

## 薬学的知見を必要とした回帰タスクに対する合成臨床データの影響に関する検討

### A Study on the Influence of Synthetic Clinical Data on Regression Tasks Requiring Pharmacological Knowledge

○宮野咲紀<sup>1</sup>, 関弘翔<sup>2</sup>, 辻泰弘<sup>3</sup>, 細野裕行<sup>2</sup>  
 \*Saki Miyano<sup>1</sup>, Hiroto Seki<sup>2</sup>, Yasuhiro Tsuji<sup>3</sup>, Hiroyuki Hosono<sup>2</sup>

**Abstract:** In this paper, we investigate the influence of synthetic clinical data, which do not take pharmacological knowledge into account, on the learning of a regression task that requires pharmacological knowledge. The use of synthetic clinical data improved the accuracy of the regression, but also suggested the need for synthetic clinical data to take pharmacological findings into account.

#### 1. まえがき

近年、リアルワールドデータ (RWD) の利活用が注目されており、医薬品開発の効率化や臨床研究の効率化について検討が行われている。RWD は各病院などの臨床現場で収集される臨床データである。RWD の利活用においては、複数組織間のデータ共有が困難であることが効率化を妨げる要因の1つとなっている。そこで我々は AI を用いて RWD の合成データ (合成臨床データ) を生成することで解決できると考えた。合成臨床データを医薬品開発の効率化に用いるためには、薬物動態などの薬学的知見を考慮する必要がある。

本研究の目的は、薬学的知見を考慮した合成臨床データ生成を行うことである。本稿では、薬学的知見を考慮していない合成臨床データが、薬学的知見を必要とした回帰タスクの学習に与える影響の検証を行う。

#### 2. ANN-PK による薬物血中濃度の予測

薬学的知見を必要とした回帰タスクに薬物血中濃度 (Drug concentration: CONC) の推定を採用し、学習するモデルには我々のグループで提案している ANN-PK<sup>[1]</sup>を採用した。ANN-PK は母集団薬物動態 (Population Pharmacokinetics: PopPK) とニューラルネットワーク (ANN) を組み合わせ、ANN により薬物動態パラメータの一つのクリアランス (CL) を予測し、経時的な CONC を PK モデルで推定する手法である。

#### 3. 使用データ

検証には、免疫抑制剤シクロスポリン (36 名分, 合計 89 ポイント) の臨床データ<sup>[2]</sup>を用いた。患者属性として、投与量 (Dose) [mg], 投与後経過時間 (Time1~Time3) [h], 性別 (SEX), 年齢 (AGE) [year], 体重 (TBW) [kg], 身長 (HT) [m], 血液検査値である肝機能値 (ALT, AST) [U/L], 腎機能値 (SCR) [mg/dL], および薬物血中濃度 (CONC) [mg/L]がある。本稿では、Train データと Test データの比率を 9 対 1 として検証を行った。

#### 4. 検証方法

合成臨床データを生成する手法として、Generative Adversarial Networks (GAN) に我々が考案した改良を加えたモデル<sup>[3]</sup>を採用している。この GAN は薬学的知見を考慮せずに合成臨床データを生成する。

ANN-PK の学習に用いるデータを複数用意した。用意したデータのパターンは以下の3つである。

- A) Train データのみ (合計 80 ポイント)
- B) Train データと合成データの混合データ (合計 160 ポイント)
- C) 合成データのみ (合計 80 ポイント)

パターン B の混合データは Train データと Test データを 1 対 1 で混ぜたデータである。

検証方法は、まず Train データを用いて GAN を学習し、学習済み GAN を用いて合成データの生成を行う。その後、各パターンのデータを ANN-PK に入力し学習を行う。ANN-PK 学習後、Test データをそれぞれの ANN-PK のモデルに入力し、推定した CONC の精度にどのような差と影響が生まれるのかを検証する。

検証結果の評価方法として、ANN-PK の評価指標である二乗平均平方根誤差 (RMSE), XAI である SHAP<sup>[4]</sup>を用いる。SHAP は機械学習モデルの出力に寄与した入力属性を可視化する手法である。SHAP を用いることで、各パターンのデータが ANN-PK の学習にどのような影響を及ぼしているか解釈が可能となる。

#### 5. 結果および考察

まず、パターン B およびパターン C で用いるデータの生成を行った。Fig.1, 2 に Train データおよび合成データの箱ひげ図を示す。これらより、年齢や体重、身長などの患者属性については、Train データと似た統計量を持つデータであることが確認できた。また、ALT や SCR 等の血液検査値の統計量が実データと異なっていることがわかる。CONC について見ると、高い血中濃度を持つデータが生成できていないことが判明した。

1: 日大理工・院 (後)・情報, 2: 日大理工・教員・情報, 3: 日大薬・教員・薬学

Table 1 に各パターンのデータを用いて学習した ANN-PK の RMSE を示す。

これより、RMSE が最も低いのはパターン B の時であった。このことから、少ない Train データで ANN-PK を学習する際に、合成臨床データで Train データの増強を行うことによって ANN-PK のさらなる精度向上が可能であるということが分かった。

