192　創薬至適化合物を選出する深層ランク学習	大上　雅史

【目的】創薬標的への強い活性を持つ新薬候補化合物の創成や、薬剤候補化合物の薬物動態パラメータの正確な推定は、医薬品研究開発の加速やコスト削減に直接繋がる重要課題である。本研究ではこれまであまり予測手法が確立されていなかった体内動態パラメータとして、薬剤の血液胎盤関門の透過性に着目し、これを予測するための学習手法を開発した。当初は深層ランク学習技術を活用することを検討していたが、ウェブアプリ等の形式でユーザーが迅速にかつ簡便に予測値を得られるようにするというアプリケーション実装を考慮した結果、勾配ブースティング木による予測手法構築を採用した。

【方法】血液胎盤関門透過性を表す指標である logFM および CI を対象に、それぞれのデータセットを用いた予測モデルの構築を行った。logFM データセットは *in vivo* 測定として臍帯血と母体血の血中濃度比（F/M 比）から得られたデータであり、Takaku らが文献から収集した 55 化合物の logFM 値を用いた。CI データセットは、*ex vivo* 実験であるヒト胎盤灌流実験によって化合物が母体側から胎児側へ移動する速度から求められる血液胎盤関門透過性の評価方法であり、コントロール化合物であるアンチピリン（antipyrine）に対するクリアランスの比率 CI（clearance index）のデータである。CI データは Giaginis らが文献から収集した 88 化合物の CI 値を用いた。予測モデルの構築では、化合物の特徴ベクトルとして mordred 記述子を用いた。各データセットは、それぞれ先行研究と同一の分割方法にて訓練データとテストデータに分割し、予測モデルの学習および検証に用いた。予測モデルの構築には勾配ブースティング木の実装の一つである LightGBM を用い、optuna の LightGBM Tuner によってパラメータチューニングを行った。評価指標は決定係数 R^2 と根平均二乗誤差 RMSE を用いた。また、構築した LightGBM による予測モデルを用いて、テキストボックスに入力された化合物の SMILES 文字列から CI と LogFM の予測値を計算するウェブサービスを構築した。フロントエンドは Vue.js フレームワークを用い、バックエンドは Python と Flask を用いてシングルページアプリケーションとして構築した。あわせて https://pbpredictor.net/api/predict?smiles=[str] の URI によって SMILES 文字列から logFM と CI の予測値をリクエスト可能な REST API を実装した。

【結果】提案手法の LightGBM による予測は、logFM と CI の双方で従来手法の予測精度を上回る結果となった。特に、従来手法ではテストデータの精度が訓練データに比べて良くなかったが、提案手法ではテストデータでも良好な予測性能を示しており、汎化性能が良いことが示唆された。構築したウェブサービスは https://pbpredictor.net からアクセス・利用が可能である。

予測ウェブサービス PBPredictor の概観 https://pbpredictor.net



192. 創薬至適化合物を選出する深層ランク学習

大上　雅史

東京工業大学 情報理工学院 情報工学系

Key words：血液胎盤関門透過性，教師あり学習，勾配ブースティング木，胎児母体血中濃度比

緒　言

創薬標的への強い活性を持つ新薬候補化合物の創成や、薬剤候補化合物の薬物動態パラメータの正確な推定は、医薬品研究開発の加速やコスト削減に直接繋がる重要課題である。医薬品開発プロセスにおいて生化学実験は欠かすことができないが、計算機による予測に基づいた化合物スクリーニングは、実験の時間的・金銭的コストを大幅に削減できる可能性がある。計算機によるスクリーニングには、昨今ではAI技術（機械学習技術）が積極的に用いられるようになった。具体的には、未知化合物の活性・体内動態等のパラメータを、既知化合物や上市薬から予測モデルを学習し適用することによって推定することである。本研究ではこれまであまり予測手法が確立されていなかった体内動態パラメータとして、薬剤の血液胎盤関門の透過性に着目し、研究を実施した。本研究課題は、当初は深層ランク学習技術を活用することを検討していたが、ウェブアプリ等の形式でユーザーが迅速にかつ簡便に予測値を得られるようにするというアプリケーション実装を考慮した結果、結果的に勾配ブースティング木を採用した。

