

空間データモデリングによるニューロンデータ検索の高速化

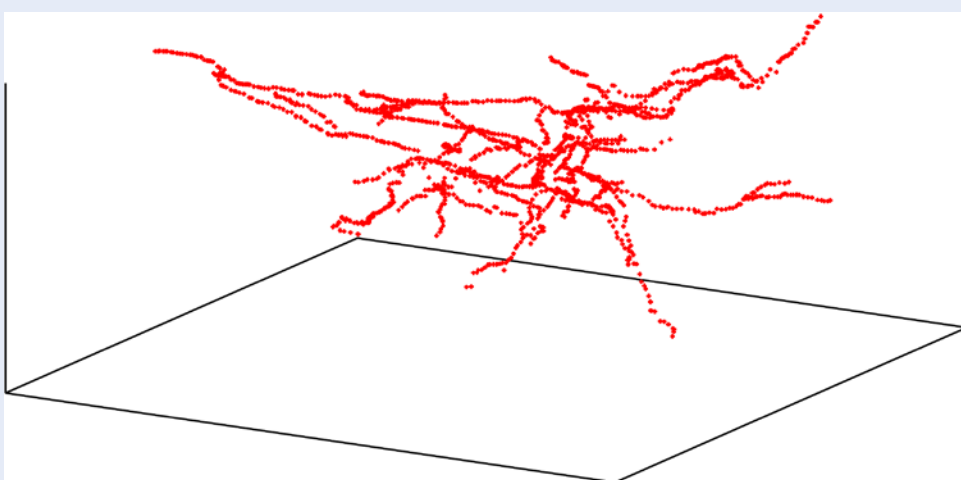
～インタラクティブなデータを高速検索～

大阪大学大学院情報科学研究科
天方 大地

1 研究動機

神経ネットワークやニューロンの機能の理解をサポート

- 計算機による高速（シミュレーション）分析
- 計算機が取り扱いやすい空間データに近似



← 1つのデータ = 大量の点
+ ネットワークには大量のデータ

3次元点によるニューロンの描画

神経ネットワークには、ハブとなるニューロンが存在

- ニューロンの作用 → 空間的近接性が関係
- 信号伝達メカニズムの解析に不可欠なデータ → 検索対象

3 本研究の効果と未来展望

「データ = 点集合」とした際の分析ツールへの注目

- これまでは「データ = 1つの点」といったモデルが主流であったが、機械学習・コンピュータビジョンの普及で、1つのオブジェクト（データ）が点集合で表されるようになり、そういったモデルで分析する実践性と重要性を提示
- 分析を高速に行うためには、新たなデータ構造が必要であることを提示
- 分析ツールは要件によって異なるため、本研究を皮切りに新たな分析ツール（オペレータ）、およびそのオペレータを高速に処理するアルゴリズム・理論が提案される可能性

5 提案データ構造：BIGrid

2つのグリッド（small grid & large grid）から構成

- Small grid：1辺の長さが $\frac{r}{\sqrt{3}}$ のグリッド
 - セル中にマッピングされ得る点のオブジェクトIDを圧縮ビットセットで管理
- Large grid：1辺の長さが $[r]$ のグリッド
 - $I(c_K^l)$ ：セル c_K^l にマッピングされた点をオブジェクトIDごとに管理する転置リスト
 - $\mathbf{b}(c_K^l)$ ：セル中にマッピングされた点のオブジェクトIDを圧縮ビットセットで管理
 - $\mathbf{b}^{adj}(c_K^l)$ ：隣接セル中にマッピングされた点のオブジェクトIDを圧縮ビットセットで管理

定理

- 下界値： $\tau^{low}(o_i) = |\bigvee_{K \in o_i, L} \mathbf{b}(c_K^L)| - 1 \leq \tau(o_i)$
- 上界値： $\tau(o_i) \leq |\bigvee_{K \in o_i, L} \mathbf{b}^{adj}(c_K^L)| - 1$

ニューロンデータは密であるため、ビットセットの圧縮率が高い
→ 下界値と上界値の計算（bitwise OR）を高速に実行可能

6 アルゴリズム & 評価実験

アルゴリズムの概要

1. インデックスをオンラインで構築
2. 各オブジェクト o の下界値 $\tau^{low}(o)$ および $\tau_{\max}^{low} = \max \tau^{low}(o)$ を計算
3. 各オブジェクト o の上界値 $\tau^{upp}(o)$ を計算し、 $\tau^{upp}(o) < \tau_{\max}^{low}$ であれば枝刈り
4. $\tau^{upp}(o) \geq \tau_{\max}^{low}$ である o を上界値の降順にソート
5. ソート順にスコアを計算（ $\tau(o^*) \geq \tau^{upp}(o)$ となった時点で計算終了）

