

# スパース統計モデルによる薬物-疾患ネットワークの予測

岩田浩明<sup>†1</sup> 山西芳裕<sup>†1</sup>

近年の新薬開発の行き詰まりを打開する方法として、既存薬の作用を徹底的に調べあげ新たに薬効を見つけ出し、本来それが開発された疾患とは別の疾患に対する治療薬として再開する創薬研究が注目を浴びている。本研究では、薬物に関するケミカルなデータと疾患に関するフェノタイプのデータから潜在的な薬物と疾患の関係を網羅的に予測する機械学習の手法を開発した。教師付き学習のアルゴリズムにスパース統計モデルを導入することで特徴抽出を可能にし、予測された関係に関与している薬物の標的タンパク質と疾患のフェノタイプの組み合わせを同定する点が独自の点である。また抽出された標的タンパク質の生物学的妥当性を Pathway Enrichment 解析を用いて評価した。本発表ではこれらの結果と、既知のデータから学習して生成したモデルを適用することで得られる新規の薬と疾患の関連解析の結果について紹介する。

## 1. 背景

新薬の開発は、病気の症状や原因についての情報を集め、必要とされる薬について調査・分析をすることから始まる。しかしながら、従来の手法では、新しい薬を完成させるためには膨大な資金と膨大な時間が必要であり、認可される新薬が減少している。このような新薬開発の行き詰まりを打開する方法として、既存薬の新しい薬効を発見し、その既存薬を別の疾患治療薬として開発する研究、ドラッグリポジショニングが注目されている[1]。既存薬はすでに人での安全性と体内動態が確認されているため、承認されるまでの期間が短く、また低コストで新薬の開発が可能となる。本研究では、薬物に関するケミカルなデータと疾患に関するフェノタイプのデータを融合し、潜在的な薬物と疾患の関連ネットワークを網羅的に予測する機械学習の手法を開発した。

## 2. データセット

DrugBank データベース[2]などの公共データベースから様々なデータセットを構築した。薬と疾患の関連データは、367の薬と131の疾患の間に728の既知の関連が得られた。薬と標的タンパク質の相互作用データは、367の薬と559のタンパク質から1369のペアの相互作用が得られた。疾患とフェノタイプの関連データは、131の疾患と974のフェノタイプから1695の関連が得られた。

## 3. 手法

薬と疾患の各ペアを薬の標的タンパク質と疾患のフェノタイプ情報を用いて高次元の特徴ベクトルで表現した。その特徴ベクトルのセットに対して、機械学習の分類手法を適用した。機械学習の手法はロジスティック回帰モデルとサポートベクターマシンの2つを検討した。また教師付き学習のアルゴリズムにスパース統計モデルを導入することで特徴抽出を行う。パラメータの最適化学習の過程において、オーバーフィッティングを回避するため L1 罰則項を

負荷した。罰則項の種類としては L1 と L2 がよく利用されるが、L1 はモデルのスパース性が高く、L2 はスパース性が低いことが知られている。薬物・疾患の未知の関係の予測だけでなく、薬物・疾患間ネットワークに関与している薬物の標的タンパク質と疾患のフェノタイプの組み合わせを同定することが狙いである。

## 4. 結果

提案手法を既存の薬物に関するケミカルなデータと疾患に関するフェノタイプのデータに適用し、潜在的な薬物と疾患の関連ネットワークの予測を行った。クロスバリデーション実験を行うことで、提案手法の予測精度を評価した。最適化された特徴の重みでゼロ以上の数が L1 罰則項を適用した手法の方が L2 罰則項を適用した手法よりも少なくスパース性な結果が得られた。また、その特徴を用いて Pathway Enrichment 解析を行うと、あるフェノタイプに特異的に関連性のあるパスウェイが見つかった。

同定したパラメータを用いて薬物と疾患の未知の関連について網羅的予測を行った。その結果、データベースのアノテーション情報によるとリーズナブル結果が得られた。既存のデータベースでは対応付けられていなかった薬物と疾患の未知の関係が網羅的に得られた。例えば、アルツハイマー病とアノテーションされているある薬物が既存のデータベースとは異なる認知症に対応付けられた。また新たに対応付けられたものの中には、標的タンパク質とフェノタイプの情報を元に妥当な結果であると結論付けることが可能なものもあった。

結果の詳細は、本研究会発表にて行う。

## 参考文献

- 1) Chong, C.R. and Sullivan, D.J.: New uses for old drugs. *Nature*, 448 (7154), 645-646 (2007).
- 2) Knox, C., et al.: DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic acids research*, 39(suppl 1), D1035-D1041 (2011).

<sup>†1</sup> 九州大学生体防御医学研究所  
Kyushu University, Medical Institute of Bioregulation