母体に投与された薬剤が胎児にどの程度移行するかは、胎盤にある血液胎盤関門をどの程度透過するかという指標で評価される。薬剤候補となる化合物の血液胎盤関門透過性を知ることは安全な薬剤を開発する上できわめて重要であり、また妊婦の治療における指針にもなりうる［1］。しかし、ヒトの胎盤は他の動物モデルでは再現されない種特異的性質を持っており、ヒトの組織モデルを用いた評価が必要となっている［2］。

化合物の血液胎盤関門透過性の評価には、出産時の臍帯血（胎児血）にどの程度化合物が含まれるかを調べた胎児－母体血中濃度比（F/M 比）や、ヒト摘出胎盤を用いた胎盤灌流モデルによる *ex vivo* 実験が用いられる。F/M 比は *in vivo* データとして血液胎盤関門透過性の直接的な指標となるが、原理的に安全であるとわかっている薬剤のデータしか得られないという制限がある。一方の胎盤灌流モデルによる *ex vivo* 実験であればどのような化合物でも測定できるものの、あくまで間接的な指標であることに注意が必要である。

本研究では、現在までに得られている F/M 比のデータや胎盤灌流モデルによる実験データから、血液胎盤関門透過性を予測することを試みた。これまでに機械学習による血液胎盤関門透過性予測についてはいくつかの事例が存在する［3～6］。しかし、いずれも予測精度は不十分であり、また予測手法を利用できるソフトウェア等も存在していなかった。そのため、本研究では機械学習による予測を広く利用できるように比較的軽量な計算で高精度な予測が可能な勾配ブースティング木（Gradient Boosting Decision Tree：GBDT）に着目した。GBDT を用いて、2 種類の公開データ（出産時の臍帯血および母体血中の薬物濃度測定で得られる *in vivo* データ（logFM）、およびヒト胎盤灌流実験で得られる *ex vivo* データ（CI））からそれぞれの回帰モデルを構築し、より高精度な予測を行い、さらに構築した回帰モデルによる予測を実際に提供するウェブサービスを構築することを目的とした。本研究の一部は情報処理学会バイオ情報学研究会で報告を行った［7］。

方　法

1. データセット

本研究では、血液胎盤関門透過性を表す以下の 2 つの指標を対象に、それぞれのデータセットを用いた予測モデルの

構築を行った。

1）logFM データセット：ヒトにおける化合物の胎盤移行を直接測定することは困難であるが、*in vivo* 測定として臍帯血と母体血の血中濃度比（F/M 比）のデータが存在する。logFM は F/M 比の常用対数をとった値である。本研究では Takaku らが文献から収集した 55 化合物の logFM 値を用いた［3］。このデータセットでは、Takaku ら ［3］が多重線形回帰により、また Wang ら［4］が Lasso 回帰による予測手法を提案している。本研究では、Wang らの予測を比較対象として取り上げた。

2）CI データセット：*ex vivo* 実験であるヒト胎盤灌流実験によって、化合物が母体側から胎児側へ移動する速度（クリアランス）を求めることで血液胎盤関門透過性を評価できる。コントロール化合物であるアンチピリン（antipyrine）に対するクリアランスの比率を CI（clearance index）といい、CI によって化合物の胎盤移行が評価される。胎盤灌流実験は介入を必要としないため、logFM に比べてデータが取りやすいという利点がある。本研究では Giaginis らが文献から収集した 88 化合物の CI 値を用いた［5］。このデータセットでは、Giaginis ら［5］および Zhang ら［6］が Partial Least Square を用いた予測手法を提案している。本研究では、Zhang らの予測を比較対象として取り上げた。

