

IoT が拓く未来  
2019 年度採択研究者

|                  |
|------------------|
| 2021 年度<br>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

2021 年度では、(1)ユークリッド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、(2)逆内積最大探索、および(3)内積空間におけるカーディナリティ推定に取り組んだ。

(1)は 2019 年度の成果である密度ベースクラスタリングの動的データへの拡張である。この研究では、データセットが新たなデータの追加および任意のデータの削除を許可する場合において、クラスタリング結果を厳密に更新するアルゴリズムを設計している。このアルゴリズムの理論的性質を導入し、実データを用いた実験においても理論的性質を示している。また、既存の厳密解に対して 100 倍以上高速であり、既存の近似解アルゴリズムよりも高速であることを示した。

(2)は、サブシステム等で集められたデータから行列分解モデルを用いて推定されたユーザの嗜好情報が与えられた際、どのユーザがクエリとなるアイテムを好むかを計算する問題であり、これまでに取り組まれていなかったものである。本問題の最も簡単な解法は、全てのユーザに対して最大内積探索アルゴリズムを適用するものである。しかし、この方法では解に含まれないユーザに対しても最大内積探索を行ってしまい、非常に非効率的である。そこで本研究では、解に含まれないユーザを効率的にフィルタリングするアルゴリズムを提案した。実データを用いた実験から、ベースラインアルゴリズムよりも最低で 100 倍以上高速であることを示した。本研究内容は、推薦システム分野のトップ会議である RecSys に採択されている<sup>1)</sup>。

(3)は、あるクエリおよび閾値が与えられた際、クエリとの内積が閾値以上となるベクトルの数を推定する問題である。本問題も(2)と同様に、クエリがどれだけのユーザに好まれるものであるかを調べるツールとして利用可能である。本問題の既存研究は誤差が非常に大きく非実践的であった。そこで本研究では、少ないサンプル数で非常に小さい誤差を与える推定アルゴリズムを提案し、実データを用いてその有効性を検証した。本研究の内容は、DEIM2022 で発表しており<sup>2)</sup>、スポンサー賞を受賞している。

### 【代表的な原著論文情報】

- 1) Daichi Amagata and Takahiro Hara, “Reverse maximum inner product search: How to efficiently find users who would like to buy my item?”, Proc. RecSys, pp.273-281, 2021.
- 2) 平田 皓平, 天方 大地, 原 隆浩, “高速かつ高精度な内積空間におけるカーディナリティ推定”, DEIM フォーラム 2022, 2022.