

Soft Confidence-Weighted Learning を用いたミカエリス定数の推定

近藤 翔太[†] 松田 健[†] 大梶 弘順[†]

[†]静岡理工科大学総合情報学部

1. はじめに

ミカエリスメンテンの式は酵素 E の反応速度 v と基質 S の濃度 $[S]$ との関係をモデル化したものであり以下の式で定義される:

$$\frac{1}{v} = \frac{K_m}{V_{max}} \frac{1}{[S]} + \frac{1}{V_{max}}.$$

ここで, V_{max} は最大反応速度を表す定数であり, K_m はミカエリス定数と呼ばれる実数である. ミカエリス定数を求めることで酵素と基質の親和性を調べることができたため, ミカエリスメンテンの式は薬の代謝速度の予測[3]や、薬物の副作用に関する予測[2]に広く利用されている.

ミカエリスメンテンの式は直線の方程式で表されているが, 実験によって得られるデータ $\frac{1}{v}$, $\frac{1}{[S]}$ の値は一般的に直線上には並ばないため, K_m の値には最小二乗法がよく用いられている.

本研究では, K_m の値を Soft Confidence-Weighted Learning と呼ばれる機械学習のアルゴリズムを応用して求める方法を提案し, 最小二乗法を用いて K_m の推定をしたときの違いを比較することで提案手法の有効性について検討した.

2. Soft Confidence-Weighted Learning

SCW[1]は線形な超平面を表す式 $\sum w_i x_i$ の正負によってデータを2つのクラスに分類するための機械学習のアルゴリズムである.

ここで x_i は与えられたデータに対応する実数であり, w_i は実数値をとるパラメータである.

SCW ではパラメータ w_i は平均が μ_i , 分散 σ_i である正規分布に従うものと考え, パラメータ w_i を決定する正規分布の平均と分散は以下のルールで更新される:

$$\mu_{t+1} = \mu_t + \alpha_t \cdot y_t \cdot \sigma_t \cdot x_t$$

$$\sigma_{t+1} = \sigma_t - \beta_t \cdot \sigma_t \cdot x_t^T \cdot x_t \cdot \sigma_t$$

Estimation of Michaelis constant with Soft Confidence-Weighted Learning

[†]Shota Kondo [†]Takeshi Matsuda [†]Kojune Ohsugi
Shizuoka Institute of Science and Technology[†]

$$\alpha_t = \min \left\{ C, \max \left\{ 0, \frac{1}{v_t \zeta} \left(-m_t \psi + \sqrt{m_t^2 \frac{\phi^4}{4} + v_t \phi^2 \zeta} \right) \right\} \right\}$$

$$\beta_t = \frac{\alpha_t \phi}{\sqrt{u_t} + v_t \alpha_t \phi}$$

$$u_t = \frac{1}{4} \left(-\alpha_t v_t \phi + \sqrt{(\alpha_t)^2 (v_t)^2 \phi^2 + 4 v_t} \right)^2$$

$$v_t = x_t^T \sigma_t x_t, m_t = y_t (\mu_t, x_t)$$

$$\phi = \Phi^{-1}(\eta), \psi = 1 + \frac{\phi^2}{2}, \zeta = 1 + \phi^2$$

SCW はオンラインアルゴリズムであるため, パラメータの学習は1つずつのデータを用いて行われる.

3. 提案手法

この章では, SCW を用いてミカエリス定数を推定する方法を提案する. SCW では原点を通る直線を用いて学習を行うため, 与えられたデータの平均座標を原点とする空間でパラメータの学習を行う. 以下がアルゴリズムの概要である.

- ① n 個のデータ $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ が与えられたとする.
- ② データの平均座標 (x_r, y_r) を求め, 平均座標が原点になるようにデータを平行移動する. 平行移動後のデータを $\{(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)\}$ とおく.
- ③ 与えられたデータの共分散行列の最大固有値に対応する固有ベクトルを求め, その固有ベクトルを傾きとし, 原点を通る直線の方程式 $l: v = a_u$ を考える.
- ④ n 個のデータ $\{(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)\}$ のうち $v > a_u$ を満たすデータには $z = 1$ というラベルを付け, $y < a_u + b$ を満たすデータには $z = -1$ というラベルを付け, SCW を用いて直線の方程式 l を更新する.
- ⑤ 直線 l を元の座標系に平行移動し, ミカエリス定数を推定する直線の方程式 $y = ax + b$