2. 予測モデルの構築

化合物の特徴ベクトルとして、mordred 記述子を用いた。mordred は RDKit ライブラリを介して 1,613 種類の 2D 記述子と 213 種類の 3D 記述子が計算できる。3D 記述子の計算に必要な立体配座生成には、RDKit の ETKDGv2 法を用いた。

各データセットは、それぞれ先行研究と同一の分割方法［4, 6］にて訓練データとテストデータに分割した。回帰モデルの構築には勾配ブースティング木の実装の一つである LightGBM を用いた。また、パラメータチューニングは訓練データの Leave-One-Out 交差検定（LOOCV）の結果に基づき、optuna の LightGBM Tuner を用いて行った。ただし事前に確認したところ、特に logFM の予測で過学習が起きており、訓練データの LOOCV は決定係数 $R^2=1$ となりテストデータの R^2は極端に悪化するという結果であった。そのため、Lasso 回帰による特徴選択を訓練データの LOOCV で事前に行い、計 5 個の記述子を選択して LightGBM の入力とした。

3. 評価指標

決定係数 R^2 と根平均二乗誤差 RMSE を用いた。optuna の最適化指標には LOOCV での RMSE 値を用いた。

4. ウェブサーバー実装

構築した LightGBM による予測モデルを用いて、テキストボックスに入力された化合物の SMILES 文字列から CI と LogFM の予測値を計算するウェブサービスを構築した。フロントエンドは Vue.js フレームワークを用い、バックエンドは Python と Flask を用いてシングルページアプリケーションとして構築した（図 1）。あわせて https://pbpredictor.net/api/predict?smiles=[str]の URI によって SMILES 文字列から logFM と CI の予測値をリクエスト可能な REST API を実装した。

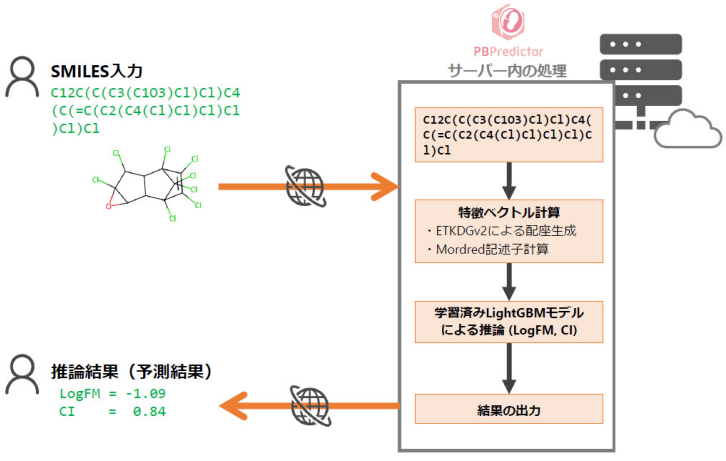


図 1. 予測ウェブサービス構築の概要図

結果および考察

1. 血液胎盤関門透過性の予測結果

logFM 値の予測結果を図 2 に、CI 値の予測結果を図 3 に示す。提案手法の LightGBM による予測は、logFM と CI の双方で従来手法の予測精度を上回る結果となった。特に、従来手法ではテストデータの精度が訓練データに比べて良くなかったが、提案手法ではテストデータでも良好な予測性能を示しており、汎化性能が良いことが示唆される。

