

2 値分類境界からの距離に基づいた細胞内重要酵素抽出法に関する研究

柿崎 遼 指導教員：作村 諭一

1 はじめに

人間の運動を制御するための情報伝達は脊髄に沿った神経軸索により行われている。そのため、脊髄への強い外力により脊髄の軸索が損傷を受けてしまうと、情報伝達機能が失われ、運動機能や感覚機能に障害を及ぼす。脊髄を含む中枢神経の軸索は、一度傷ついてしまうと再生は困難とされており、未だ有効な治療法は確立されていない。そのため、脳からの信号を四肢へ伝達する人工膜を脊髄に装着する技術の開発 [1] や軸索伸長を阻害するたんぱく質の同定 [2] などさまざまな研究が行われている。

一方、軸索の伸長・退縮はリン酸化酵素反応経路の最終出力とみなすことが出来る。ゆえに、酵素の応答を薬剤（阻害剤）で操作することにより、軸索の長さを変えることが可能である。そして、どの酵素が軸索伸長に大きな影響を与えているのかを特定することが出来れば、軸索を再生するための薬剤の開発に役立つ。一般的に、酵素のネットワーク要素（説明変数）の方が薬剤（条件）よりも多いため、パラメータの正則化などの条件を追加しなければ、回帰問題として解くことは難しい。本研究では、生化学反応方程式に基づく人工データを作成し、様々な酵素阻害度を持つ薬剤を適用した際の軸索伸長度合いを計測する。そこから、機械学習法であるサポートベクターマシンを用い、薬剤を伸長クラスと退縮クラスの 2 クラスにわけ、分類境界から特徴ベクトル（薬剤）までの距離に着目し、最適な酵素阻害パターンを推定することを考える。

2 生化学反応の連鎖による軸索伸長

2.1 薬剤による軸索長の操作

軸索の伸長・退縮はさまざまな分子が相互に作用することで制御されているが、特にリン酸化酵素が重要な役割を担っており、リン酸化酵素ネットワークの最終出力とみなすことができる。リン酸化酵素には軸索の伸長を促進させるものや、退縮を引き起こすものが存在し、それらの酵素の働きのバランスにより最終的な軸索長が決定される。ゆえに、薬剤を用いて酵素の応答を操作することができる。

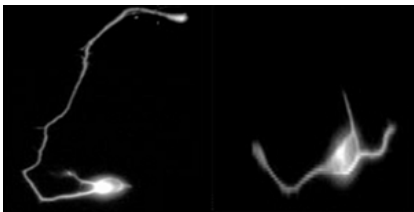


図 1 薬剤で異なる酵素を抑えたときの軸索長の比較 (Malami 大学 Vance Lemmon 研究室より)

2.2 人工データの作成

本研究の手法を実験データに適用する前に、手法の有効性のある程度事前に評価しなければならない。

そこで、これらと神経伸長を数理モデルとして人工データを作成する。生化学反応方程式を用いることでリン酸化酵素間の

反応を表現すると

$$\frac{da_i}{dt} = P_i - D_i \quad (1)$$

となる。ここで、 a_i は i 番目のリン酸化酵素がリン酸化された割合を表しており、 $0 \leq a_i \leq 1$ である。 P_i と D_i はそれぞれリン酸化速度と脱リン酸化速度を表している。Michaelis-Menten 式による近似を用いるとリン酸化速度は、

$$P_i = \sum_{j \in pho} \frac{k_{ji} a_j (1 - a_i)}{K_{ji} + (1 - a_i)} + \sum_{j \in dep} \frac{k_{ji} (1 - a_j) (1 - a_i)}{K_{ji} + (1 - a_i)} \quad (2)$$

