

Topic modeling in multimedia: algorithms and applications

Lei Xie · Jia Zeng · Zhi-Qiang Liu

Published online: 13 December 2014
© Springer-Verlag Berlin Heidelberg 2014

1 Introduction

The past decade has seen a rapid development of probabilistic topic models notably probabilistic latent semantic index (PLSI) and latent Dirichlet allocation (LDA). Originally, topic modeling methods have been used to find thematic word clusters called topics from a collection of documents. Since the bag-of-words (BOW) representations have been widely extended to represent both images and videos, topic modeling techniques have found many important applications in the multimedia area. Typical examples include natural scene categorization, human action recognition, multi-label image annotation, part of speech annotation, topic identification and spoken document segmentation. The advantage of topic models lies in their elegant graphical representations and efficient approximate inference algorithms. In the meanwhile, many real-world systems use topic modeling methods to automatically do the feature engineering job. However, different applications require investigating different topic models and inference algorithms to improve the overall performance. Therefore, how to design proper topic models and efficient inference algorithms for specific multimedia appli-

cations still remains a challenging problem. At the meantime, multimedia applications often have large-scale data sets, and thus big topic modeling techniques are still an urgent need.

This special issue aims to bring together researchers and technologists engaged in the development of topic modeling techniques for information processing, emerging multimedia applications and user-centric human computer interaction. We received over twenty papers and finally accepted eight papers after strict peer-review.

Developing efficient topic models is essential for practical applications. The paper from Yan et al. titled “Communication-efficient algorithms for parallel latent Dirichlet allocation”, describes novel communication efficient algorithms for parallel LDA, namely CE-PLDA. As we know, communication cost is a major bottleneck for large-scale parallel learning of LDA. To solve this problem, this paper introduces Zipf’s law and proposes parallel LDA algorithms that communicate only partial important information at each learning iteration. Evaluation on large-scale data sets demonstrates that the proposed algorithms can greatly reduce communication and computation costs to achieve a better scalability.

Topic models have been shown to be quite useful in many natural language processing (NLP) tasks. In this special issue, Tang et al. propose a statistical word sense aware topic model in the paper titled “Statistical word sense aware topic model”. In the NLP research community, LDA is frequently used to model semantic relations between surface words, which measures the contribution of a surface word to each topic and regards a surface word to be identical across all documents. This is not always the case in real-world applications because a surface word may present different signatures in different contexts. Therefore, Tang et al. believe that disambiguating word senses for topic models can enhance their discriminative capabilities. Experiments show that their proposed word sense aware model outperforms the baselines significantly in

L. Xie (✉)
School of Computer Science,
Northwestern Polytechnical University, Xi’an, China
e-mail: lxie@nwpu.edu.cn

J. Zeng
School of Computer Science and Technology,
Soochow University, Suzhou, China
e-mail: j.zeng@ieee.org

Z.-Q. Liu
School of Creative Media, City University of Hong Kong,
Hong Kong, SAR, China
e-mail: zq.liu@cityu.edu.hk

document clustering and improves the word sense induction as well against a standalone non-parametric model.

In real-world text classification, we usually do not know the categories beforehand or only know a part of the categories. Open-categorical text classification methods are highly desired. In the paper titled “Open-categorical text classification based on multi-LDA models”, Fu et al. try to solve this problem by introducing an open-categorical approach based on multi-LDA models. Experiments on the collected real-world data show that their approach outperforms the state-of-the-art supervised and semi-supervised SVM methods.

Another paper from Wu et al. titled “Sentence extraction with topic modeling for question-answer pair generation”, uses LDA to extract appropriate sentences for automatic question-answer (QA) pair generation. The idea is quite straightforward: they use a topic model to help determine if an article is of the same topic as a specific domain of interest. QA pairs are generated from the selected articles. Experiments show that, using the proposed topic modeling approach, a significant improvement of acceptance rate on the generated questions is achieved.

Topic segmentation aims to partition a multimedia document, e.g., a broadcast news program, a public lecture and a meeting recording, into topically coherent small segments. This task usually serves as a necessary prerequisite for many down-stream tasks. The paper from Chen et al. titled “Topic segmentation on spoken documents using self-validated acoustic cuts”, tries to solve two practical problems in spoken document segmentation: (1) the number of topics in a document has to be known prior to segmentation and (2) the segmentation heavily relies on a rich-source speech recognizer. They introduce a self-validated acoustic cuts approach (SACuts) that can determine the topic number in a document automatically and perform segmentation

on acoustic representation directly without a large vocabulary speech recognizer. Evaluation on a broadcast news topic segmentation task shows the superiority of the proposed approach. Parameters tuning is also crucial in the segmentation performance. However, previous methods use manual empirical tuning to choose parameters. The paper titled “NestDE: generic parameters tuning for automatic story segmentation”, introduces a practical general-purpose parameter tuning method. This approach is proven to be parameters-robust and generic enough to optimize the most usual types of parameters for the given corpus and the evaluation criterion.

Topic models also show great promise in solving computer vision problems. Liu et al. introduce such an approach for scene recognition. In the paper titled “Learning topic of dynamic scene using belief propagation and weighted visual words approach”, they describe an extension of the previously proposed topic model by belief propagation (TMBP) model. Specifically, they introduce the prior information of visual words and scenes into the model and experiments on the static and dynamic scenes show its effectiveness. Another paper from Zhang et al. titled “Multiple pedestrian tracking based on couple-states Markov chain with semantic topic learning for video surveillance”, proposes a new couple-states Markov chain-based multiple pedestrian tracking framework. By incorporating the topic model with latent semantic analysis with the hidden state of particle filtering inference, the tracking approach is more effective and accurate on benchmark videos.

Finally, we hope that this collection of papers will draw broad interest from readers. We would like to thank all the authors for their contributions and the reviewers for their outstanding cooperation and constructive suggestions.