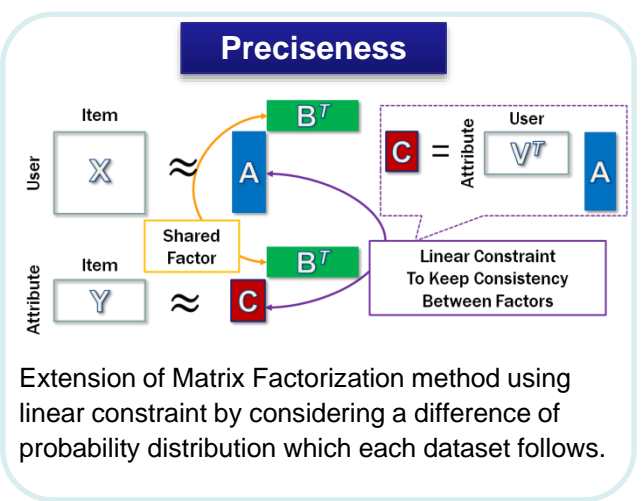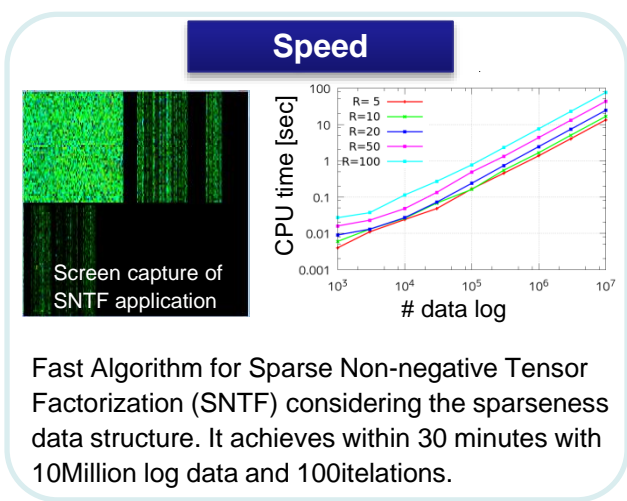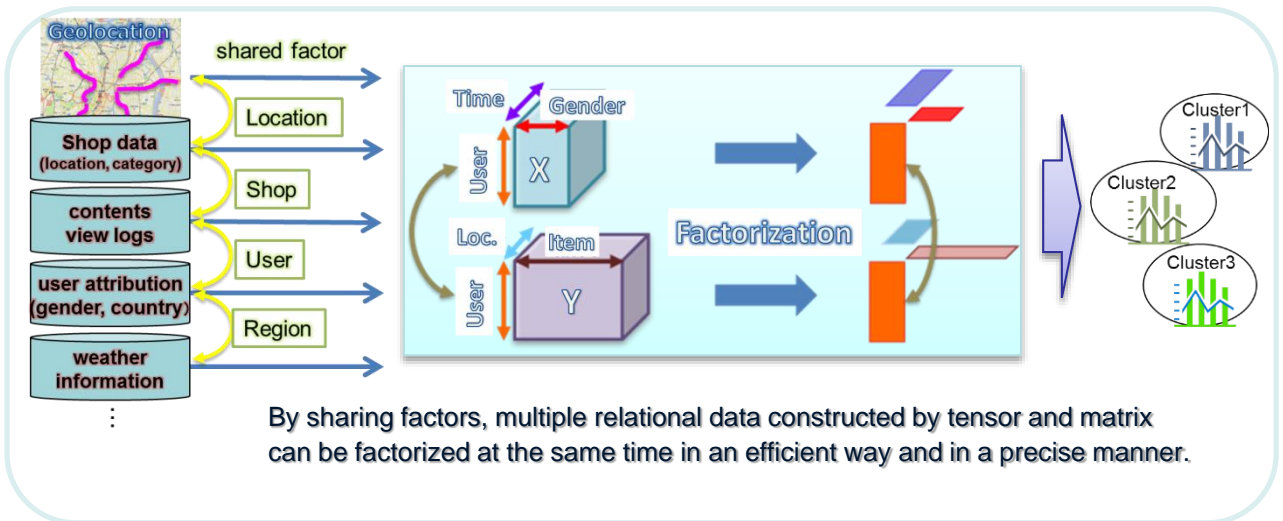# 02 Finding various factors hidden in data
~ Advanced and fast high-dimensional multiple factorization ~

Abstract

Our research focuses on finding various factors hidden in data. Suppose that we have a dataset of purchase records in which each purchase is represented by hashed user ID, user attribute, purchased good, purchase time, and store. Our method analyses the dataset and finds some specific tendencies, e.g., business persons tend to buy burgers for lunch at convenience stores, in an efficient way and in a precise manner. For efficient computations, we optimize the data structure and algorithm so that sparse data entries are aligned and accessed sequentially. For preciseness, we introduce some constraints that represents the relationship between a user ID and its attribute. Our proposed method contributes to discovering new insights from ever increasing behavior logs of humans and machines.

By sharing factors, multiple relational data constructed by tensor and matrix can be factorized at the same time in an efficient way and in a precise manner.

Speed

Screen capture of SNTF application

Fast Algorithm for Sparse Non-negative Tensor Factorization (SNTF) considering the sparseness data structure. It achieves within 30 minutes with 10Million log data and 100itelations.

Preciseness

Extension of Matrix Factorization method using linear constraint by considering a difference of probability distribution which each dataset follows.

Related works

Matsubayashi, T. et al., "Brand-Choice Analysis using Non-negative Tensor Factorization," Transactions of the Japanese Society for Artificial Intelligence, to appear.

Kohjima, M. et al., "Non-negative Multiple Matrix Factorization for Consumer Behavior Pattern Extraction by Considering Attribution Information", Transactions of the Japanese Society for Artificial Intelligence, to appear.

Kohjima, M. et al., "Non-negative Matrix Factorization for Inconsistent Resolution Data Analysis", The 29th Annual Conference of the Japan Society of Artificial Intelligence (JSAI2015), 2015.

Contact

Tatsushi Matsubayashi Machine Learning and Data Science Center
E-mail : matsubayashi.tatsushi(at)lab.ntt.co.jp

Machine **L**earning **C**enter
and Data Science