が得られる。

提案アルゴリズムは少し煩雑な計算過程を必要とするように感じられるかもしれないが、直線の傾きを正規分布の平均と分散の値を用いて決定するため、与えられたデータから得られる直線の方程式の傾きの範囲を求めることができる。最小二乗法を用いて推定する場合との相違点となる。

4. 実験と結果

本研究では、表1のデータに前章で提案したアルゴリズムを適用し、ミカエリス定数を推定する。

表1 EDTA非存在下での濃度と速度のパラメータ

S(mM)	0.05	0.1	0.2	0.5	1	2	5	10
1/S	20.0	10.0	5.00	2.00	1.00	0.50	0.20	0.10
V($\mu\text{mol/min}$)	1.02×10^{-4}	5.29×10^{-4}	8.54×10^{-4}	2.03×10^{-3}	3.71×10^{-3}	4.54×10^{-3}	9.53×10^{-3}	1.42×10^{-2}
1/V	1.46×10^4	1.89×10^3	1.17×10^3	4.93×10^2	2.70×10^2	2.20×10^2	1.05×10^2	7.00×10^1

SCWと最小二乗法を比較したのが図1である。

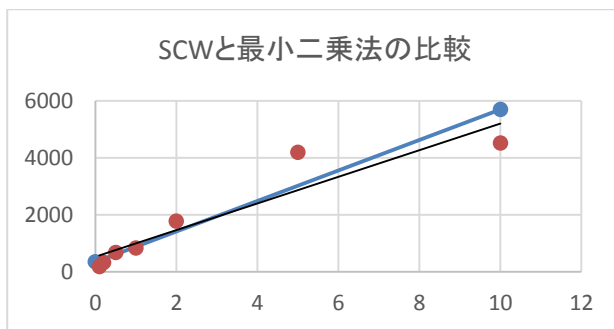


図1 青の直線：SCW、黒の直線：最小二乗法

表1で与えたデータに対して3章で提案したアルゴリズムを用いて K_m 値と V_{max} 値を推定した結果は次の通りである。

$$y = 535.4114x + 339.3235351$$

$$K_m = \frac{535.4114}{339.3235351} = 1.577878763529362 \approx 1.58(\text{mM})$$

$$V_{max} = \frac{1}{339.3235351} = 0.0029470399089922 \approx 2.95 \times 10^{-3} \left(\frac{\mu\text{mol}}{\text{min}} \right)$$

一方、最小二乗法を用いて K_m 値と V_{max} 値を推定した結果は次の通りである。

$$K_m = \frac{185.64}{104} = 1.7850 \approx 1.76(\text{mM})$$

$$V_{max} = \frac{1}{104} = 0.0096154 \approx 9.62 \times 10^{-3} \left(\mu \frac{\text{mol}}{\text{min}} \right)$$

提案手法と最小二乗法を用いた場合の直線の当

てはまりの良さを調べるために、データの各点と直線の距離を測ってその距離の総和をとる事で、どちらの方法の方が直線の当てはまり具合が良いか調べた。結果は

提案手法：5.43

最小二乗法：11.70

となり、今回用いたデータでは提案手法の方が直線の当てはまり具合は良いという結果となった。

5. 考察とまとめ

本研究ではSCWと呼ばれる機械学習の手法を用いてミカエリス定数を推定するためのアルゴリズムを提案した。一般的に、ミカエリス定数の推定には最小二乗法が利用されるが、提案手法と最小二乗法を用いてミカエリス定数を求めるための直線の方程式に対して点と直線の距離をそれぞれ計算した結果、提案手法の方が最小二乗法と比較してよく直線に当てはまっていることを確かめることができた。提案手法はSCWを利用しているため、SCWの性質から、はずれ値を含むデータが与えられたときも適切にパラメータ推定を行うことができる可能性をもつと考えられる。この点について検証することが今後の課題である。

謝辞

本論文を作成するにあたり、静岡理工科大学理工学部物質生命科学科の遠藤氏、大坂氏、大瀧氏から実験データを賜り、松田研究室ゼミ生の大沢氏にSCWの指導を賜りました。ここに感謝の意を表します。

参考文献

- [1] Jialei Wang et al, "Exact Soft Confidence-Weighted Learning", International Conference on Machine Learning, pp. 121-128, 2012
- [2] 白石 文秀 (2012) 「代謝反応ネットワーク解析の意義と方法」『化学と生物』第59巻第2号, 103-110 項.
- [3] 山舘 周恒 (1997) 「酵素反応速度パラメータ算出法の信頼性に関する研究」『臨床化学』第26号, 31-37 項.