IoT が拓く未来
2019 年度採択研究者

2019 年度
実績報告書

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

2019 年度では，メトリック空間における(1)高次元データに対する範囲検索(指定したクエリオブジェクトからある範囲内に存在する全てのオブジェクトを検索する)問題，および(2)アウトライアー検出問題，の2つの問題に取り組んだ．IoT データは何らかの特徴やパターンを持つことが一般的であり，範囲検索はそれらを発見する(データマイニング・分析を行う)上で必須の技術である．また，IoT データは大量のノイズや外れ値を含むことも一般的であり，アウトライアー検出は，それらを効果的に発見・除去する(データクリーニングを行う)ための技術である．これらの検索・検出を大量の IoT データに対して高速に実行するアルゴリズムとデータ構造を開発した．

まず(1)に関して，高次元データに対する「正確な」範囲検索の高速化は難しく，全てのデータにアクセスする方法(シーケンシャルスキャン)が最速であるとされている．そのため，近似解を求めるアルゴリズムの開発が盛んに行われており，本研究も近似解を出力するアルゴリズムを開発した．提案アルゴリズムは，距離の短いオブジェクト間にリンクを持たせた近接グラフと呼ばれるデータ構造を事前に作成する．クエリが指定されたら，その近接グラフのあるノードからクエリに近づいた後，範囲内のデータを貪欲に辿る貪欲法により，範囲検索を実行する(図1)．提案アルゴリズムは，シーケンシャルスキャンや木構造に基づくインデックスを用いたアルゴリズムよりも 100 倍程度高速であり，近似解を出力する代表的なアルゴリズムである LSH よりも 50 倍程度高速かつより高い精度である．

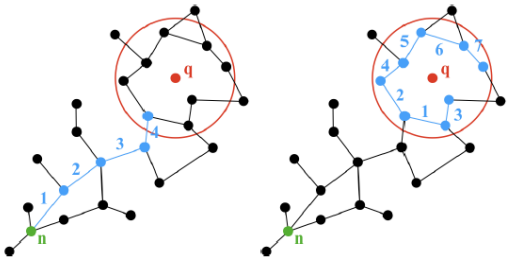

図 1：近接グラフを用いた範囲検索

(2)は，ある範囲内に一定数のデータがないデータをアウトライアーと定義し，全てのアウトライアーを検索する問題である．本問題は，範囲検索がベースとなっているため，(1)で開発した技術を応用することにより，高速にアウトライアーを発見できることを確認している．