また、Fig.3, 4, 5 に各パターンで学習した ANN-PK の SHAP による可視化結果を示す。

SHAP の可視化結果は縦軸の上部が最も ANN-PK の CL 予測に優位に影響する属性であり、下部になるにつれて結果に対する重要度が下がっていく。Fig.3 の結果を見ると、AGE と TBW が ANN-PK の結果に優位であり、AGE が高いほど CL は低く、TBW が高いほど CL は高い傾向がある。これは薬学的に妥当な結果である<sup>[3]</sup>。Fig.4 の結果では、SEX が浮上しているが、AGE と TBW の優位性は変動がなく、薬学的な妥当性を維持していると言える。Fig.5 の結果では、AGE と TBW が CL に与える影響の傾向は概ね妥当であるが、AGE の重要度が大きく低下したことが分かる。このことから、薬学的知見を考慮していないデータのみで学習を行う場合、ANN-PK は薬学的な妥当性を十分に満たさないことが分かった。以上より、合成臨床データを ANN-PK の学習に用いる際には、薬学的知見を考慮したデータを用いる必要があると言える。

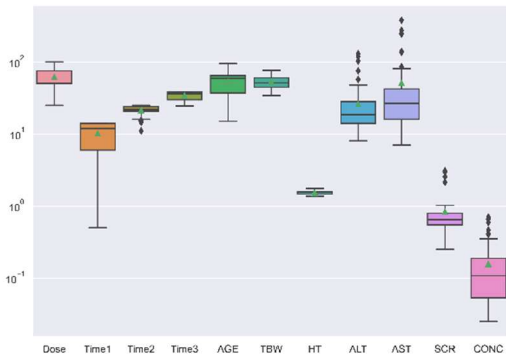


Figure 1. Box plot of Train data

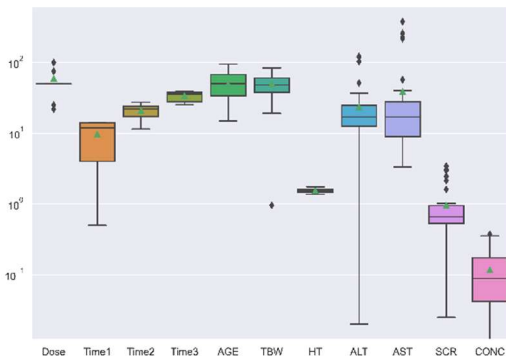


Figure 2. Box plot of synthetic data

Table 1. RMSE of ANN-PK trained for each pattern

Pattern	RMSE[ng/mL]
A	43.909
B	33.625
C	70.122

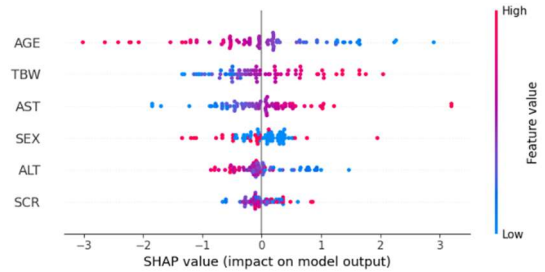


Figure 3. ANN-PK visualization results for Pattern A

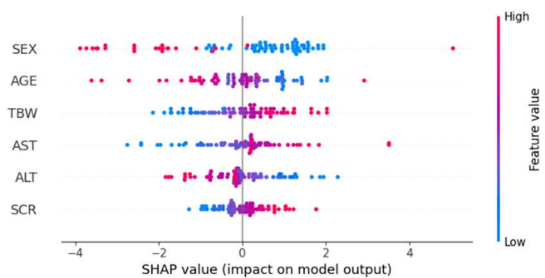


Figure 4. ANN-PK visualization results for Pattern B

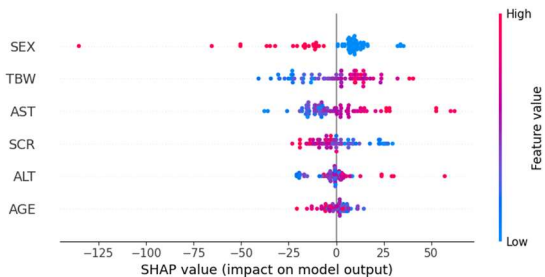


Figure 5. ANN-PK visualization results for Pattern C

## 6. まとめ

本稿では、薬学的知見を考慮していない合成臨床データが、薬学的知見を必要とした回帰タスクの学習に与える影響の検証を行った。合成臨床データを用いて Train データの増強を行うことで学習精度向上が見込めると判明した。今後は薬学的知見を考慮した合成臨床データ生成に取り組む。

## 参考文献

- [1] C, Ogami, et al., C PT: Pharmacometrics & Systems Pharmacology, 10(7), pp.760-768, 2021.
- [2] Y. Tsuji, et al., Biological and Pharmaceutical Bulletin, 38(9), pp.1265-1271, 2015.
- [3] 宮野咲紀, 関弘翔, 辻泰弘, 細野裕行, 令和5年度電気学会全国大会, 3-090, 2023.
- [4] Scott M. Lundberg et al., Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017.