02 Extracting common patterns from multiple data sets
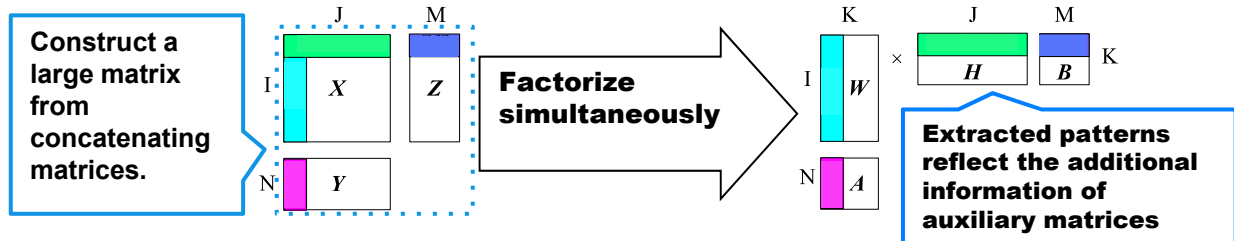## ～Non-negative multiple matrix factorization: NM2F～

Abstract

We propose a novel matrix factorization method for heterogeneous data called Non-negative Multiple Matrix Factorization (NM2F), which utilizes parts of the data as auxiliary matrices that share the row or column indices of the target matrix. The performances of the factorization are improved by decomposing the target and auxiliary matrices simultaneously, since the auxiliary matrices provide additional information about the bases and coefficients. We examined NM2F in the real data experiments. The effect of the auxiliary matrices was confirmed by the improved NM2F performance. It was also confirmed that the bases obtained from the real data that NM2F exhibited intuitive and consistent patterns thanks to the non-negative constraint.

Analyzing multiple data simultaneously yields intuitive and consistent patterns from heterogeneous data sets

✓ Non-negative Multiple Matrix Factorization (NM2F)
⇒ NM2F factorize target matrix X and auxiliary matrices Y and Z simultaneously

Construct a large matrix from concatenating matrices.

Factorize simultaneously

**Extracted patterns reflect the additional information of auxiliary matrices**

Application: Analyzing the latent structures of a picture book data set
·The picture book data set consists of term frequency, target age, and part of speech matrices.

✓ **Pattern 1 (Yellow):**
Picture books targeted at crawlers are extracted. Those books frequently use interjections ("さあ", "ほら", etc.).
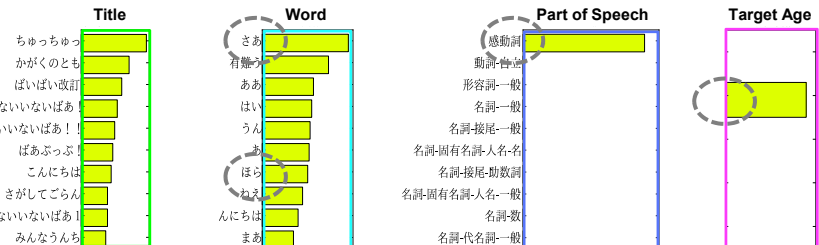
✓ **Pattern 2 (Red):**
Picture books targeted at three-years-old children are extracted. Those books frequently use proper nouns.
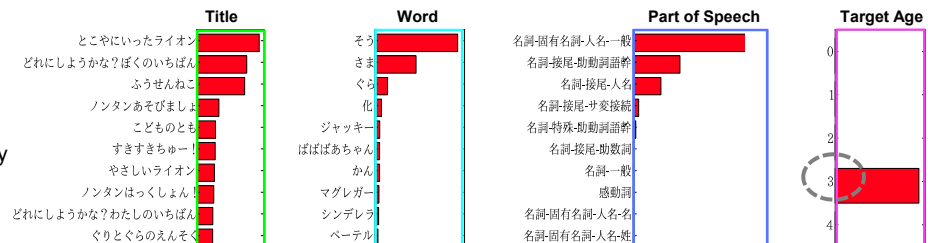
Related work

[1] 竹内孝, 石黒勝彦, 小林哲生, 藤田早苗, 平博順, "複合非負値行列因子分解(NM2F)による絵本データセットからの多角的パターン抽出," 人工知能学会全国大会, 2014 (in Japanese).
[2] Koh Takeuchi, Katsuhiko Ishiguro, Akisato Kimura, and Hiroshi Sawada. "Non-negative multiple matrix factorization," In *Proc. IJCAI,* 2013.
[3] Koh Takeuchi, Ryota Tomioka, Katsuhiko Ishiguro, Akisato Kimura, and Hiroshi Sawada. "Non-negative multiple tensor factorization," In *Proc. ICDM*, 2013.

Contact

Koh Takeuchi   Learning and Intelligent Systems Research Group, Innovative Communication Laboratory
E-mail : takeuchi.koh{at}lab.ntt.co.jp (Please replace {at} with @)