

点過程データの解析における最適化手法について

鮎川 矩義

東京理科大学 工学部 情報工学科

概要: 点過程データとは、あるイベントの発生時刻を時系列順に並べたものである。マーク付き点過程データとは点過程データの各イベントにマークと呼ばれる高次元ベクトルが付与されたものをいう。点過程データやマーク付き点過程データとしてモデル化される現実の事象としては、地震活動、株価変動、SNS サービスのつぶやき、脳のニューロンの活動、などが挙げられる。

点過程データやマーク付き点過程データの解析では、データ間の「距離」が頻繁に利用され、基本的かつ重要な役割を果たしている。よく知られた距離として、Victor and Purpura (1997) によって提唱された、VP 距離がある。VP 距離は2つの点過程データ間の距離を、一方をもう一方に移し変えるのに必要となる費用の最小値として定義する。移し変える際に使える操作は、イベントの「移動」・「挿入」・「削除」の3つであり、各操作には費用が定義される。そのため、VP 距離は、編集距離とも呼ばれる。Suzuki, Hirata, Aihara (1997) は、VP 距離のマーク付き点過程データへの拡張を提案し、その有用性を示している。

VP 距離や拡張 VP 距離の計算は、本質的には、組合せ最適化問題である。しかしながら、最適化手法に基づいた計算方法については十分検討されていないのが現状である。本発表では、おもに、つぎの2つの成果について報告する：

- (1) VP 距離と拡張 VP 距離の計算にハンガリー法を適用する方法を提案する。数値実験を行ない、提案手法の有効性を検証する。
- (2) VP 距離に基づいたメディアン問題を導入する。これは、与えられた複数の点過程データに対し、VP 距離の総和を最小にする点過程データを求める問題である。このメディアン問題に対し、整数最適化に基づいた厳密解法を提案する。数値実験を行ない、提案手法の有効性を検証する。

(1) は、平田祥人氏（筑波大学）との共同研究に基づく。(2) は、鈴木祥平氏、志賀大祐氏、池辺淑子氏（東京理科大学）との共同研究に基づく。