と表現される。リン酸化状態による項と脱リン酸化による項からなっている。 k_{ji}, K_{ji} は速度定数である。一般的に、リン酸化酵素に比べ、脱リン酸化酵素は反応対称の選択性が小さく、加水分解による自己脱リン酸化反応も進む。そのため、脱リン酸化速度 D_i は他のリン酸化酵素の影響を受けないと仮定し、

$$D_i = \frac{k_i^{dep} a_i}{K_i^{dep} + a_i} \quad (3)$$

とした。 k_i^{dep}, K_i^{dep} は速度定数である。

軸索長 L の時間変化は、伸長と退縮を施す項からなっており、

$$\frac{dL}{dt} = v_{L+} - v_{L-} \quad (4)$$

と表現した。 v_{L+} は軸索を伸長させる速度、 v_{L-} は軸索を退縮させる速度であり、それぞれ、

$$v_{L+} = \sum_{j \in pho} k_{jL} a_j + \sum_{j \in dep} k_{jL} (1 - a_j) \quad (5)$$

$$v_{L-} = \left\{ \sum_{m \in pho} k_{mL} a_m + \sum_{m \in dep} k_{mL} (1 - a_m) \right\} L \quad (6)$$

とモデル化した。 k_{jL} と k_{mL} は速度定数である。軸索伸長には物理的な細胞骨格などに制限があるため、その影響を v_{L-} が L に依存する形で表現している。酵素反応経路、パラメータ、初期値は乱数で決定した。

3 解析手法

既存の薬剤を人工データの酵素ネットワーク適用し、その伸長度から薬剤に「伸長」ラベルと「退縮」ラベルを付ける。そして、薬剤をサポートベクターマシン (SVM) によって学習させ、「伸長」クラスと「退縮」クラスの 2 クラスに分類する。そこから、分類境界から特徴ベクトル（薬剤）までの距離に着目し、より軸索が伸長する薬剤を推定する。

3.1 サポートベクターマシン (Support Vector Machine, SVM)

SVM はパターン分類手法の 1 つで、データ群を 2 クラスに分類する超平面を決定する学習器である。2 次元データであれば直線によって分類することができ (図 2)、3 次元データであれば平面によって分類される。データがベクトル \mathbf{x} で表現されているとき、次の符号関数によって分離することができる。

$$y = \text{sign}(\mathbf{w}^T \mathbf{x} + b) \quad (7)$$

(7) 式の w, b はパラメータであり, 関数 $\text{sign}(g)$ は $g < 0$ のとき -1 をとり, $g > 0$ のとき 1 をとる. また, 図 2 の $w^T x + b = 0$ となる直線を分離超平面と呼び, 分離超平面に最も近い各クラスのデータ (サポートベクトル) と分離超平面との距離をマージンと呼ぶ. マージンは $\frac{1}{|w|}$ で表され, パラメータ w, b はマージンの最大化によって決定される.

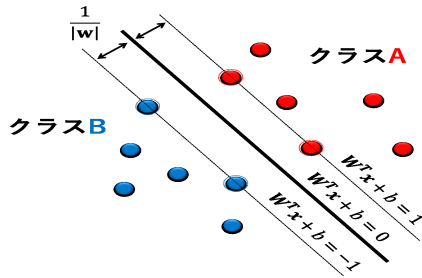


図 2 サポートベクターマシンの概念図

3.2 学習に用いる薬剤データ

SVM の学習に用いるデータの例を図 3 に示す. 行は薬剤の番号を表しており, 列は酵素の番号を表している. 酵素の阻害度は百分率で表現されており, 100% は無阻害, 0% は阻害度最大を表す. また, 薬剤を適用した際の伸長度は百分率で表現しており, 100% を超えた場合は「伸長」ラベル, 100% を下回った場合には「退縮」ラベルを付ける. 酵素数は 200, 薬剤数は 200 (伸長する薬剤 100, 退縮する薬剤 100) として学習を行った.

薬の index	伸長度 (%)	酵素の index							
		1	2	3	4	5	6	7	8 ...
1	279.9	99	97	98	92	95	100	100	96
2	207.9	99	97	98	92	95	100	100	96
3	207.3	99	99	99	97	98	100	100	97
4	173.1	100	100	100	100	98	100	100	100
5	130.5	98	97	98	94	99	93	82	33
6	129.5	100	100	99	100	100	100	99	100
7	126.1	100	100	100	99	93	100	100	99
8	116.9	99	98	98	96	95	90	99	98
9	100.3	98	98	99	98	91	100	99	95
10	97.8	100	100	100	90	95	100	100	98
11	97.1	100	99	99	96	102	98	94	90
12	86.7	100	99	99	96	97	100	99	97

図 3 学習に用いる薬剤データ

3.3 分類境界からの距離と薬剤効果の関係

分類境界から薬剤までの距離に着目する. SVM によって薬剤を「伸長」クラスと「退縮」クラスに分類したとき, 分類境界から距離が遠い薬剤の阻害パターンに, 軸索伸長に関する重要な情報が含まれているのではないかと考えた (図 4).

