

異種データ間の隠れた関連性を探り出す ～教師なしオブジェクトマッチング～

どんな研究

異なる種類のデータ集合の間で**関連するオブジェクトを自動的に見つける**研究です。例えば、意味が同じ英語と日本語の単語の対応付け（辞書自動作成）、画像と文の対応付け（説明文作成）、複数のデータベースのIDの対応付け（名寄せ）に利用できます。

どこが凄い

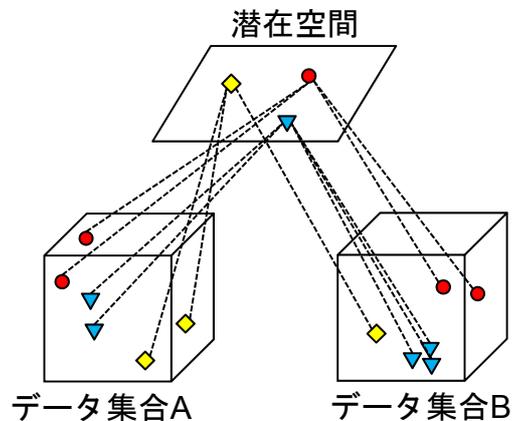
これまでは人手で対応付けたデータや異種データ間の類似尺度が必要でした。提案法は潜在空間に全データを埋め込むことによって、**対応データや類似尺度なしで**関連性を発見できます。また多対多対応がある場合や、2つ以上のデータ集合が与えられた場合でも適用できます。

目指す未来

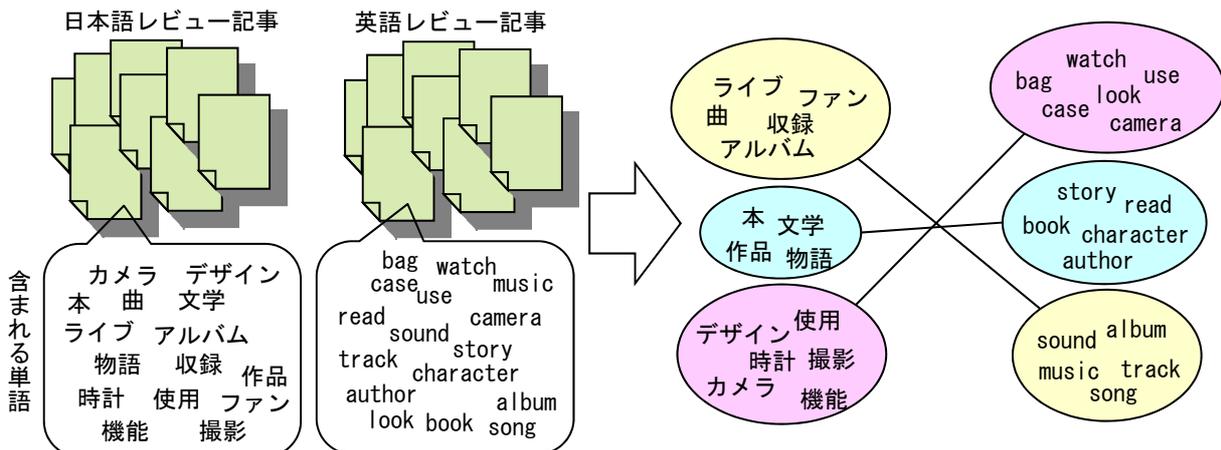
現在、大量のデータ集合が多様な用途で収集・管理されています。そのため異なるデータ集合の間の関係を解析することが困難でした。本技術により**複数のデータ集合の統一的解析**が可能となり、これまで知りえなかった因果関係や知識の発見につながります。

一見対応関係のない複数のデータ集合を、性質を保存したまま共通の空間（潜在空間）に埋め込むことによって、要素間の対応を付けることができます。

- ✓ 教師データが不要
- ✓ 特徴数・データ数・統計的性質が異なるデータにも適用可能
- ✓ 多対多の関連性を自動的に推定



【応用例】日本語と英語の商品レビュー記事集合を用い辞書や対訳文なしで関連する単語を対応付け



関連文献

[1] T. Iwata, T. Hirao, N. Ueda, "Unsupervised cluster matching via probabilistic latent variable models," in *Proc. The Twenty-Seventh AAAI Conference on Artificial Intelligence (AAAI-13)*, 2013.

連絡先

岩田 具治 (Tomoharu Iwata) 協創情報研究部 創発知能環境研究グループ
E-mail: iwata.tomoharu{at}lab.ntt.co.jp ({at}の部分をもに置き換えてください)