評価実験

距離の閾値 r とオブジェクト数を変えた評価（実行時間）

2 取り組んだ問題

入力：

- オブジェクト集合 O （ $|O| = n$ ）
- 各オブジェクト $o_i \in O$ は、1個以上の3次元点集合 P_i
- 距離の閾値 r

オブジェクトのスコア：

$$\tau(o_i) = |\{o_j \mid o_j \in O - \{o_i\}, p \in P_i, p' \in P_j, \text{dist}(p, p') \leq r\}|$$

- グラフにおける度数に相当

出力：

- $o^* = \arg \max_o \tau(o_i)$
- グラフにおける度数中心性に相当

4 プロジェクトの成果

1. 最もインタラクトしているオブジェクトの定式化

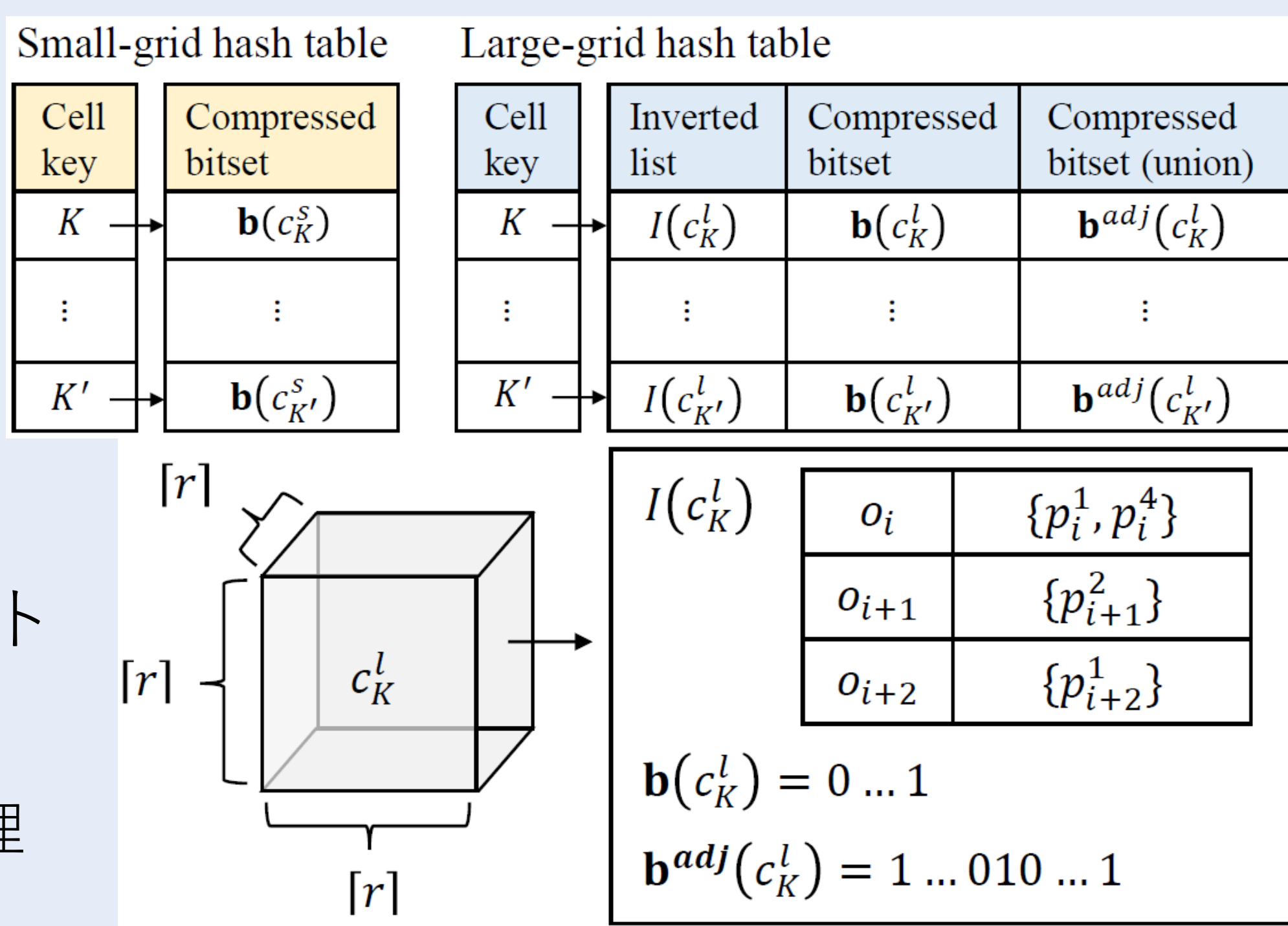
- 離散数学からセマンティックも保証

2. 新たなデータ構造とアルゴリズムの開発

- 素朴なアルゴリズムに対して**何十倍も高速**
- 分散処理にも対応

3. 論文発表（データベース分野のトップ会議に採択）

- IEEE International Conference on Data Engineering 2019
- ニューロン以外にも適用先があることを紹介



あるセル c_K^L の構造

