

**CHENNAI – PONDICHERRY**

**Joint Latent Dirichlet Allocation for Social Tags**

**Abstract**

Social tags, serving as a textual source of simple but useful semantic metadata to reflect the user preference or describe the web objects, has been widely used in many applications. However, social tags have several unique characteristics, i.e., sparseness and data coupling (i.e., non-IIDness), which makes existing text analysis methods such as LDA not directly applicable. In this paper, we propose a new generative algorithm for social tag analysis named joint latent Dirichlet allocation, which models the generation of tags based on both the users and the objects, and thus accounts for the coupling relationships among social tags. The model introduces two latent factors that jointly influence tag generation: the user's latent interest factor and the object's latent topic factor, formulated as user-topic distribution matrix and object-topic distribution matrix, respectively. A Gibbs sampling approach is adopted to simultaneously infer the above two matrices as well as a topic-word distribution matrix. Experimental results on four social tagging datasets have shown that our model is able to capture more reasonable topics and achieves better performance than five state-of-the-art topic models in terms of the widely used point-wise mutual information metric. In addition, we analyze the learnt topics showing that our model recovers more themes from social tags while LDA may lead the topic vanishing problems, and demonstrate its advantages in the social recommendation by evaluating the retrieval results with mean reciprocal rank metric. Finally, we explore the joint procedure of our model in depth to show the non-IID characteristic of social tagging process.