Docket #: S19-336

Assumption-Free Contrastive Component Analysis for Finding Data Trends

Researchers at Stanford have developed, for the first time, a component analysis algorithm that does not require any assumption on the data structure or data generation process to find out the important components or trends in data. **Since the new method is not based on any assumption on underlying data structure, it can be applied in any fields and on any type of data**. Data visualization, exploration, compression (dimension reduction), classification, clustering, etc. are critical tasks in various scientific and engineering fields. Such operations require finding the important data components or trends that best describe the variation in the data. There are several methods available (e.g., principal component analysis or PCA) that can serve these purposes in very specific cases based on the data structure or data acquisition methods. However, existing methods are based on some assumptions of underlying data structure and fail when these assumptions are not fulfilled. The new method does not impose any assumption on data structure or components and still is able to explore the data and ascertain the important components.

Stage of Development
Ongoing. The researchers have shown that the technique is effective at revealing the underlying dominant trends in datasets of protein expression and single-cell RNA sequencing, computed tomography, electroencephalography and wearable physiological sensors. (see *Nature* publication below).

Applications

- Data visualization, exploration, compression, clustering, classification or other application in any field, where finding the important components or trends is

required

Advantages

- Existing methods (e.g., PCA, ICA, t-SNE, MDS) may not be applicable in many cases of scientific interest, where data structure assumptions are not satisfied partially or fully.

Publications

- Islam, M.T., Xing, L. A data-driven dimensionality-reduction algorithm for the exploration of patterns in biomedical data. *Nat Biomed Eng* (2020). https://doi.org/10.1038/s41551-020-00635-3

Innovators

- Lei Xing
- Md Tauhidul Islam

Licensing Contact

Matthew Ryan

Licensing Manager

Email