

Two Heads Better Than One: Pattern Discovery in Time-Evolving Multi-aspect Data

Jimeng Sun¹, Charalampos E. Tsourakakis², Evan Hoke⁴, Christos Faloutsos²,
and Tina Eliassi-Rad³

¹ IBM T.J. Watson Research Center

² Carnegie Mellon University

³ Lawrence Livermore National Laboratory

⁴ Apple Computer, Inc.

Data stream values are often associated with multiple *aspects*. For example, each value observed at a given time-stamp from environmental sensors may have an associated type (e.g., temperature, humidity, etc) as well as location. Time-stamp, type and location are the three aspects, which can be modeled using a tensor (high-order array). However, the time aspect is special, with a natural ordering, and with successive time-ticks having usually correlated values. Standard multiway analysis ignores this structure. To capture it, we propose *2 Heads Tensor Analysis* (2-heads), which provides a qualitatively different treatment on time. Unlike most existing approaches that use a PCA-like summarization scheme for all aspects, 2-heads treats the time aspect carefully. 2-heads combines the power of classic multilinear analysis (PARAFAC [1], Tucker [5], DTA/STA [3], WTA [2]) with wavelets, leading to a powerful mining tool. Furthermore, 2-heads has several other advantages as well: (a) it can be computed incrementally in a streaming fashion, (b) it has a provable error guarantee and, (c) it achieves significant compression ratio against competitors. Finally, we show experiments on real datasets, and we illustrate how 2-heads reveals interesting trends in the data.

This is an extended abstract of an article published in the Data Mining and Knowledge Discovery journal [4].

References

1. Harshman, R.: Foundations of the parafac procedure: model and conditions for an explanatory multi-mode factor analysis. UCLA working papers in phonetics 16 (1970)
2. Sun, J., Papadimitriou, S., Yu, P.: Window-based tensor analysis on highdimensional and multi-aspect streams. In: Proceedings of the International Conference on Data Mining (ICDM) (2006)
3. Sun, J., Tao, D., Faloutsos, C.: Beyond streams and graphs: Dynamic tensor analysis. In: KDD (2006)
4. Sun, J., Tsourakakis, C.E., Hoke, E., Faloutsos, C., Eliassi-Rad, T.: Two heads better than one: Pattern discovery in time-evolving multi-aspect data. Data Mining and Knowledge Discovery 17(1), 111–128 (August 2008)
5. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. Psychometrika 31(3) (1966)