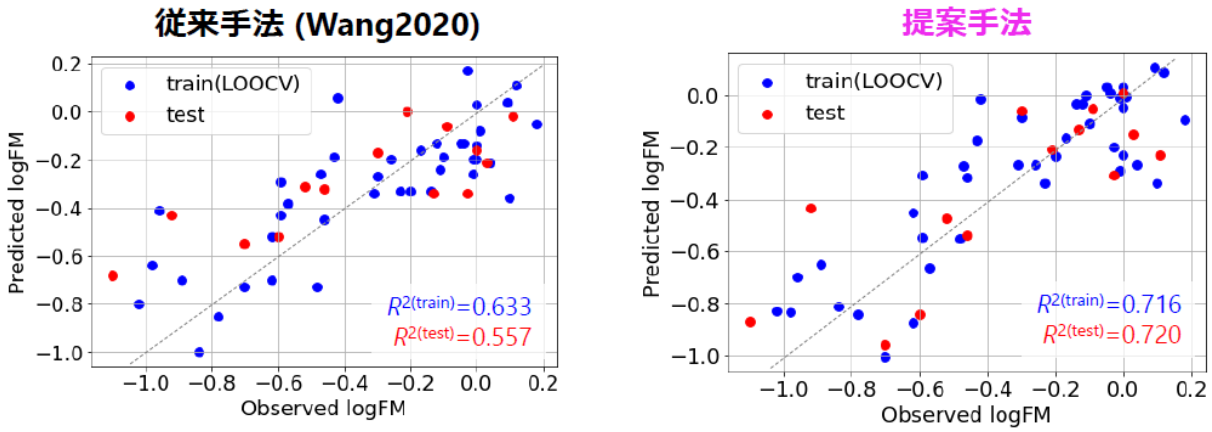


図 2. LogFM の予測結果の散布図

従来手法 [4] の予測結果 (左)、提案手法による予測結果 (右)。

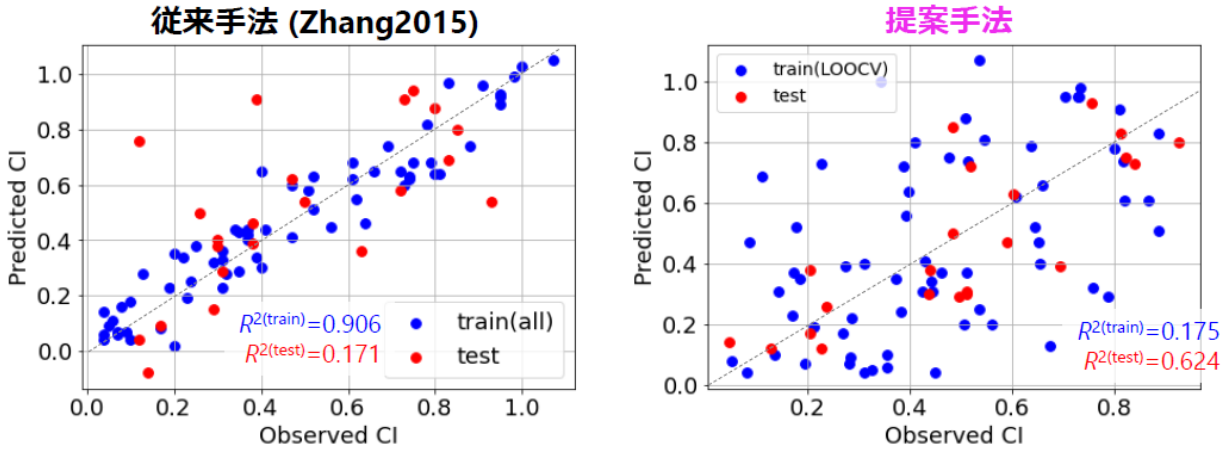


図 3. CI の予測結果の散布図

従来手法 [6] の予測結果 (左)、提案手法による予測結果 (右)。

2. 選択された記述子の考察

一般に、胎盤を透過しやすい薬剤は、分子量が小さく (<500)、脂溶性で、血漿タンパク質結合率が低い、という傾向が知られている [8]。提案手法で選択された記述子のほとんどは空間電荷に関わる記述子であり、脂溶性との関わりが考えられる。また、ZMIC1 は原子間結合の分岐やサイズと関係しており、分子量との相関が考えられる。Lipinski はリピンスキー則 (rule of five) を満たすか否かを表す記述子であり、分子量や脂溶性と直接関わっている。AATSC1c と ZMIC1 については従来手法 [4] でも選択された記述子であり、より血液胎盤関門透過性との関係が期待される。なお、血液脳関門透過性についても類似する法則性が示唆されているが [9]、血液脳関門透過性の予測値 (ADMET

