

論文 / 著書情報
Article / Book Information

題目(和文)	薬剤標的タンパク質予測における行列因子分解モデルの改良とその拡張
Title(English)	Improved Matrix Factorization Model for Drug-Target Prediction and Its Enhancement Techniques
著者(和文)	伴兼弘
Author(English)	Tomohiro Ban
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11118号, 授与年月日:2019年3月26日, 学位の種別:課程博士, 審査員:秋山 泰,篠田 浩一,村田 剛志,関嶋 政和,石田 貴士
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11118号, Conferred date:2019/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

論文要旨

THESIS SUMMARY

専攻：	計算工学	専攻	申請学位（専攻分野）：	博士	（ 工学 ）
Department of			Academic Degree Requested	Doctor of	
学生氏名：	伴 兼弘		指導教員（主）：	秋山 泰	
Student's Name			Academic Supervisor(main)		
			指導教員（副）：		
			Academic Supervisor (sub)		

要旨（和文 2000 字程度）

Thesis Summary (approx.2000 Japanese Characters)

近年、既存薬の再利用や副作用回避を目的として、薬剤と標的タンパク質の相互作用 (Drug-Target Interaction, DTI) を網羅的に予測する手法の研究開発が盛んに行われてきた。DTI 予測は、薬剤の化学構造ならびに標的タンパク質の構造情報、そして既知の相互作用情報を用いて数理モデルを学習する機械学習手法と、標的タンパク質の立体構造情報を用いて物理化学的シミュレーションを行うドッキング計算手法に大分することができる。機械学習手法では、計算に要する時間が比較的短いため、大規模な予測が可能である。しかし、既知の情報の偏りから相互作用の発見が困難なものが存在し、それらによる予測精度の低下が問題となっていた。また、ドッキング計算手法においては、探索空間がタンパク質の形状により変化することから、従来の探索グリッドの配置方法では予測精度が悪化することが問題とされていた。本論文では、機械学習手法ならびにドッキング計算手法における予測精度の高精度化を実現することを目標として、ベイズ統計を用いた行列因子分解モデルに基づく新規予測手法ならびにドッキング計算の新規グリッド配置方式を提案した。

はじめに、薬剤とタンパク質の相互作用を予測する既存研究について述べている。機械学習に基づく手法では、カーネル回帰モデルが最も初期に提案され、その後、行列因子分解モデルの有効性が示され、行列因子分解に基づく様々な手法が提案されてきたことが述べられている。特にロジスティック行列因子分解モデルを用いた NRLMF は最先端手法の一つとして知られており、本研究の基礎となる手法であるため、その詳細について記述されている。また、分子シミュレーションに基づく手法は、標的タンパク質の立体構造情報を活用することで、薬剤とタンパク質の結合状態とその結合エネルギーを予測できるため、機械学習による予測の後処理として、相互作用部位の予測の役割を担うことが多いことを述べている。

次に、ベイズ統計を用いた行列因子分解モデルに基づく従来の手法に対して、共役事前分布を用いた予測スコアの再計算を導入することにより、相互作用に関する情報が乏しい状況下において予測精度を改善できる新規手法を提案している。従来手法 NRLMF の予測において偽陰性となる薬剤と標的タンパク質の組は、既知の相互作用の情報が少ない場合に特に多く表出する傾向を発見したことから、提案手法ではそれを解消する方向にモデルを修正することが考案されている。提案手法 NRLMF β では、NRLMF の予測スコアとその特徴量から一意に定まるベータ分布の期待値を、新たな予測スコアとする。計算機実験を通じて、偽陰性となる組の割合が減少し、全体的に予測精度が改善されたことが述べられている。

また、行列因子分解モデルにおけるハイパーパラメータの探索を、ベイズ最適化に基づき高速化する方法について述べている。行列因子分解モデルは、他の手法と比較してハイパーパラメータの数が多く、その探索に時間を要する。従来はグリッドサーチが用いられていたが、ベイズ最適化を適用することで、精度を落とさずに、ハイパーパラメータ探索の時間を削減できることを述べている。

一方、ドッキング計算の性能を向上することを目的として、予測された結合部位を被覆するように複数のグリッドを配置する方式を提案している。これによりドッキング計算の予測精度が改善することが述べられている。

行列因子分解モデルの弱点であった新規化合物に対する予測を、深層学習の技法により解決できる可能性について説明している。深層学習では、以前の行列に基づいて作成した学習済みのモデルを使うことで、新規化合物に対する特徴ベクトルを計算することができる。また、行列因子分解モデルの尤度関数に基づいて、深層学習モデルの損失関数を設計できる可能性についても述べている。

以上を要するに、本論文は、薬剤と標的タンパク質の間の相互作用を予測する手法を提案しており、特に相互作用に関する情報が乏しい状況下における予測精度を改善できる点において新規性が高い。また、提案された手法は薬剤標的相互作用予測に限らず、行列因子分解モデルに基づく、他の多くの分野のデータ解析にも応用が可能と考えられ、広く利用されることが期待される。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).

論文要旨

THESIS SUMMARY

専攻：	計算工学	専攻
Department of		
学生氏名：	伴 兼弘	
Student's Name		

申請学位 (専攻分野)：	博士 (工学)
Academic Degree Requested	Doctor of
指導教員 (主)：	秋山 泰
Academic Supervisor(main)	
指導教員 (副)：	
Academic Supervisor(sub)	

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words)

Drug target Interaction (DTI) prediction is one of the most important processes in drug discovery. The purpose of DTI prediction is to identify off-target proteins that cause some side effects and to discover new medical effects with regard to chemical compounds. Machine learning and docking simulation are known as two main approaches. In the machine learning approach, a large-scale prediction is possible since the time required for calculation is short. However, there are some difficulties in finding interactions due to unevenness of the known information, and deterioration of prediction accuracy by them is a problem.

In this thesis, we proposed a novel DTI prediction model called Beta-distribution-rescored NRLMF (NRLMF β). By introducing the recalculation of the prediction score using conjugate prior distribution for the conventional method based on matrix factorization model, it is possible to improve prediction accuracy in situations where information on interactions is poor. Since we found a tendency to express a large number of pairs of drugs and target proteins that are false negatives in the prediction of NRLMF especially when there is little information on known interactions, in the proposed method, a model has been devised to correct. In the proposed method NRLMF β , the expectation value of the beta distribution uniquely determined from the feature quantity and the score of NRLMF are set as new prediction scores. Through computer experiments, it is stated that the percentage of false negatives is reduced, and overall prediction accuracy has been improved.

In summary, this thesis proposes a method to predict the interactions between drugs and target proteins, and is highly novel in that prediction accuracy can be improved especially in a situation where information on interaction is poor. In addition, the proposed method is not limited to drug target interaction prediction, but is expected to be applicable to data analysis in many other fields based on a matrix factorization model and is expected to be widely used.

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note：Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1 copy of 800 Words (English).

注意：論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).