Inter-Battery Topic Representation Learning

Cheng Zhang, Hedvig Kjellström and Carl Henrik Ek

{chengz, hedvig}@kth.se,carlhenrik.ek@bristol.ac.uk

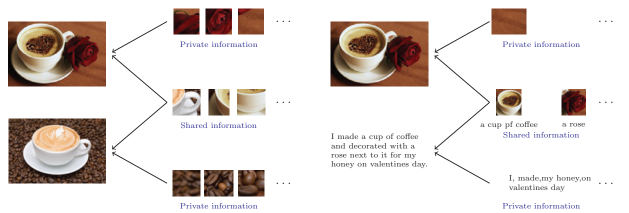
KTH VETENSKAP OCH KONST ROYAL INSTITUTE OF TECHNOLOGY

University of BRISTOL

Abstract

In this work, we present the Inter-Battery Topic Model (IBTM). Our approach extends traditional topic models by learning a factorized latent variable representation. The structured representation leads to a model that marries benefits traditionally associated with a discriminative approach, such as feature selection, with those of a generative model, such as principled regularization and ability to handle missing data. The factorization is provided by representing data in terms of aligned pairs of observations as different views. This provides means for selecting a representation that separately models topics that exist in both views from the topics that are unique to a single view. This structured consolidation allows for efficient and robust inference and provides a compact and efficient representation. Learning is performed in a Bayesian fashion by maximizing a rigorous bound on the log-likelihood. The model is then evaluated in both uni- and multi-modality settings on two different classification tasks with off-the-shelf convolutional neural network (CNN) features which generate state-of-the-art results with extremely compact representations. Additionally, we introduce a novel application with IBTM in healthcare.

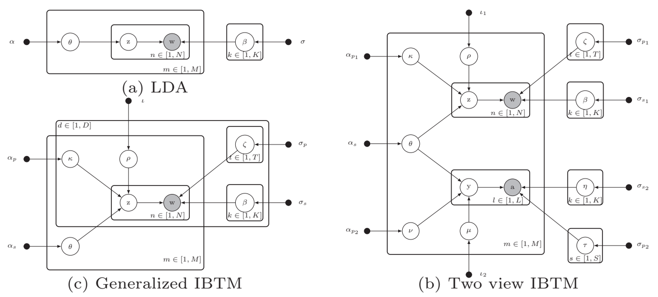
Introduction



(a) Uni-modal example (b) Multi-modal example

Factorized representations in different scenarios, for example, (a) gives an example of modeling "a cup of coffee" images. Different images with a cup of coffee all share certain patterns, such as cup handles, cup brims, etc. Moreover, each image also contains patterns that are not immediately related to the "cup of coffee" label, such as the rose or the coffee beans. These can be considered as private or instance-specific for each image. (b) gives an example of modeling the image and its caption. Different modalities describe the same content as "a cup of coffee" and "a rose". However, the wooden table pattern is not described in the caption and words such as "I made", "my honey" etc. do not correspond to the content of the image. This information can be considered as private or modality-specific.

Model



(a) LDA

(c) Generalized IBTM

(b) Two view IBTM

The key of IBTM is that we assume that topics are factorized. Taking two views scenario as an example, we do not force topics from two views to be matched completely since commonly each view has its view specific information. Hence, in our model, there is a shared topic distribution between two views for each document and there is a private topic distribution for each view respectively. As shown in the figure (b) above, $\theta \sim Dir(\alpha_s)$ is the shared per topic distribution for each document for both views, and $\beta \sim Dir(\sigma_{s1})$ and $\eta \sim Dir(\sigma_{s2})$ are the per shared topic words distribution for each view. $\kappa \sim Dir(\alpha_{p1})$ and $\nu \sim Dir(\alpha_{p2})$ are the private per document topic distribution for each view respectively, and $\zeta \sim Dir(\sigma_{p1})$ and $\tau \sim Dir(\sigma_{p2})$ are the per private topic word distribution for each view. To determine how much information is shared and how much information is private, partition parameters $\rho \sim Beta(\iota_1)$ and $\mu \sim Beta(\iota_2)$ are used for each view. In this case, to generate topic assignment for each word, z is sampled as

$$z \sim Mult([\rho * \theta; (1-\rho) * \kappa]). \quad (1)$$

Similarly, y is sampled as:

$$y \sim Mult([\mu * \theta; (1-\mu) * \nu]). \quad (2)$$

IBTM can be easily generalized to multi-view scenarios as shown in the graphical representation (c) above.

Mean Field Variational Inference is used to compute the latent variables of the model.

References

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. *Latent Dirichlet allocation* Journal of Machine Learning Research, 3:993-1022, 2003

[2] C. Chemudugunta, P. Smyth, M. Steyvers. *Modeling general and specific aspects of documents with a probabilistic topic model.* In: NIPS. Volume 19. (2006) 241-248

[3] Y. Zheng, Y.J. Zhang, H. Larochelle. *Topic Modeling of Multimodal Data: An Autoregressive Approach.* In: CVPR. (2014) 1370-1377

[4] C. Zhang, H. Kjellström, C. H. Ek and B. C. Bertilson *Diagnostic Prediction Using Discomfort Drawing with IBTM* Machine Learning in Health Care (MLHC), 2016.