Predictor で計算）と logFM や CI との強い相関は得られなかった。

3. ウェブサービスの利用例

構築したウェブサービスの概観を図 4 に示す。テキストボックスに SMILES 文字列を 1 化合物 1 行で入力し、submit ボタンを押すと logFM と CI の予測値が表示される。なお、SMILES として不適合な文字列が入力された場合は INVALID SMILES と表示されるようにした。

また、API リクエストの例として、https://pbpredictor.net/api/predict?smiles=CC(=O)NC1=CC=C(C=C1)O のように URL を指定すると、{"logFM":0.0919, "CI":0.5330, "smiles":"CC(=O)NC1=CC=C(C=C1)O"} のように json 形式で値が返ってくるようになっている。

4. 結論

本研究は、勾配ブースティング木 (GBDT) を用いて化合物の血液胎盤関門透過性を予測する手法を提案した。GBDT の枠組みの一つである LightGBM を用いて予測モデルを構築した結果、従来手法と比較して良好な予測性能を示した。提案手法はテストデータでも予測精度が低下しなかったことから、汎化性能もある程度担保できているものと考えられる。LightGBM は計算が軽量であることも利点の一つであり、ウェブサーバを介して予測結果を即座に返すようなサービスにも適している。本研究では PBPredictor と名付けた血液胎盤関門透過性予測ウェブサービスを試作し、公開した。なお、logFM と CI の両方を同時に考慮するマルチラベル学習の検討や、血液脳関門透過性や皮膚透過性など他の物性指標との関係、催奇形性や胎児毒性などの毒性との関係に関する考察が不十分であり、また化合物の部分構造に起因する血液胎盤関門透過性への影響を示唆できれば、薬剤設計においてより安全な部分構造の提案なども可能になると考えられ、これらは今後の課題である。

