Author Contributions Checklist Form

This form documents the artifacts associated with the article (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper **does not** involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

We analyze the following dataset in Section 6.

Chicago crime dataset: this dataset consists of crime counts reported in the city of Chicago, ranging from January 1st, 2001 to December 11th, 2017. The observed tensor is an order-3 tensor with entries representing the log counts of crimes from 24 hours, 77 community areas, and 32 crime types.

Availability

☒ Data **are** publicly available

☐ Data **cannot be made** publicly available

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available data* section, below.

Publicly available data

☐ Data are available online at:

☒ Data are available as part of the paper's supplementary material.

☐ Data are publicly available by request, following the process described here:

Chicago dataset can be accessed through the R package, SmoothTensor. Use data ltns in the pacakge. Alternative way is to use raw source of data from http://frostt.io/tensors/chicago-crime/ and use the code in data/Chicago_data_processing.R

We also provide the dataset in https://github.com/Chanwoost/Smooth-tensor-estimation-with-unknown-permutations

☐ Data are or will be made available through some other mechanism, described here:

Non-publicly available data

Discussion of lack of publicly available data:

Description

File format(s)

☐ CSV or other plain text:
☒ Software-specific binary format (.Rda, Python pickle, etc.): Chicago.RData
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (described here):

Data dictionary

☐ Provided by the authors in the following file(s):
☒ Data file(s) is (are) self-describiing (e.g., netCDF files)
☐ Available at the following URL:

Additional information (optional)

Part 2: Code

Abstract

We provide our codes in https://github.com/Chanwoost/Smooth-tensor-estimation-with-unknown-permutations/tree/main/code

We build R functions in code/functions_blk.R that implement Borda count algorithm and other alternative algorithms. R functions in code/functions_asym.R is for the assymetric version of functions_blk.R. We use these R functions for all the simulations and real data analyses in the paper.

We provide the R scripts for reproducing Figure 3-5, Figure S1-3, Table 2-3, and Table S2. We describes the R scripts in code/README.md.

We produce the outputs from R scripts (Figure3.R, Figure4.R, Table3.R) and upload them as zip format.

Description

Code format(s)

☒ Script files
- ☒ R ☐ Python ☐ Matlab
- ☐ Other:

☒ Package
- ☒ R ☐ Python ☐ MATLAB toolbox
- ☐ Other:

☐ Reproducible report
- ☐ R Markdown ☐ Jupyter notebook
- ☐ Other:

☐ Shell script

☐ Other (described here):

Supporting software requirements

Version of primary software used

R version 3.5.1

Libraries and dependencies used by the code

Matrix version 1.2-18.
rTensor version 1.4.1.
softImpute version 1.4.
methods version 4.0.3.
stats version 4.0.3.

Libraries for figures:
ggplot2 version 3.3.3.
ggpubr version 0.4.0.
dplyr version 1.0.5.
plyr version 1.8.6.
rgl vesion 0.105.22.
RcolorBrewer version 1.1-2
colorRamps version 2.3.
plot.matrix version 1.5.2.

Supporting system/hardware requirements (optional)

Parallelization used

☒ No parallel code used
☐ Multi-core parallelization on a single machine/node
Number of cores used:
☐ Multi-machine/multi-node parallelization
Number of nodes and cores used:

License

☒ MIT License (default)

☐ BSD
☐ GPL v3.0
☐ Creative Commons
☐ Other (described here):

Additional information (optional)

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:

☐ Any numbers provided in text in the paper

☐ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))

☐ All tables and figures in the paper

☒ Selected tables and figures in the paper, as explained and justified here:

We provide all tables and figures in the paper except Figure 1-2 and Table 1, which are provided as illustrations of the model and the estimation method.

Workflow details

Location

The workflow is available:

☐ As part of the paper's supplementary material

☒ In this Git repository: https://github.com/Chanwoost/Smooth-tensor-estimation-with-unknown-permutations

☐ Other:

Format(s)

☐ Single master code file

☐ Wrapper (shell) script(s)

☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

☒ Text file (e.g., a readme-style file) that documents workflow

☐ Makefile

☐ Other (more detail in 'Instructions' below)

Instructions

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

☐ <1 minute

☐ 1-10 minutes

☐ 10-60 minutes

☐ 1-8 hours

☐ >8 hours

☒ Not feasible to run on a desktop machine, as described here:

The total run-time depends on the combination of the number of simulation replicates, dimensions, group sizes, models, and data types in consideration. For examples, Figure 3 considers 300 combinations and Figure 4 considers 200 combinations. Because each model fitting is separate from the other, we run these simulations in parallel in server with four cores. In the codes, we provide an example of 1 combination (for example, BATCH = 1 out of 1:300 in Figure3.R). Outputs of all combinations are in /code/Data_*.zip files.

Additional documentation (optional)

Notes (optional)