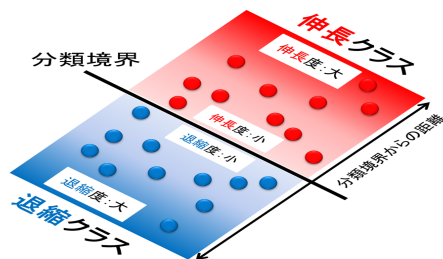


図 4 分類境界からの距離と薬剤効果の関係

そこで, 本研究では分類境界からの距離が伸長方向に最も遠くなるような仮定の薬剤をランダム探索によって推定することを考える. 図 5 は, ランダム探索の概念図である. 左図は, 一般的な直角座標系上の酵素阻害パターン, 右図は, 非線形変換によって線形分離できる空間上の酵素阻害パターンである.

初めに, 「伸長」クラスの薬剤を 1 つ選択する. その薬剤の直角座標系における酵素阻害パターンに微小乱数を付加し, 変換座標系において分類境界からの距離を評価する. 乱数付加前よりも距離が遠くなれば, 新規の薬剤とする. このサイクルを距離の変化がなくなるまで繰り返すことにより, 分類境界からの距離が最も遠くなる仮想薬剤を推定する.

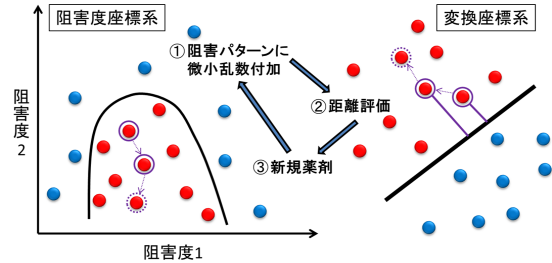


図 5 ランダム探索の概念図

4 推定結果

ランダム探索において, 距離が遠い仮想薬剤が見つかるたびに, 人工データの酵素ネットワークに適用し, 軸索伸長度合いを計測した. 図 6 は, 分類平面からの距離と軸索伸長度合いとの関係を表しており, 横軸は分類境界から適用した仮想薬剤までの距離, 縦軸は軸索伸長度合い, 100% の線は薬剤適用前の軸索長を表す. 図 6 より, 分類境界からの距離が遠い薬剤のほうが軸索が伸長していることが確認できる.

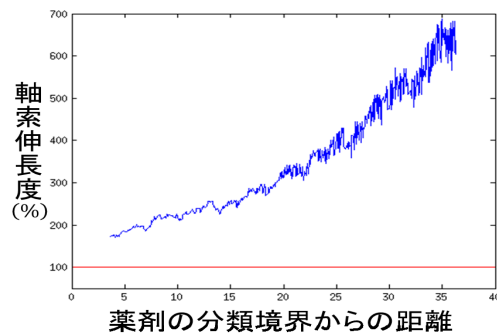


図 6 分類境界からの距離と軸索伸長度合い

5 まとめ

分類境界からの距離が遠い薬剤ほど軸索が伸長するという結果になり, 反応経路が不明であっても酵素阻害パターン (入力) と軸索伸長度合い (出力) から, SVM によって作成された分類境界からの距離に基づいて, より軸索が伸長する酵素阻害パターン (仮想薬剤) の推定が可能であることが示唆される.

参考文献

- [1] I. R. Mineev, P. Musienko, A. Hirsch, Q. Barraud, N. Wenger, E. M. Moraud, et al., "Biomaterials. Electronic dura mater for long-term multimodal neural interfaces," Science, vol. 347, pp. 159-63, 2015.
- [2] J. Nagai, Y. Kitamura, K. Owada, N. Yamashita, K. Takei, Y. Goshima, et al., "Crmp4 deletion promotes recovery from spinal cord injury by neuroprotection and limited scar formation," Sci Rep, vol. 5, p. 8269, 2015.