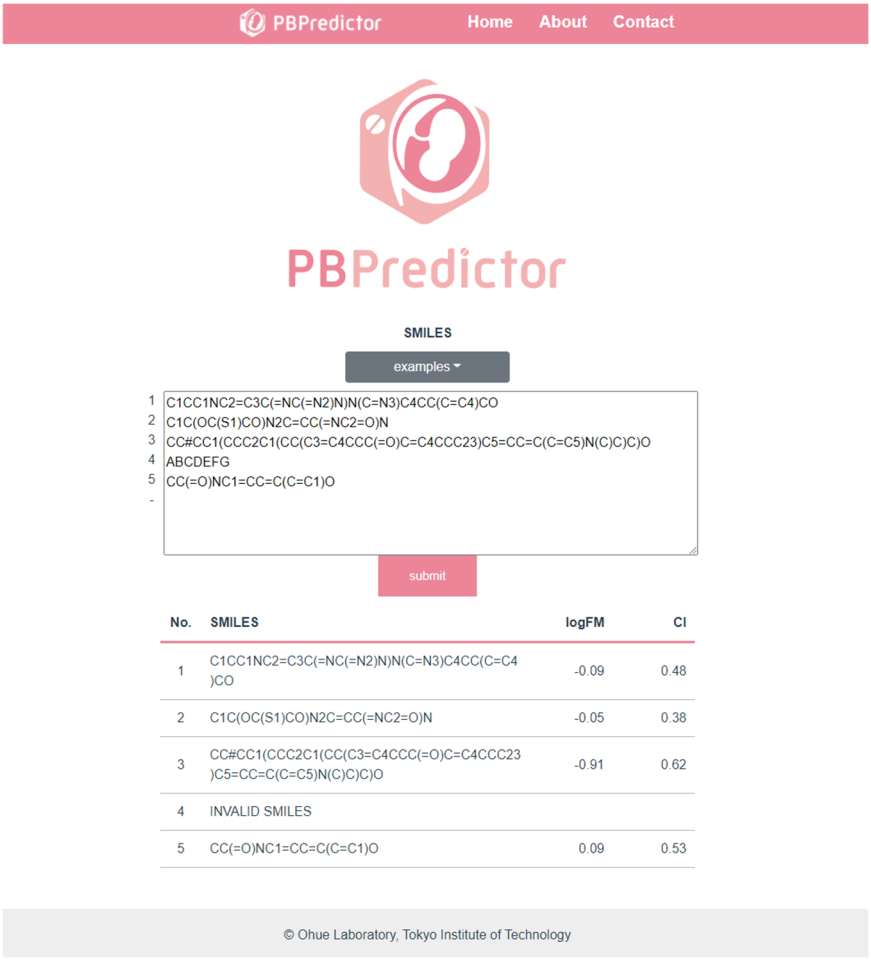


図 4. 予測ウェブサービス PBPredictor の概観

共同研究者

本研究の共同研究者は、東京工業大学情報理工学院情報工学系の古井海里である。

文　献

1) Myren M, Mose T, Mathiesen L, Knudsen LE. The human placenta--an alternative for studying foetal exposure. Toxicol In Vitro. 2007 Oct;21(7):1332-40. doi: 10.1016/j.tiv.2007.05.011. Epub 2007 Jun 7. PMID: 17624715.

2) Giaginis C, Theocharis S, Tsantili-Kakoulidou A. Current toxicological aspects on drug and chemical transport and metabolism across the human placental barrier. Expert Opin Drug Metab Toxicol. 2012 Oct;8(10):1263-75. doi: 10.1517/17425255.2012.699041. Epub 2012 Jul 11. PMID: 22780574.

3) Takaku T, Nagahori H, Sogame Y, Takagi T. Quantitative structure-activity relationship model for the fetal-maternal blood concentration ratio of chemicals in humans. Biol Pharm Bull. 2015;38(6):930-4. doi: 10.1248/bpb.b14-00883. PMID: 26027836.

4) Wang CC, Lin P, Chou CY, Wang SS, Tung CW. Prediction of human fetal-maternal blood concentration ratio of chemicals. PeerJ. 2020 Jul 21;8:e9562. doi: 10.7717/peerj.9562. PMID: 32742813; PMCID: PMC7380269.

5) Giaginis C, Zira A, Theocharis S, Tsantili-Kakoulidou A. Application of quantitative structure-activity relationships for modeling drug and chemical transport across the human placenta barrier: a multivariate data analysis approach. J Appl Toxicol. 2009 Nov;29(8):724-33. doi: 10.1002/jat.1466. PMID: 19728316.

6) Zhang YH, Xia ZN, Yan L, Liu SS. Prediction of placental barrier permeability: a model based on partial least squares variable selection procedure. Molecules. 2015 May 7;20(5):8270-86. doi: 10.3390/molecules20058270. PMID: 25961165; PMCID: PMC6272791.

7) 大上雅史, 古井海里. GBDT による化合物の血液胎盤関門透過性予測, 情報処理学会研究報告, 2021-BIO-65(8):1-3, Mar 2021.

8) Joshi AA, Vaidya SS, St-Pierre MV, Mikheev AM, Desino KE, Nyandege AN, Audus KL, Unadkat JD, Gerk PM. Placental ABC Transporters: Biological Impact and Pharmaceutical Significance. Pharm Res. 2016 Dec;33(12):2847-2878. doi: 10.1007/s11095-016-2028-8. Epub 2016 Sep 19. PMID: 27644937; PMCID: PMC5384842.

9) Abbott NJ, Patabendige AA, Dolman DE, Yusof SR, Begley DJ. Structure and function of the blood-brain barrier. Neurobiol Dis. 2010 Jan;37(1):13-25. doi: 10.1016/j.nbd.2009.07.030. Epub 2009 Aug 5. PMID: 19664713.