Vol.2022-MPS-138 No.49
Vol.2022-BIO-70 No.49
2022/6/29

多様な実験設定におけるランク学習を用いた化合物スクリーニングの性能評価

古井 海里[1]　大上 雅史[1,a)]

概要：情報検索分野で発展してきたランク学習手法が，リガンドベースのバーチャルスクリーニング (LBVS) に活用されている．ランク学習は順序関係を学習する機械学習の枠組みであり，異なる環境の実験データを統合するのに適しているという利点が注目されている．我々の取り組みで，複数の環境のアッセイデータが得られる状況において，新規標的に対するランキング予測が回帰モデルよりも予測精度の面で優れていることを明らかにした．しかし，標的と同ファミリーのタンパク質や標的そのものに関するアッセイ情報が全く無い，あるいは少し存在するなどの様々な状況下において LBVS にランク学習が適しているかは未知であった．また，従来研究で用いられていた NDCG (Normalized Discounted Cumulative Gain) 指標は他のモデルと比較して優れているかのみを評価するため，予測モデルがランダムな結果よりも悪い結果を出したかどうか考慮できないという問題があった．本研究は，多様なアッセイ情報の保有状況を想定した学習データを用いて，ランク学習手法の LBVS 性能を検証した．結果として，ランク学習手法はランキング予測において回帰と同等以上の予測精度であり，特に標的と関連するアッセイが複数存在する状況でランク学習のデータ統合が有効である可能性が示唆された．さらに，ランダムな予測を基準とした予測性能を評価する目的で新たに提案したランキング指標「Normalized Enrichment Discounted Cumulative Gain (NEDCG)」が，複数のテストデータについて予測結果の良し悪しを評価するのに適していた．

1. 序論

新規薬剤が承認されるまでの開発コストや開発期間は年々増加しており，1 薬剤あたりおよそ 3 千億円必要であるという推計 [1] や，リード化合物の同定から臨床試験までに開発期間が 10 年以上であるという報告 [2] がある．新規薬剤開発の初期段階では，バーチャルスクリーニングが行われる．バーチャルスクリーニングとは，創薬の標的となるタンパク質に対する活性を持つ化合物を，計算機によって大規模な化合物ライブラリーから探索することである．バーチャルスクリーニングの予測精度を向上したり，バーチャルスクリーニングの適用範囲を広げて様々な活性データに対して使えるようにしたりすることは，新規薬剤開発に必要なコストを少しでも削減するのに欠かせない．

バーチャルスクリーニングの手法の 1 つとして，既にアッセイで得られた活性情報を活用するリガンドベースの手法がある．リガンドベースのバーチャルスクリーニング (Ligand-based virtual screening, LBVS) は主に，回帰予測 [3, 4] や分類予測 [5–7] などの機械学習手法に基づいている．しかし，近年では活性値の順序関係に基づくランク学習を LBVS に適用するアプローチが提案されている [8–16].

ランク学習を LBVS に適用する利点は，大きく分けて 2 つある [12, 13, 16]．1 つは，ランキング予測に関して回帰タスクよりも予測精度が高いという点である [12]．創薬では，バーチャルスクリーニングによって選ばれた活性を持つ可能性のある化合物は，最終的にアッセイを行うことで実際に活性を持つか確かめられる．そのため，真に望ましいバーチャルスクリーニングとは正確な活性値を予測することそのものではなく，予測の上位に少しでも活性の強い化合物を順序付けることであると言い換えることができる．したがって，順序に基づいて予測を行うランク学習はバーチャルスクリーニングの目的に適している．もう 1 つは，ランク学習は比較可能なグループ内の順序関係に基づいて学習するため，異なる分布の実験情報を統合するのに適しているという点である．50%阻害濃度 IC_{50} など，生化学アッセイに基づく親和性指標は環境によって値が大きく変動するため，回帰手法では異なる環境のアッセイデータを統合するのは困難である [12]．新規標的を対象としたバーチャルスクリーニングでは，標的と関連するタンパク質とのアッセイを複数用いることで，ランキング予測精度

