

よく似たデータを瞬時に発見

～近傍グラフを用いた高速類似探索～

どんな研究？

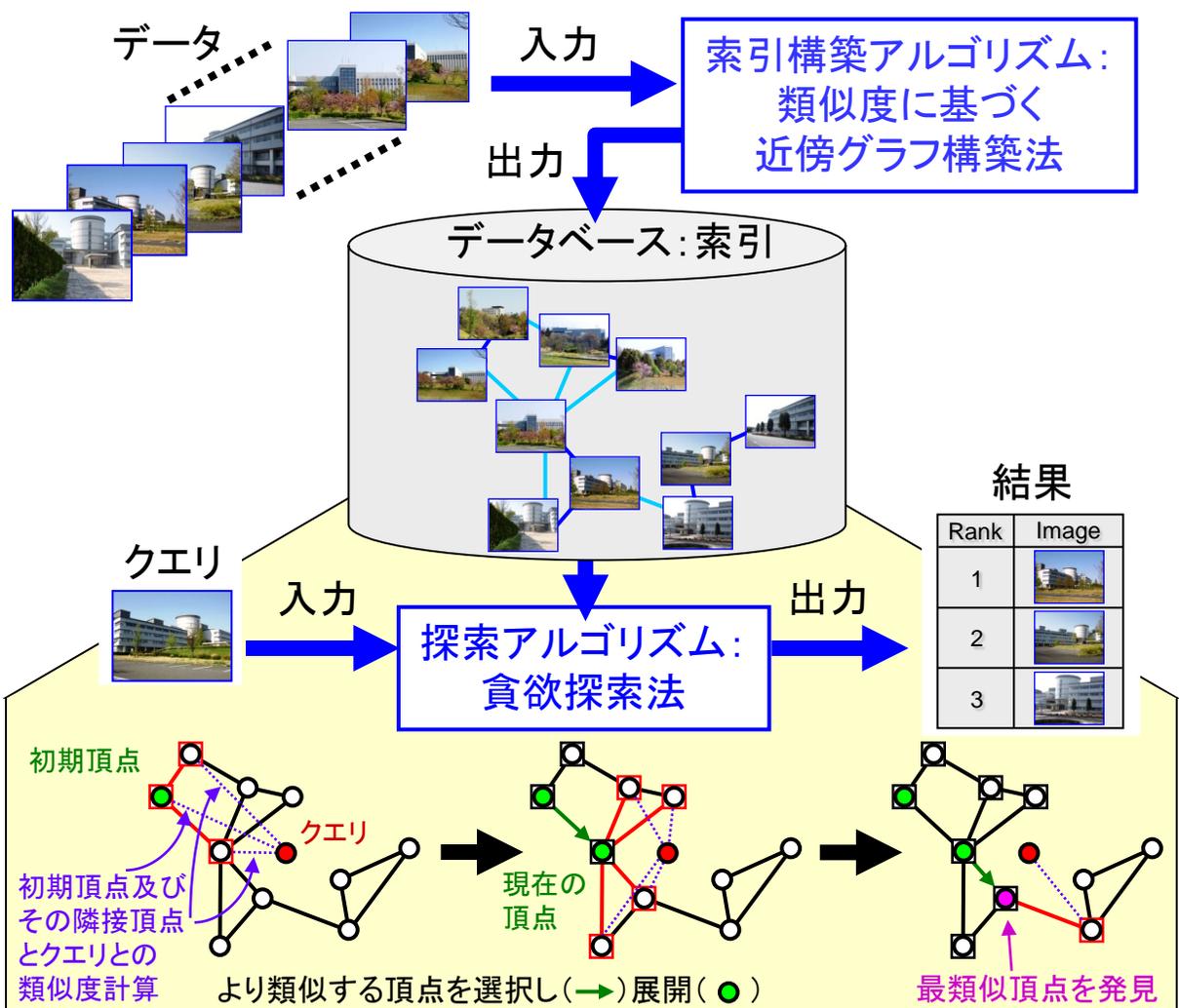
多様で大規模なデータから、似ているデータを瞬時に見つけることは困難な問題です。この展示では、文書、画像、音声のような様々な大規模データから、与えたデータに類似するものを高速に発見する技術を、データの構造化と探索法とに焦点を絞り、紹介します。

どこが凄い？

データから構築した近傍グラフを索引とし、このグラフに貪欲探索法を適用した提案法は、高速、精度保証可能、多様なデータに適用可能、結果の可視化という特徴を有する方法です。100万文書データに、90%の成功率を設定した場合、僅か0.05%の500文書との類似比較で最類似文書を発見します。

どんな風に役立つ？

複数種類のデータ（文書、画像、音声音響信号、記号列など）を組み合わせたマルチメディア・データや使用する方が自分で設定した類似度を利用した探索が可能になります。その他にも、データマイニング、パターン認識など様々な分野への波及効果が期待できます。



関連文献

- [1] K. Aoyama, S. Watanabe, H. Sawada, Y. Minami, N. Ueda, and K. Saito, "Fast similarity search on a large speech data set with neighborhood graph indexing," ICASSP2010
- [2] K. Aoyama, K. Saito, H. Sawada, and N. Ueda, "Fast approximate similarity search based on degree-reduced neighborhood graphs," KDD2011

連絡先

青山一生 (Kazuo Aoyama) 協創情報研究部 知能創発環境研究グループ
E-mail : aoyama.kazuo[at]lab.ntt.co.jp ({at} の部分を @ に置き換えてください)