データを圧縮、高速計算

東京大学　講師　松井勇佑

課題名 圧縮線形代数：
データ圧縮による省メモリ
高速大規模行列演算

東京大学 THE UNIVERSITY OF TOKYO

問題

大量データ → 処理（線形代数） → 応用

データ量が多すぎるので　☹大量メモリ消費。遅い。

- 現状、大規模処理は力技（大量サーバ）
- Google等の大企業しか取り組めない

どのように未来を切り拓くか

大量データ → コード → 処理（圧縮線形代数） → 応用

(1) 圧縮して　(2) 線形代数演算

- データ圧縮した世界での高速演算体系
- 全てのエンジニア/研究者に**便利な道具**を提供

成果（ACT-I + ACT-I加速）

【道具１：実用】圧縮世界での高速処理
- クラスタリング [ACMMM 17]
  - ✓ 2017当時世界最速級・コード公開
- 部分探索 [ACMMM 18]

【道具２：理論】添字計算
- 添え字（青い数字）に関係を持たせる
- 新しい理論的枠組み

関係を理論的に整理

【アウトリーチ】
- CV分野国内最大学会 (MIRU) および世界最大学会でチュートリアル (CVPR, 6月予定)
- その他招待講演（東北大、産総研、DeNA、CyberAgent）

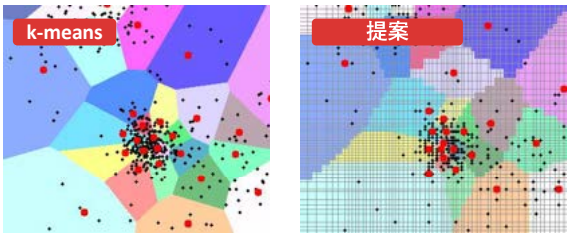
CVPR SEATTLE WASHINGTON JUNE 16-18 2020

技術的詳細の要約

x → $k^* = \underset{k}{\operatorname{argmin}} \|x - c_k\|_2^2$ → $k^* = 32$

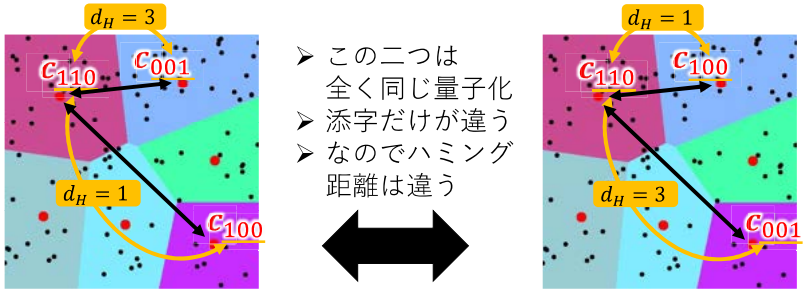
- ベクトル量子化でデータを圧縮
- 「圧縮データ」に対し、**圧縮したまま演算**を行いたい
  - ✓【道具１：実用】圧縮したまま高速探索。高速クラスタリング
  - ✓【道具２：理論】添字に関係を持たせる（これまでにない**操作**）

【道具１：実用】クラスタリング



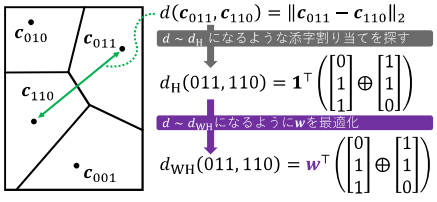
- 圧縮したデータの世界でk-means
  - ✓ データを格子上の点におき、格子上の世界でクラスタリング
- k-meansに比べ，10x から 100x 高速。10x から 100x メモリ効率良
- コード公開。プレスリリース。
- 手法：
  1. 圧縮コードに対する探索 [Matsui+, ICCV15]
  2. 圧縮コードに対する平均計算 [Matsui+, ACMMM 17]

【道具２：理論】添字に意味をもたせる



- この二つは全く同じ量子化
- 添字だけが違う
- なのでハミング距離は違う

- 「添字割り振り」の自由度がある
- この自由度を使って、**新たな関係性**を埋め込む
- すると、これまでにない数学体系が作れる（かも）
- ✓ 例：
  これまでと全く同じベクトル量子化器だが、**添字をじっと眺めるとなぜか別の距離尺度が計算できる**
- ✓ より応用的な例：
  商品の類似画像検索で、圧縮された特徴ベクトルを**じっと眺めるとなぜかカテゴリ検索が出来る**



ユークリッド距離そのものを埋め込む例：
- ベスト割り振りを求める（置換の最適化）
- さらにベストな重みを求める

ユークリッド以外（別の距離尺度）を埋め込めるかもしれない