[1] 東京工業大学 情報理工学院 情報工学系
Department of Computer Science, School of Computing, Tokyo Institute of Technology
[a)] ohue@c.titech.ac.jp

が向上することが示唆されており [12, 16]，このように環境の異なる複数のアッセイを統合して学習に利用できるランク学習は，従来手法では難しいとされていた新規標的への化合物スクリーニングに特に効果があるという点で注目されている．

ここで，新規標的を目的とした LBVS にランク学習を適用するにあたり，本稿で取り組んだ既存研究の課題を 2 つ挙げる．

まず，先行研究では Normalized Discounted Cumulative Gain (NDCG) [17–19] と呼ばれる評価指標を用いてランキング予測の性能を評価している [12–14] が，NDCG が実際に LBVS に適しているかは十分に議論されていない．NDCG は，異なるモデルを比較してどちらが優れているか評価したり，ハイパーパラメータのチューニングを行ったりするのには適している．しかしながら，NDCG は本来情報検索分野のランク学習性能評価で用いられていた指標であり，バーチャルスクリーニングで想定する状況に適さない点がある．例えば，NDCG の値がそれ単体で意味を持つのは，正しい順序を予測したときの NDCG = 1 と，列挙した全ての目的変数が 0 のときの NDCG = 0 の場合に限る．ランダムな予測を行った時（すなわち学習するための知識が何もない状態）の NDCG の値は評価するデータごとに異なる．したがって，NDCG の値単体では予測モデルがどの程度有益な改善をもたらしているか分からない．例えば情報検索分野では，既に十分な規模のデータセットがある事が多いため，ランダムな予測よりは改善していることが想定される．しかし，ランク学習が対象とする新規標的のための LBVS の場合では，アッセイ済みの化合物は少なくデータセットを増やすのが難しいため，分布外の分子の活性を予測できないという状況が多く，モデルがランダムな予測より改善しているかは重要な情報である．また，バーチャルスクリーニングでは化合物のランキングにおける enrichment の効果が重要視されてきた [20]．しかし，前述した先行研究 [12–14] ではこのような分野の違いを考慮せずに，NDCG をランキング予測に対するバーチャルスクリーニングの評価指標として用いている．

また，既存研究における新規標的に対する LBVS の実験設定は極端なケースのみ評価されており，より一般的な状況を考慮すべきである．Zhang ら [12] や我々の以前の研究 [16] によって，新規標的に対する LBVS ではランク学習手法が回帰学習手法よりも適していると結論付けられている．これは，ランク学習が異なる環境のアッセイに関する情報を別々に学習できるという特性が，新規標的と関連するタンパク質のアッセイを利用した LBVS（標的タンパク質と同じファミリーのタンパク質に関するアッセイ情報をスクリーニングに利用する）に寄与するためである．一方で，新規標的との関連性の低いタンパク質を学習のみを活用した場合でも，いくつかの標的で学習が上手くいく場合があると言及している [12] が，予測精度が実際にどの程度良い値であるかについて定量的な分析はしていない．この 2 つのケースは極端であり，実際には既に数十件程度標的に関する阻害剤が得られているケースや，同じタンパク質ファミリーについてのアッセイが十分に無いケースなどがあり得るため，より多様な実験設定でランキング予測を評価する必要がある．

本研究では，新規標的を目的としたバーチャルスクリーニングに優れたランク学習が望ましい予測精度を達成するか，回帰学習による手法より優れているか等を検証した．Zhang らの学習戦略 [12] を拡張して，新規標的に関するアッセイが僅かにある場合や，標的タンパク質との関連性の低いタンパク質のアッセイを含めた場合など，複数の実験設定がランキング予測性能に与える影響を評価した．

