

特別研究報告書

時系列間距離とその検索数推定への応用

指導教員 山下信雄 教授

京都大学工学部情報学科
数理工学コース
平成30年4月入学

長塩 幸大

令和4年1月28日提出

摘要

株価や音声など、時刻によって変化する情報を持つデータのことを時系列データと呼ぶ。時系列データの活用方法として、データの分類、クラスタリング、予測などが挙げられる。そのような活用方法においてデータ間の距離が重要な役割を果たしている。これまでに時系列間距離として、ユークリッド距離と動的時間伸縮法に基づく距離が主に使われてきた。しかしながら、これらの時系列間距離には問題点が存在する。ユークリッド距離はデータの長さが異なる場合に定義できず、また短期的に時間軸上のずれを生じるデータに対しては非常に脆い。一方、動的時間伸縮法に基づく距離は計算量が大きく、また距離の公理を満たさず、カーネル法などの機械学習の手法に応用できない場合がある。

本報告書では、既存の時系列間距離の欠点を解消するため、機械学習などの分野で確率分布間の距離を表すのによく用いられている Wasserstein 距離に着目する。Wasserstein 距離は最適輸送問題の解として得られ、距離の公理を満たすことが知られている。また、Wasserstein 距離の計算は、通常であれば膨大な時間がかかるが、データを次元の確率分布とみなせる場合は高速に求められることが知られている。一方で時系列間距離に Wasserstein 距離をそのまま適用する場合、時系列データの確率分布への変換によってデータの規模についての情報が失われ、また距離の公理の非退化性を満たさない。

そこで、本報告書では、新たな時系列間距離として Wasserstein 距離に時系列データの要素の総和の項を追加したものを提案する。さらに、提案した距離が距離の公理を満たすことを証明し、数値実験によって時系列データの予測において提案した距離と既存の距離の性質について考察する。