

1126-68-199

Michael W Berry* (mberry@tennessee.edu), 401 Min H. Kao Building, 1520 Middle Drive, Knoxville, TN 37996. *Concept Extraction Using Separable Nonnegative Matrix Factorization.*

In large-scale text mining applications such as tweet classification there is need for fast yet robust techniques to summarize or track concepts without prior knowledge of the content. A numerical procedure referred to as separable nonnegative matrix factorization (SNMF) has previously been shown to be quite successful in solving solving hyperspectral unmixing problems. We have adapted this method for automated summarization of time-sensitive documents, especially social media. SNMF is designed to extract a cone spanned by a small subset of the columns of the input nonnegative data matrix. In the context of text mining, this translates to the extraction of key documents (tweets) that may well summarize a current stream and thereby reduce the human effort of reading enormous numbers of documents to extract meaning (concepts) from the stream. In essence, SNMF can function as an unsupervised learner that does not require prior labeling or metadata. (Received January 13, 2017)