本稿の主張は以下の 3 点である．

- 従来の NDCG による評価は比較のみに基づいて判別している．ランキングに基づいた LBVS では，新規に提案した NEDCG (Normalized Enrichment DCG) のように，ランダム予測からの改善を考慮することが重要である．
- 新規標的に対する LBVS の複数の実験設定において，回帰学習手法との比較を行った．結果として，ランク学習は回帰学習と同程度か上回る予測精度であり，ランク学習は新規標的に対する LBVS に適している．
- 新規標的に関するアッセイがある場合や，標的との関連性が高い（同じタンパク質ファミリーに属する）タンパク質のアッセイが十分にあるとき，ランキング予測は成功する．また，これらのアッセイの組み合わせによって，バーチャルスクリーニング性能は大きく向上させることができる．標的と関連しないタンパク質のアッセイ情報を含めることは，基本的に予測精度を下げる要因になった．

2. 手法

2.1 Normalized DCG (NDCG)

Normalized Discounted Cumulative Gain (NDCG) [17–19] は，ランク学習モデルの性能を測定するための主要な評価指標の 1 つであり，情報検索分野において，ランキング性能を評価するのに用いられている [17–19]．N 件のグループの上位 K 件についての NDCG@K は以下で求められる．

$$\begin{aligned} \mathrm{NDCG@}K &= \sum_{i=1}^{K} \frac{G_i}{D_i} \\ G_i &= \frac{gain_i}{\mathrm{maxDCG}} \\ gain_i &= 2^{y_i} - 1 \\ D_i &= \log_2(i+1) \end{aligned}$$

ここで, i は予測順位, y_i は i の目的変数, maxDCG は正規化定数で, 正しい予測をしたときの Discounted Cumulative Gain (DCG) である．NDCG は，ランキング上位 K 件に正しいものを列挙できるほど 1 に近づき，最も悪い場合は 0 になる．

2.2 Normalized Enrichment DCG (NEDCG)

2.1 節で説明した NDCG は，異なるモデルを比較してどちらが優れているか評価したり，ハイパーパラメータのチューニングを行ったりするには適する．しかし，NDCG はそれ単体でどのくらい有益な予測をしているか見積もることができないという欠点がある．分類タスクで用いられる AUROC (Area Under the Receiver Operating Characteristic Curve) は 0.5 であれば，その予測モデルはランダムな予測と変わらないことを意味する．我々は AUROC の着想を元にして，NDCG からランダムな予測（学習を行う前の状態）の影響を取り除いた Normalized Enrichment DCG (NEDCG) を新たに提案した [16]．上位 K 件に対する NEDCG@K を以下に示す．

$$\mathrm{NEDCG@}K = \frac{\mathrm{DCG@}K - \mathrm{randomDCG@}K}{\mathrm{maxDCG@}K - \mathrm{randomDCG@}K} \quad (1)$$

$$\mathrm{DCG@}K = \sum_{i=1}^{K} \frac{gain_i}{\log_2(i+1)} \quad (2)$$

ここで, randomDCG@K はランダムな予測をしたときの上位 K 件に関する DCG である．実際には, randomDCG@K は N 件のグループの平均ゲイン $gain_{mean}$ によって以下のように算出する．

$$\mathrm{randomDCG@}K = \sum_{i=1}^{K} \frac{gain_{mean}}{\log_2(i+1)}$$

$$gain_{mean} = \frac{1}{N}\sum_{j=1}^{N} gain_j$$

2.3 GBDT と LambdaMART

GBDT（Gradient boosting decision tree, 勾配ブースティング決定木）は，決定木による弱学習器を繰り返しアンサンブルしてコスト関数を最小化する機械学習アルゴリズムである．GBDT は近年機械学習コンペティション等で広く用いられており，XGBoost [21] や LightGBM [22] などの実用的な実装がいくつか存在する．本研究では GBDT の実装として LightGBM を用いた．

LambdaMART は，情報検索分野で広く用いられているランク学習アルゴリズムである [23]．lambdarank [24] というランキング損失関数で GBDT を学習する．lambdarank は，NDCG を直接最適化することを目的としている．ここで，N 件のグループについての lambdarank 損失関数の定義を以下に示す．

$$l(\boldsymbol{y}, \boldsymbol{s}) = \sum_{y_i > y_j} \rho_{ij} \left|G_i - G_j\right| \log\left(1 + e^{-\sigma(s_i - s_j)}\right)$$

$$\rho_{ij} = \left|\frac{1}{D_i} - \frac{1}{D_j}\right|$$

ここで，s_i と s_j は順序 i，j に関する予測したランキングスコアである．lambdarank は $\Delta\mathrm{NDCG}_{ij} = \rho_{ij}|G_i - G_j|$, すなわち，順序 i と j を入れ替えたときの NDCG の変化に基づいてペナルティを与えることにより，NDCG を改善するように順序関係を学習する．