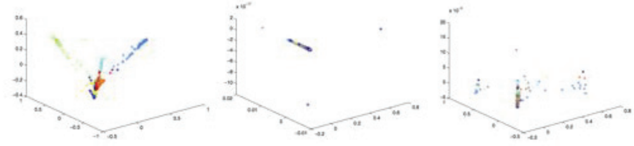
Experiment

- **LabelMe Dataset**

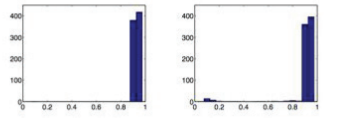
Images and Images

In this experiment, we explore the scenario in which only one modality is available. Both views are bag-of-CNN Conv5_1 feature representations of the image data. For each document, two training images from the same class are randomly paired.

DocNADE [3]	SupDocNADE [3]	Full SVM	PCA15 SVM	LDA15	SWB15 [2]	IBTM15
81.97%	83.43%	87%	80.88%	85.25%	59.88%	**89.75%**



(a) θ, Shared (b) κ, Private (c) ν, Private



(a) Hist of ρ (Img) (b) Hist of μ (Img)

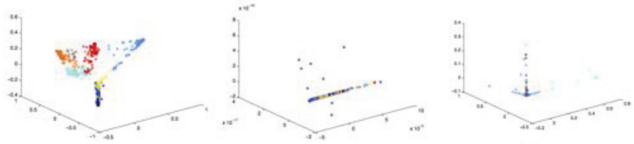
The figure above visualizes the shared topic representation (θ) and private topic representations (κ and ν). The documents of different classes are colored differently and the plots show the first three principal components after applying PCA on the per document topic distributions for all the training data. The figure on the left shows the histogram over partition parameters.

Images and Annotations

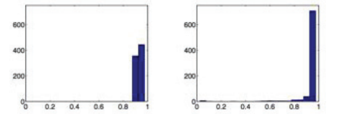
In this experiment, we explore the scenario when two different modalities are available for different views. We use the bag-of-Conv5_1 representation of images as the first view and the image annotations as the second view.

Full SVM	PCA 15	LDA15	SWB 2V [2]	IBTM15 1V	IBTM15 2V
87.63%	84.88%	85.38%	61%	**89.38%**	**95%**

The performance comparison for the image-annotation experiment for the LabelMe dataset.



(a) θ, Shared (b) κ, Private (c) ν, Private



(a) Hist of ρ (Img) (b) Hist of μ (Ann)

The figure above visualizes the topic representation and the figure on the left shows the histogram over partition parameters. The figures indicate that most annotation information is more essential. This is consistent with the intuition of the relative noise levels in image vs annotation data.

- **Other**

IBTM is also evaluated with different features. The results show that more information can be shared with higher quality of features. We also evluatate IBTM over synthetic data and Leeds butterfly dataset where consistent results can be found in the paper.

Additional Application

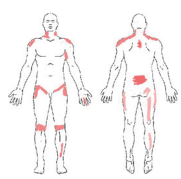
Application on diagnostic Prediction

The goal of this application [4] is to prediction possible diagnostic results given a discomfort drawing image.

Dataset A dataset with 174 authentic patient discomfort drawings were collected from clinic records with diagnostic labels from medical experts. The clinic is specialized in diagnosing unspecific pain and discomfort and patients referred often have neuropathic pain syndromes.

	$K=5$	$K=10$	$K=20$	$K=30$
F-measure	$34.31 \pm 1.35\%$	$36.7 \pm 1.37\%$	$38.32 \pm 1.1\%$	$38.56 \pm 1.23\%$

A typical example of predictive performance.



36 Prd: L neck dcf; L shoulder impingement; R shoulder impingement; L shoulder dcf;R shoulder dcf; L upper trapezius dcf; R upper trapezius dcf; Lumbago;L crest of the ilium dcf; R crest of the ilium dcf; L adductor tendonitis; R back thigh dcf; L PFS; R PFS; R calf dcf; L back thigh dcf; R anterior knee dcf;Coccydynia; L anterior knee dcf; R medial knee dcf;

L C4 Rdc; R C4 Rdc; L C6 Rdc; L C7 Rdc; L L5 Rdc; R L5 Rdc; R S1 Rdc; L S1 Rdc; L S2 Rdc; R S2 Rdc; DLI C3-C4; DLI C5-C6; DLI C6-C7; DLI L4-L5;DLI L5-S1; DLI S1-S2;

27 GT: L neck dcf; R neck dcf; L shoulder impingement; R shoulder impingement; L shoulder dcf; R shoulder dcf;; Interscapular dcf; L PFS; R PFS; Lumbago; L crest of the ilium dcf; R crest of the ilium dcf; L adductor tendonitis; R adductor tendonitis; R sciatica; L shin discomfort; R side thigh dcf;

L C5 Rdc; R C5 Rdc; L C7 Rdc; L L5 Rdc; R L5 Rdc; R S1 Rdc; DLI C5-C6; DLI C6-C7;DLI L4-L5; DLI L5-S1;

Acknowledgement

This research has been supported by the Swedish Research Council (VR), Stiftelsen Promobilia.