

IoT が拓く未来
2019 年度採択研究者

2020 年度 年次報告書

天方 大地

大阪大学 大学院情報科学研究科
助教

超高速 IoT ビッグデータ解析のための分散アルゴリズム基盤

§ 1. 研究成果の概要

2020年度では主に、(1)メトリック空間における距離に基づくアウトライア検出問題、(2)高次元空間における高速・高精度な最近傍検索問題、および(3)メトリック空間における動的データに対する密度ベースクラスタリング、に取り組んだ。

(1)は、 k 最近傍までの距離が最も遠い n 個のデータを検索するものである。この研究では、一度の計算で大量のアウトライアでないデータをフィルタする技術を開発し、単純な並列化も可能とした。実データを使用した実験から、既存技術よりも 10 倍以上高速であることを確認した。

(2)は、機械学習技術の普及による高次元データの増加とその高速検索の需要の高まりによって、より重要化されている問題である。我々は、局所鋭敏型ハッシュ関数 (locality-sensitive hashing) と近接グラフ (データをノードとし、距離の小さいデータにエッジを加えたもの) の2つのデータ構造を利用することによって、既存の技術よりも高速・高精度なアルゴリズムを開発した。このアルゴリズムは、ハッシュ関数を用いて、解との距離が小さいデータを定数時間で検索し、それを近接グラフにおける探索の初期位置として設定する。その後、初期位置から解に近づきデータを貪欲に辿る。これは、既存の近接グラフアルゴリズムにはなかった、理論的な精度のコントロールを可能とした初のアルゴリズムである。

(3)では、IoT 環境で想定されるストリームデータの高速クラスタリング技術を開発した。クラスタリングアルゴリズムとして、Density-Peaks Clustering (DPC) を採用し、メトリック空間 + 動的データを想定した際の DPC の高速性と正確性に関する理論を導出した。また、高速に高精度なクラスタリング結果を導くアルゴリズムを開発し、実データを用いた実験から、既存技術よりも高速・高精度であることを確認した。

【体系的な原著論文情報】

- 1) “Fast and Exact Outlier Detection in Metric Spaces: A Proximity Graph-based Approach”, ACM SIGMOD Conference on Management of Data (SIGMOD), Jan. 2021(Accepted).
- 2) “高次元空間における効率的な近似 k Nearest Neighbor 検索アルゴリズム”, 第 13 回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2021), 2021 年 3 月.