3. 実験

本実験では，以下のランク学習・回帰学習手法の 2 つの予測モデルについて比較を行う．

(1) lambdarank（ランク学習）
　lambdarank 損失関数を用いた GBDT モデル
(2) regression（回帰学習）
　L_2 loss を用いた GBDT モデル

3.1 複数の実験設定でのランク学習評価

まず，3 通りの訓練データの設定について示す．

実験 A　標的タンパク質・単一アッセイ（30–100 化合物程度）のデータセット
実験 B　標的と同一ファミリーのタンパク質・単一アッセイのデータセット
実験 C　標的と同一ファミリーのタンパク質・複数のアッセイデータセット

標的タンパク質を直接訓練データとして用いる実験 A や訓練データ数が十分にある実験 C が，実験 B よりも予測精度が高いことが期待される．

さらに，各実験に対して 2 つの条件を試みた．

Case 1　訓練データにデータを追加しない
Case 2　訓練データに標的の関連性の低いタンパク質とのアッセイ情報を追加

Case 2 は，異なるファミリーの実験情報を追加することが予測精度の向上に寄与するかを調べるために行う．以降，実験 A，Case 1 の設定を実験 A1 と表記する．

また，新規標的に関連する親和性情報が全くない状況を想定した実験 D（Case 2 で追加されるデータセットのみで学習する）と，実験 A1 と実験 C1 の学習に有効と考えられるデータセットを十分に利用した実験 E を行う．

実験 D　標的タンパク質と関連する情報はなく，標的と異なるファミリーのタンパク質とのアッセイ情報のみを用いたデータセット
実験 E　実験 A1（標的タンパク質について単一アッセイ）と，実験 C1（標的と同一ファミリーのタンパク質について複数のアッセイ）を組み合わせたデータセット

実験 D の条件は Zhang らの戦略に基づく [12]．実験 D の設定でランキング予測性能がランダム予測よりも改善する

Vol.2022-MPS-138 No.49
Vol.2022-BIO-70 No.49
2022/6/29

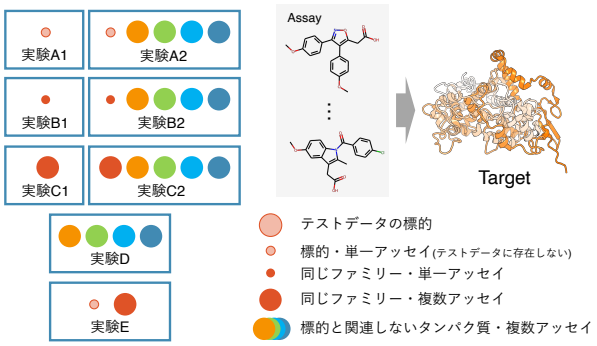

図 1 各実験設定の概念図

表 1 テストデータに用いるデータセットと，実験 A，実験 B の学習に用いる単一アッセイのサイズ

Target name	Assay size
Cyclooxygenase-1 (CO-1)	34
Cyclooxygenase-2 (CO-2)	36
Estrogen receptor alpha (ER-α)	107
Estrogen receptor beta (ER-β)	108
Monoamine oxidase A (MO-A)	40
Monoamine oxidase B (MO-B)	53

ならば，どのような標的に対しても万能に対応できる学習モデルが構築できることを意味する．

図 1 に，本実験で行う実験の概念図を示す．

3.2 データセット

Matsumoto ら [15] のデータセットを参考に，IC_{50} についての生化学アッセイ情報を 24 タンパク質に関する ChEMBL データベース [25] から収集した．目的変数は IC_{50} の負の対数を取り，$pIC_{50}(= -\log_{10} IC_{50})$ とし，inactive と表記されている化合物については文献 [14] を参考に $pIC_{50} = 0$ とした．また，アッセイは inactive でないデータが 10 件以上，5 件以上の異なる pIC_{50} を持つ化合物に限り，アッセイ内での化合物の重複は除去した．

