



Contribution ID: 18

Type: Poster

## Factor analysis and correlated topic model for multi-modal data integration

Integrating information across various data modalities can be beneficial for gaining valuable insights into underlying phenomena. Numerous methods exist for multi-modal data integration, ranging from linear matrix factorization-based approaches to nonlinear methods employing i.e. deep generative models. However, integrating data becomes particularly challenging when one or more modalities exhibit complex structures, such as textual, image-based, or spatial. In such cases, existing strategies often rely on preprocessing structured data as an initial step.

We present FACTM, a novel method that leverages a Bayesian probabilistic graphical model to address this challenge. Our approach combines two features. Firstly, it employs a multi-modal factor analysis (FA) to integrate information and identify common latent factors shared across all modalities, including structured data. Secondly, it uses a correlated topic model (CTM) to uncover the structure of the complex data. Specifically, the CTM part identifies meaningful clusters and shares information about the observation-wise changes of population fractions of specific clusters with the FA component of the model. Importantly, our model extracts information from complex modalities and runs factor analysis simultaneously, allowing both components of the model to potentially enhance each other's performance. Optimal parameters are determined using Bayesian variational inference.

In the poster, we will provide a detailed description of the model, along with results demonstrating its practical application. Specifically, we will present results obtained from two datasets: one containing medical multi-omics data with spatial single-cell modality, and another comprising videos with a complex modality consisting of short textual statements.

**Primary author:** LAZECKA, Malgorzata (University of Warsaw)

**Co-author:** OKSZA-ORZECOWSKI, Kazimierz

**Session Classification:** Break + Posters session

**Track Classification:** Poster