲線の下側面積である。0 から 1 までの値をとり、値が大きいほど精度が良いとされる。

4. 実験結果・考察

3 つの実データセットを用いた学生モデリングの結果を表 1 に示す。全てのモデルにおいて、データセットの 80% を train データとして学習に使い、残りの 20% を test データとして評価に使用した。また、最も精度の良かったモデルの AUC の数値を強調して表示している。

データセットとして ASSISTment2009 と EdNet を用いた場合は、DeepFM と EdNet が KT 手法の 3 つの精度を上回る結果となった。一方で、データセットに ASSISTment2015 を用いると、DeepFM と FiBiNET の精度は KT 手法を下回った。以上の結果から、ユーザ番号や問題番号、過去に消化した宿題の番号などの基本的な情報をはじめ、リスニングの音声を再生した回数、解説ビデオを視聴した回数などの学習行動の情

報は学生モデリングに有用であると考えられる。一方で、豊富なデータを生成できない環境 (IT 技術に精通していない小中学校など) では、問題番号と正答結果の組のみで学習できる KT 手法が優れていると考えられる。

5. まとめ・今後の課題

本研究では、推薦システムでの CTR 予測モデルとして提案された DeepFM と FiBiNET を用いて学生モデリングを行い、学生モデリングのデファクトスタンダードである Knowledge Tracing (DKT, DKVMN, SAKT) と比較を行った。

今後は、DeepFM や FiBiNET の説明可能性を SHAP [7] などの手法で可視化することによる、実世界へのモデルの応用を目指す。

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] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas Guibas, and Jascha Sohl-Dickstein, “Deep knowledge tracing”, In NIPS’15 Proceedings of the 28th International Conference on Neural Information Processing Systems, Volume 1, Vol. 28, pp. 505-513 (2015).
- [2] Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung, “Dynamic key-value memory networks for knowledge tracing”, In WWW’17 Proceedings of the 26th International Conference on World Wide Web, WWW 2017, pp. 765-774 (2017).
- [3] Shalini Pandey and George Karypis, “A self-attentive model for knowledge tracing”, In 12th International Conference on Educational Data Mining, EDM 2019, pp.384-389 (2019).
- [4] Steffen Rendle, “Factorization Machines”, In the 10th IEEE International Conference on Data Mining, pp. 995-1000 (2010).
- [5] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, Xiuqiang He, “DeepFM: A Factorization-Machine based Neural Network for CTR Prediction” In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, pp.1725-1731 (2017).
- [6] Tongwen Juang, Zhiqi Zhang, Junlin Zhang, “FiBiNET: Combining Feature Importance and Bilinear feature Interaction for Click-Through Rate Prediction”, In 13th ACM Conference on Recommender Systems, (2019).
- [7] Scott M. Lundberg, Su-In Lee, “A Unified Approach to Interpreting Model Predictions”, In NIPS’17 Proceedings of the 31th International Conference on Neural Information Processing Systems, pp. 4768-4777 (2017).