また，テストデータとして**表 1** の 6 つのタンパク質を選択した．実験 B1 の CO-1 のアッセイがテストデータの時，CO–2 の 36 化合物が学習で用いられることを意味する．

化合物の記述子として，mordred [26](ver 1.2.0) による 1-D/2-D 化合物記述子 (1,613 次元) を利用した．ただし，訓練データの半数以上の化合物で Null 値を取る特徴量を除去し，残りの 1447 記述子を用いた．また，複数タンパク質を訓練データに含む場合はタンパク質情報として PyBioMed [27] (ver 1.0) によって生成した CTD 記述子を 147 次元の特徴量を加えた．

3.3 モデルの学習

GBDT モデルの学習は，チューニング時はラウンド数を 100，評価用の学習時には `early_stopping=1000` とし，検証の NDCG@10 最大となるラウンド数の 1.1 倍のラウンドのときのモデルで予測を行う．

ハイパーパラメータのチューニングのときは 3-fold の交差検証を行う．このとき，データセットの特性に合わせて，分割を以下のように工夫する．

- 単一アッセイ (実験 A1, 実験 B1) のデータセット: 全体をランダムに分割する．
- 標的と関連する複数アッセイ (実験 C) のデータセット: アッセイ単位で分割する．
- 標的と関連しないタンパク質に関するアッセイ (実験 D) のデータセット: タンパク質単位で分割する．

なお，異なる条件のデータセットを組み合わせた時はそれぞれの方法でデータセットの分割を行った後に組み合わせる．例えば実験 E（テストデータと同じ標的の単一アッセイ+同じファミリーの複数アッセイ）の場合は，同じ標的の単一アッセイに関してはランダムに，同じファミリーの複数アッセイについてはアッセイ単位で分割し，それらを結合させる．

GBDT のハイパーパラメータは，木の最大の深さ，決定木の葉の数，葉に割り当てるデータの最小数をグリッドサーチで探索した．また，学習率については，パラメータチューニング時には 0.1，学習時には 0.05 を用いた．lambdarank の損失関数を用いるとき，`lambdarank_truncation_level` を 30，label gain の刻み幅 δ を 0.01 に固定する．なお，実験 A2 および B2 では，標的と関連しないタンパク質に関するデータの重要度を低くするため，学習の重みを 99 : 1 にした．

4. 結果

各実験における予測性能の評価には，グループ内の上位 20%に関する予測性能を表す NDCG%20 と NEDCG%20 を用いる．これは，例えば 50 件の化合物を持つアッセイならば，NDCG@10, NEDCG@10 を意味する．また，以降では medianNEDCG を NEDCG 分布の中央値とする．

図 2 および**図 3** に，ランク学習モデル (lambdarank) と回帰学習モデル (regression) について，実験 A1 から E までの NDCG%20 と NEDCG%20 の箱ひげ図をそれぞれ示す．図 2，図 3 どちらについても，実験 E，実験 A2, 実験 C1 の順に予測精度が高い．

5. 考察

5.1 実験設定による予測精度への影響

図 3 の結果より，実験ごとの medianNEDCG が 0 を大きく上回るのは，実験 E（最良の結果），実験 A2, 実験 C1 であった．各実験の medianNEDCG を見ると，ランク学習は回帰学習手法と同程度か上回る予測精度であることが分かる．特に，標的タンパク質か同じファミリーの複数アッセイを含む実験 C1，実験 C2，実験 E の 3 つの実験では，ランク学習が回帰学習を上回る傾向がある．これらの実験において，ランク学習の複数アッセイ統合による効果

