Integrative Co-Embedding of Multi-View Data Sets

Haesun Park*[1], Dongjin Choit , and Barry Drake

[1]Georgia Institute of Technology – United States

Abstract

An integrative co-embedding method based on constrained low rank approximation is introduced. The method achieves knowledge fusion of multi-type data and projects the various types of objects onto a common lower-dimensional space. The goal is to produce a more informed reduced-dimensional representation that maintains both in-type and across-type semantic proximity between objects to be utilized in several key data analytics tasks such as clustering.
The effectiveness of the proposed method is illustrated using examples of document data clustering where we utilize co-embedding of papers, authors, key words, and patient profiling in healthcare data utilizing traditional medical records, as well as patients' interactions via browsing and searching on healthcare web portals. One important feature of the proposed co-embedding method is its ability to compute the basis vectors for the new embedding space. This allows embeddings for new, previously unobserved data efficiently and effectively, eliminating the need to revisit the entire data set or recomputing the embedding.

*Speaker