

We recently published a paper on a new data fusion method in IEEE Transactions on Pattern Analysis and Machine Intelligence.

For most problems in science and engineering we can obtain data sets that describe the observed system from various perspectives and record the behavior of its individual components. Heterogeneous data sets can be collectively mined by data fusion. Fusion can focus on a specific target relation and exploit directly associated data together with contextual data and data about system's constraints. In the paper we describe a data fusion approach with penalized matrix tri-factorization, called data fusion by matrix factorization (DFMF), that simultaneously factorizes many data matrices to reveal hidden associations. The approach can directly consider any data that can be expressed in a matrix, including those from feature-based representations, ontologies, associations and networks. We demonstrate the utility of DFMF for gene function prediction task with eleven different data sources and for prediction of pharmacologic actions by fusing six data sources. Our data fusion algorithm compares favorably to alternative data integration approaches and achieves higher accuracy than can be obtained from any single data source alone.

Short preprint is available at Arxiv:1307.0803. Full paper is online at IEEE.