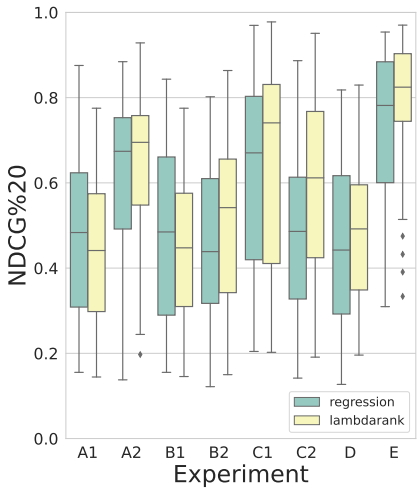

図 2 各実験設定ごとの NDCG%20 の比較

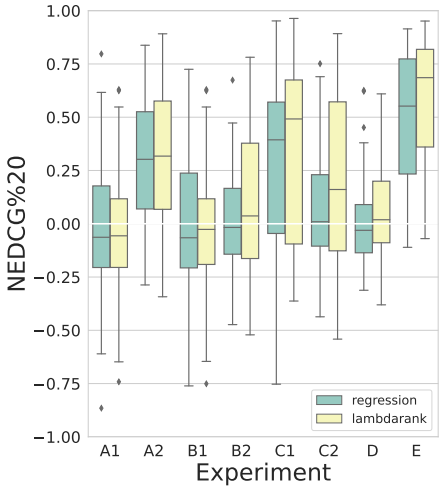

図 3 各実験設定ごとの NEDCG%20 の比較

が現れたためだと考えられる．

NEDCG が良い活性を持つ阻害剤をどの程度予測の上位に凝縮できるかを示す指標であるということを考えれば，本研究のような実験設定においてランク学習手法が回帰手法と同程度かそれ以上のバーチャルスクリーニング性能があるといえる．

5.2 NEDCG によるランキング性能評価

図 3 の我々の提案した NEDCG%20 による評価では，予測がランダムより悪いときは 0，全て正しいときは 1 になるように調整しているため，medianNEDCG が 0 付近であれば予測精度がランダム予測より悪い可能性があることを意味する．そのため，モデルの予測精度がどの程度有用な値であるかを直感的に理解することができる．一方，図 2 の既存の NDCG%20 による評価では，他の予測モデルの値と比較によってしか予測の良さが分からない．

また，NEDCG は NDCG に比べて，複数のアッセイを総合して評価するのに適している．**図 4** のアッセイごとのランダム予測をしたときの NDCG (randomNDCG) 分布

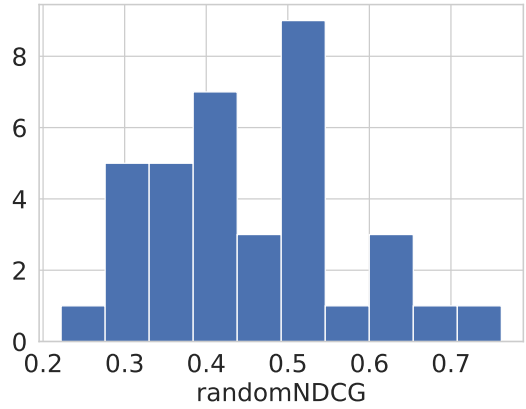

図 4 アッセイごとの randomNDCG の頻度

が示す通り，アッセイごとの randomNDCG は 0.2 から 0.7 までばらつきがある．すなわち，randomNDCG が 0.2 のとき予測モデルが NDCG を 0.5 までランキングを改善した場合と，randomNDCG が 0.7 のとき予測モデルが NDCG を 0.6 までランキングを悪化させた場合では，NDCG は後者のランキングを高く評価してしまう．そのため，複数のアッセイを総合して予測精度を評価するには，NEDCG のようにアッセイごとに存在するランダム予測のバイアスを除去することが重要である．

6. 結論

本稿では，新規標的を目的としたバーチャルスクリーニングにおいて，どのような実験設定で十分なランキング予測精度が得られるか検証した．例えば，標的タンパク質とのアッセイ情報の有無，標的タンパク質と関連しないタンパク質のアッセイ情報が学習に寄与するか等，複数の条件で回帰学習手法とランク学習手法を比較した．

