Discovering Discrete Latent Topics with Neural Variational Inference

A. Discovered Topics

Table 1 presents the topics by the words with highest probability (top-10 words) achieved by different neural topic models on *20NewsGroups* dataset.

Space	***Religion***	***Encryption***	***Sport***	***Science***
space	god	encryption	player	science
satellite	atheism	device	hall	theory
april	exist	technology	defensive	scientific
sequence	atheist	protect	team	universe
launch	moral	americans	average	experiment
president	existence	chip	career	observation
station	marriage	use	league	evidence
radar	system	privacy	play	exist
training	parent	industry	bob	god
committee	murder	enforcement	year	mistake

(a) Topics learned by GSM.

Space	***Religion***	***Lawsuit***	***Vehicle***	***Science***
moon	atheist	homicide	bike	theory
lunar	life	gun	motorcycle	science
orbit	eternal	rate	dod	gary
spacecraft	christianity	handgun	insurance	scientific
billion	hell	crime	bmw	sun
launch	god	firearm	ride	orbit
space	christian	weapon	dealer	energy
hockey	atheism	knife	oo	experiment
cost	religion	study	car	mechanism
nasa	brian	death	buy	star

(b) Topics learned by GSB.

Aerospace	***Crime***	***Hardware***	***Technology***	***Science***
instruction	gun	drive	technology	science
spacecraft	weapon	scsi	americans	hell
amp	crime	ide	pit	scientific
pat	firearm	scsus	encryption	evidence
wing	criminal	hd	policy	physical
plane	use	go	industry	eternal
algorithm	control	controller	protect	universe
db	handgun	tape	privacy	experiment
reduce	law	datum	product	reason
orbit	kill	isa	approach	death

(c) Topics learned by RSB.

Table 1. Topics learned by neural topic models on 20NewsGroups dataset.

B. Topic Diversity

An issue that exists in both probabilistic and neural topic models is redundant topics. In neural models it straightforward regularises the distance between each of the topic vectors in order to diversify the topics. Following Xie et al. (2015), we apply such topic diversity regularisation while carrying out neural variational inference. We use the cosine distance to measure the distance between two topics $a(t_i, t_j) = arccos(\frac{|t_i \cdot \beta_j|}{||t_i|| \cdot ||t_j||})$. The mean angle of all pairs of K topics is $\zeta = \frac{1}{K^2}\sum_i \sum_j a(t_i, t_j)$, and the variance is $\nu = \frac{1}{K^2}\sum_i \sum_j (a(t_i, t_j) - m)^2$. We add the following topic diversity regularisation to the variational objective:

$$\mathcal{J} = \mathcal{L} + \lambda(\zeta - \nu),$$

where λ is a hyper-parameter for the regularisation that is empirically set as 0.1. Though in practise diversity regularisation does not provide a significant improvement to perplexity (2$\sim$5 in most cases), it helps reduce topic redundancy and can be easily applied on topic vectors instead of the simplex over the full vocabulary.