

琉球大学研究紹介:

機械学習とエンリッチメント解析のハイブリッドアプローチによる疾患関連遺伝子の推定

宮田 龍太 (みやた りょうた) / 琉球大学工学部エネルギー環境工学コース・助教

〒903-0213 沖縄県中頭郡西原町千原1番地 TEL:098-895-8620 E-mail: miyata26@tec.u-ryukyu.ac.jp

キーワード：バイオインフォマティクス，遺伝子選択，機械学習，エンリッチメント解析

本研究成果は、2019/09/09~11 に東京工業大学で開催された「日本バイオインフォマティクス 2019 年年会 第8回生命医薬情報学連合大会」にて、筆頭著者である修士1年の藤澤孝太くんにより発表され、後日、大会から優秀口頭発表賞が贈呈されました (<https://www.jsbi.org/news/topics/20191010-nenkai/>)。藤澤くんの受賞は新聞記事としても取り上げて頂きました (<https://ryukyushimpo.jp/news/entry-994660.html>)。

1. はじめに

本研究では、患者と非患者で発現量に有意な差が見られる遺伝子をデータドリブンで特定する機械学習手法である「主成分分析を用いた教師なし学習による変数選択法 (principal components analysis-based unsupervised feature extraction, PCAUFE. 詳細は Taguchi 2019 *Springer* を参照)」に生物学的知見からの制約を課すことで、疾患関連遺伝子をどのくらい限定でき、かつどんな種類が選択されるか検証しました (図1)。

2. 方法と結果

具体的には、PCAUFE で選択されたプローブリストにエンリッチメント解析をかけ誤発見率 (false discovery rate, FDR) のカットオフ水準を設けることで候補を絞りました。その結果、最終的に元の 1/3 程度に候補を限定でき、かつこの操作により、複数の患者に共通して変化が見られる遺伝子と、発現量の個人差が大きい遺伝子に分類できることが明らかになりました (図2)。

3. おわりに

今後はさまざまな種類の病気に関するデータに提案手法を適用し、有用性を検証していく予定です。

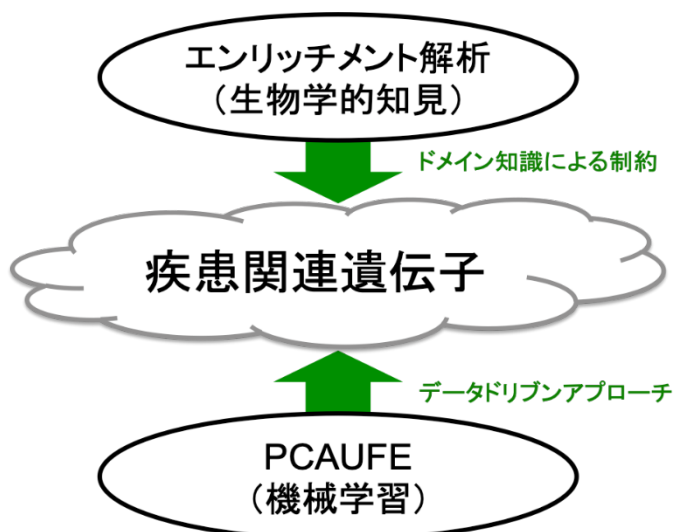


図1: 本研究で提案するアプローチの概念.

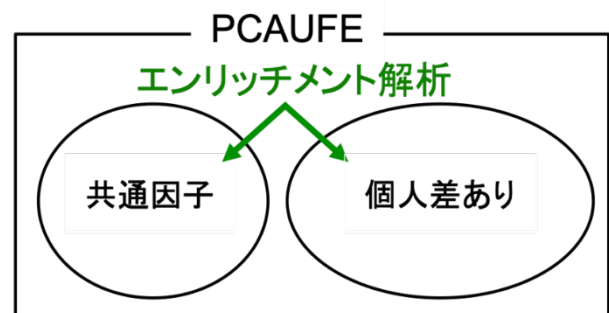


図2: 提案手法で選択された疾患関連遺伝子の解釈.