結果として，ランク学習手法は回帰学習手法と同程度かそれ以上のランキング予測精度であり，新規標的を目的としたバーチャルスクリーニングに適していることが示唆された．特に，標的タンパク質に関連したアッセイ情報が十分にあるとき，ランク学習のランキング予測性能は回帰学習手法よりも優れていた．また，標的に関連しないタンパク質データを含めることは，多くの場合で予測精度を悪化させる要因となった．それから，標的アッセイに関する情報が十分にないときは，実験 E のように同じファミリーのタンパク質情報を組み合わせることで予測精度を大きく改善できる．

我々が提案した NEDCG は，値が 0 以上であれば少なくともランダム予測よりも良いことを表すため，既存の NDCG よりも予測の良さを直感的に評価できる．

本研究では GBDT によるランク学習を用いたが，近年普及し始めている深層学習モデル [28] においてもランク学習によるバーチャルスクリーニング手法が有効かどうか検証する必要がある．深層学習モデルによるランク学習が必

ずしも GBDT や RankSVM などの記述子ベースの機械学習手法より優れているとは限らないが，各手法の特性を理解するために重要である．

また，ランク学習はケモインフォマティクスの他のタスクに対しても適用できる可能性がある．例えば，ADMET (吸収, 分布, 代謝, 排泄, 毒性) 特性の予測 [29] や，QSAR (Quantitative Structure-Activity Relationship) [15]，薬物標的親和性予測 [30] が挙げられる．これらのタスクにおいて，回帰予測をするにはデータ数が少ない場合や，環境の異なるアッセイデータを学習データとして統合する場合において，ランキング予測の有効性が期待できる．

謝辞 本研究は，JST 創発的研究支援事業 (JPMJFR216J)，JST 戦略的創造研究推進事業 ACT-X (JPMJAX20A3)，科研費 (20H04280)，AMED BINDS (JP22ama121026) の支援を受けて実施された．

