

01

Finding latent relationships between different data sets

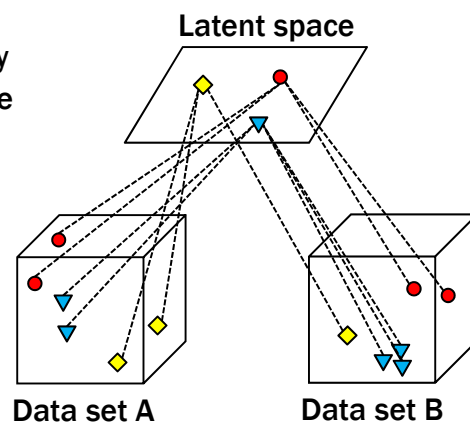
~Unsupervised object matching~

Abstract

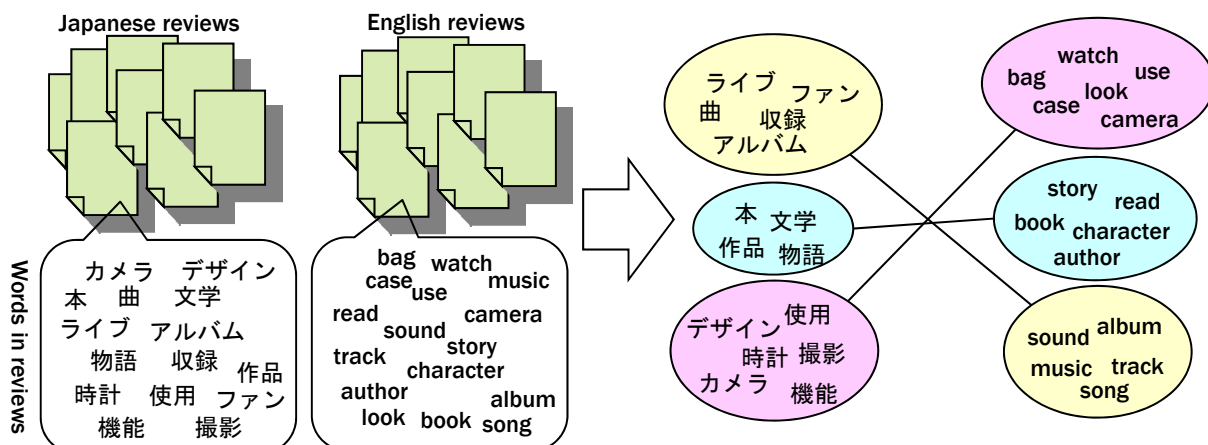
Object matching is an important task for finding correspondences between objects in different domains. It can be used for matching words in different languages, matching images and texts, and matching users in different databases. Existing object matching methods require paired data or similarity measures. However, these information might not be available because of cost or the need to preserve privacy. We propose an unsupervised object matching method that can find many-to-many matching without paired data and similarity measures by embedding all objects in a shared latent space. The proposed method enables us to analyze different data sets simultaneously, and it leads to discovery of new hidden relations and knowledge.

We can find matching over different data sets by embedding data into a shared latent space while preserving their properties.

- ✓ Unsupervised
- ✓ Applicable to data sets with different #features, #objects and properties
- ✓ Find many-to-many matching



【Application】 Matching words without dictionaries and parallel sentences using Japanese and English review articles



Related work

[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.

Contact

Tomoharu Iwata Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
E-mail : iwata.tomoharu{at}lab.ntt.co.jp (Please replace {at} with @)