Tensor Factorizations, Data Fusion & Applications

Esta presentación consta de dos charlas, realizadas por el Prof. Bro, financiado por el programa **Visiting Scholar del Plan Propio (Ref PP2017-VS02)**, y su colaboradora la Doctora Evrim Acar. La presentación está co-organizada por el **Departamento de Teoría de la Señal, Telemática y Comunicaciones** y el **Programa de Doctorado en Tecnologías de la Información y la Comunicación**.

A continuación expongo un resumen del currículum de ambos y posteriormente resumen de las charlas. Las charlas tendrán lugar en la sala de Conferencias del CITIC, el viernes día 23, con la siguiente organización:

- 10:00: Tensor data analysis (Prof. Bro)
- 11:00: Descanso
- 11:30: Data Fusion based on Coupled Matrix and Tensor Factorizations (Dra. Acar)
- 12:30: Reuniones de colaboración

Para organizar las reuniones de colaboración, se debe mandar un correo con un breve resumen a José Camacho (josecamacho@ugr.es) hasta el día 21 de febrero.

Professor Bro is a prominent figure in the field of data analysis and in particular in multivariate analysis. During the last two decades, Prof. Bro has been a principal actor in the development of multivariate analysis in chemical and biological applications in the chemometrics area, in which according to google scholar is the second top researcher. His vast research work is outstanding, presenting an H index equal to 65. Prof. Bro is the author of the most referenced tutorial of PARAFAC (1933 citations), the developer of an extensively used multi-way version of Partial Least Squares (N-PLS) and co-author of more than 200 publications mainly related to the use of multi-way analysis in real life applications. His contributions to the use of constrained modelling are also very relevant, including the use of sparse methodologies for data analysis, and he has co-authored relevant articles to the problem of data fusion, a main challenge in Big Data analyses. The international influence in the community of his course “Multi-way Analysis” and his monograph "Multi-way Analysis: applications in the chemical sciences" is widely recognized. Prof. Bro has taught courses on data analysis in dozens of public and private organizations around the world. From 2001, he is the head of ODIN, an industrial research consortium offering courses, workshops, international contacts and student collaboration.

Evrim Acar is a Senior Research Scientist at Simula Research Laboratory. Her research focuses on data mining and mathematical modeling; in particular, tensor factorizations, data fusion using coupled factorizations of higher-order tensors and matrices, and their applications in diverse disciplines. She has also been with the Chemometrics and Analytical Technology group at the University of Copenhagen since March 2011. Prior to that, she was a senior researcher at the National Research Institute of Electronics and Cryptology in Turkey and a postdoctoral researcher at Sandia National Laboratories in Livermore, California. She received her B.S. in Computer Engineering from Bogazici University (Istanbul, Turkey) in 2003 and her M.S. and Ph.D. in Computer Science from Rensselaer Polytechnic Institute (Troy, NY) in 2006 and 2008, respectively.

UNIVERSIDAD DE GRANADA

Tensor data analysis (Prof. Rasmus Bro)

We will show some of the short comings of working with matrix data. We will show that basic matrix modeling tools such as principal component analysis also known as singular value decomposition (PCA/SVD) suffer from so-called rotational freedom and are not able to recover the real underlying information. In recent years non-negative matrix factorization (NMF) has been put forward as a more adequate tool in that respect but it is shown that it does not solve the problem.

Moving to tensors, a natural extension of matrices, the problem is solved by what is called PARAllel FACtor analysis (PARAFAC). We will show that it can model data that is so small that even PCA does not apply, that it handles noise much better, that it can recover the real underlying variation that it can even apply to data sets that are not in the same subspaces that it can handle extreme amounts of noise.

It all sounds a little too magical … and it is! PARAFAC is an extremely capable and powerful method. Examples will be given; mainly coming from chemistry, food and medicine.

We will discuss possibilities and shortcomings in using tensors

Data Fusion based on Coupled Matrix and Tensor Factorizations (Evrim Acar, Ph.D.)

When the goal is to discover the underlying patterns in a complex system such as the human metabolism or brain, the complexity of the problem requires the collection and analysis of data from multiple sources. Therefore, data fusion, i.e., knowledge extraction by jointly analyzing complementary data sets, is a topic of interest in many fields. For instance, in metabolomics, analytical platforms such as Liquid Chromatography - Mass Spectrometry and Nuclear Magnetic Resonance Spectroscopy are used for chemical profiling of biological samples. Measurements from different platforms are capable of detecting different chemical compounds with different levels of sensitivity, and their fusion has the potential to provide a more complete picture of the metabolome related to a specific condition. Similarly, neuroimaging modalities such as functional Magnetic Resonance Imaging (fMRI) and electroencephalography (EEG) provide information about the brain function in complementary spatio-temporal resolutions, and their joint analysis is expected to provide better understanding of brain activities. However, data fusion remains a challenging task since there is a lack of data mining tools that can jointly analyze incomplete (i.e., with missing entries) heterogeneous (i.e., in the form of higher-order tensors and matrices) data sets, and capture the underlying shared/unshared patterns. We formulate data fusion as a coupled matrix and tensor factorization (CMTF) problem and discuss its extension to structure-revealing data fusion, i.e., fusion models that can identify shared and unshared factors in coupled data sets. In order to solve the coupled factorization problem, we use an all-at-once optimization approach, which easily extends to coupled analysis of data sets with missing entries. Numerical experiments on simulated and real coupled data sets demonstrate that while traditional methods based on matrix factorizations have limitations in terms of jointly analyzing heterogeneous data sets, the structure-revealing CMTF model can successfully capture the underlying patterns by exploiting the low-rank structure of higher-order tensors. We will show the broad impact of CMTF-based fusion models with applications from metabolomics, neuroscience and recommender systems.