参考文献

[1] A. Mullard: New drugs cost US $2.6 billion to develop, *Nat. Rev. Drug Discov.*, Vol. 13, p. 877 (2014).
[2] C. M. Song, S. J. Lim, J. C. Tong: Recent advances in computer-aided drug design, *Brief. Bioinform.*, Vol. 10, No. 5, pp. 579–591 (2009).
[3] J. Ma *et al.*: Deep neural nets as a method for quantitative structure–activity relationships, *J. Chem. Inf. Model.*, Vol. 55, No. 2, pp. 263–274 (2015).
[4] P. J. Ballester, J. B. Mitchell: A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking, *Bioinformatics*, Vol. 26, No. 9, pp. 1169–1175 (2010).
[5] E. Byvatov *et al.*: Comparison of support vector machine and artificial neural network systems for drug/nondrug classification, *J. Chem. Inf. Comput. Sci.*, Vol. 43, No. 6, pp. 1882–1889 (2003).
[6] N. Schneider *et al.*: Gradual in silico filtering for drug-like substances, *J. Chem. Inf. Model.*, Vol. 48, No. 3, pp. 613–628 (2008).
[7] F. Nigsch *et al.*: Ligand-target prediction using Winnow and naive Bayesian algorithms and the implications of overall performance statistics, *J. Chem. Inf. Model.*, Vol. 48, No. 12, pp. 2313–2325 (2008).
[8] A. M. Wassermann, H. Geppert, J. Bajorath: Searching for target-selective compounds using different combinations of multiclass support vector machine ranking methods, kernel functions, and fingerprint descriptors, *J. Chem. Inf. Model.*, Vol. 49, No. 3, pp. 582–592 (2009).
[9] T. Joachims: Optimizing search engines using clickthrough data, *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 133–142 (2002).
[10] S. Agarwal, D. Dugar, S. Sengupta: Ranking chemical structures for drug discovery: a new machine learning approach, *J. Chem. Inf. Model.*, Vol. 50, No. 5, pp. 716–731 (2010).
[11] F. Rathke *et al.*: StructRank: a new approach for ligand-based virtual screening, *J. Chem. Inf. Model.*, Vol. 51, No. 1, pp. 83–92 (2011).
[12] W. Zhang *et al.*: When drug discovery meets web search: learning to rank for ligand-based virtual screening, *J. Cheminform.*, Vol. 7, No. 1, p. 5 (2015).
[13] S. D. Suzuki, M. Ohue, Y. Akiyama: PKRank: a novel learning-to-rank method for ligand-based virtual screening using pairwise kernel and RankSVM, *Artif. Life Robot.*, Vol. 23, No. 2, pp. 205–212 (2018).
[14] M. Ohue, S. D. Suzuki, Y. Akiyama: Learning-to-rank technique based on ignoring meaningless ranking orders between compounds, *J. Mol. Graph. Model.*, Vol. 92, pp. 192–200 (2019).
[15] K. Matsumoto, T. Miyao, K. Funatsu: Ranking-Oriented Quantitative Structure–Activity Relationship Modeling Combined with Assay-Wise Data Integration, *ACS Omega*, Vol. 6, No. 18, pp. 11964–11973 (2021).
[16] K. Furui, M. Ohue: Compound virtual screening by learning-to-rank with gradient boosting decision tree and enrichment-based cumulative gain, *arXiv preprint arXiv:2205.02169* (2022).
[17] K. Järvelin, J. Kekäläinen: IR evaluation methods for retrieving highly relevant documents, *ACM SIGIR Forum*, Vol. 51, No. 2, pp. 243–250 (2017).
[18] K. Järvelin, J. Kekäläinen: Cumulated gain-based evaluation of IR techniques, *ACM Trans. Inf. Syst.*, Vol. 20, No. 4, pp. 422–446 (2002).
[19] C. Burges *et al.*: Learning to rank using gradient descent, *Proc. 22nd Int. Conf. Mach. Learn*, pp. 89–96 (2005).
[20] L. Han *et al.*: A support vector machines approach for virtual screening of active compounds of single and multiple mechanisms from large libraries at an improved hit-rate and enrichment factor, *J. Chem. Inf. Model.*, Vol. 26, No. 8, pp. 1276–1286 (2008).
[21] T. Chen, C. Guestrin: XGBoost: A scalable tree boosting system, *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 785–794 (2016).
[22] G. Ke *et al.*: LightGBM: A highly efficient gradient boosting decision tree, *Adv. Neural Inf. Process. Syst.*, Vol. 30, pp. 3146–3154 (2017).
[23] C. J. Burges: From RankNet to LambdaRank to LambdaMART: An Overview, Technical Report MSR-TR-2010-82 (2010).
[24] C. Burges, R. Ragno, Q. Le: Learning to rank with non-smooth cost functions, *Adv. Neural Inf. Process. Syst.*, Vol. 19, pp. 193–200 (2006).
[25] D. Mendez *et al.*: ChEMBL: towards direct deposition of bioassay data, *Nucleic Acids Res.*, Vol. 47, No. D1, pp. D930–D940 (online), DOI: 10.1093/nar/gky1075 (2018).
[26] H. Moriwaki *et al.*: Mordred: a molecular descriptor calculator, *J. Cheminform.*, Vol. 10, No. 1, p. 4 (2018).
[27] J. Dong *et al.*: PyBioMed: a python library for various molecular representations of chemicals, proteins and DNAs and their interactions, *J. Cheminform.*, Vol. 10, No. 1, p. 16 (2018).
[28] D. Jiang *et al.*: Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models, *J. Cheminform.*, Vol. 13, No. 1, p. 12 (2021).
[29] J. Li *et al.*: Plasma protein binding prediction focusing on residue-level features and circularity of cyclic peptides by deep learning, *Bioinformatics*, Vol. 38, No. 4, pp. 1110–1117 (online), DOI: 10.1093/bioinformatics/btab726 (2022).
[30] X. Ru *et al.*: NerLTR-DTA: drug–target binding affinity prediction based on neighbor relationship and learning to rank, *Bioinformatics*, Vol. 38, No. 7, pp. 1964–1971 (2022).