Topic Compositional Neural Language Model

Wenlin Wang, Zhe Gan, Wenqi Wang, Dinghan Shen
Jiaji Huang, Wei Ping, Sanjeev Satheesh, Lawrence Carin

Duke University and Baidu Silicon Valley AI Lab (SVAIL)

INTRODUCTION

Objective: Developing a joint framework to learn a language model and a topic model simultaneously .

Main idea:

- Constructing a neural topic model (NTM) upon the variational auto-encoder framework.
- Mixture-of-Experts (MoE) model is proposed, each expert being response to a semantic topic.
- Our TCNLM extends each weight matrix of the long short-term memory (LSTM) to be an ensemble of topic-dependent weight matrices.
- The degree to which each expert of the ensemble is used to generate the language sequence is tied to the probability of the corresponding topic.
- The entire framework can be optimized through variational inference.

Topic Compositional Neural Language Model

Neural Topic Model: A topic model takes bag-of-words representation $\boldsymbol{d} \in \mathbb{Z}_+^D$ of a document as input and recovers itself from a sparse latent representation $\boldsymbol{t}$. The generative process of our NTM is

$$\boldsymbol{\theta} \sim \mathcal{N}(\mu_0, \sigma_0^2) \qquad \boldsymbol{t} = g(\boldsymbol{\theta})$$
$$z_n \sim \text{Discrete}(\boldsymbol{t}) \qquad w_n \sim \text{Discrete}(\boldsymbol{\beta}_{z_n}), \tag{1}$$

$g(\cdot)$ is a transformation function that maps sample $\boldsymbol{\theta}$ to the topic embedding $\boldsymbol{t}$, defined here as $g(\boldsymbol{\theta}) = \text{softmax}(\hat{\mathbf{W}}\boldsymbol{\theta} + \hat{\boldsymbol{b}})$. The marginal likelihood for $\boldsymbol{d}$ is:

$$\begin{aligned} p(\boldsymbol{d}|\mu_0, \sigma_0, \boldsymbol{\beta}) &= \int_{\boldsymbol{t}} p(\boldsymbol{t}|\mu_0, \sigma_0^2) \prod_n \sum_{z_n} p(w_n|\boldsymbol{\beta}_{z_n}) p(z_n|\boldsymbol{t}) d\boldsymbol{t} \\ &= \int_{\boldsymbol{t}} p(\boldsymbol{t}|\mu_0, \sigma_0^2) p(\boldsymbol{d}|\boldsymbol{\beta}, \boldsymbol{t}) d\boldsymbol{t}\,. \end{aligned} \tag{2}$$

Neural Language Model:

- Assume a MoE language model, which consists a set of "expert networks", *i.e.*, $E_1, E_2, ..., E_T$. Each expert is itself an RNN with its own parameters corresponding to a latent topic. Let $\mathcal{W} \in \mathbb{R}^{n_h \times n_x \times T}$ and $\mathcal{U} \in \mathbb{R}^{n_h \times n_h \times T}$. All T experts work cooperatively to generate an output y_m:

$$p(y_m) = \sum_{k=1}^{T} \boldsymbol{t}_k \cdot \text{softmax}(\mathbf{V}\boldsymbol{h}_m^{(k)}) \tag{3}$$
$$\boldsymbol{h}_m^{(k)} = \sigma(\mathcal{W}[k]\boldsymbol{x}_m + \mathcal{U}[k]\boldsymbol{h}_{m-1}), \tag{4}$$

- MoE module is computational prohibitive and storeae excessive. We extend the weight matrix of RNN to be an ensemble of topic-dependent weight matrices.

$$p(y_m) = \text{softmax}(\mathbf{V}\boldsymbol{h}_m) \tag{5}$$
$$\boldsymbol{h}_m = \sigma(\mathbf{W}(\boldsymbol{t})\boldsymbol{x}_m + \mathbf{U}(\boldsymbol{t})\boldsymbol{h}_{m-1}), \tag{6}$$

- We decompose $\mathbf{W}(\boldsymbol{t})$ into a multiplication of three terms $\mathbf{W}_a \in R^{n_h \times n_f}$, $\mathbf{W}_b \in R^{n_f \times T}$ and $\mathbf{W}_c \in R^{n_f \times n_x}$, where n_f is the number of factors.

$$\begin{aligned} \mathbf{W}(\boldsymbol{t}) &= \mathbf{W}_a \cdot \text{diag}(\mathbf{W}_b \boldsymbol{t}) \cdot \mathbf{W}_c \\ &= \mathbf{W}_a \cdot (\mathbf{W}_b \boldsymbol{t} \odot \mathbf{W}_c), \end{aligned} \tag{7}$$

where $\odot$ represents the Hadamard operator. $\mathbf{W}_a$ and $\mathbf{W}_c$ are shared parameters across all topics. $\mathbf{W}_b$ are the factors.

Model Architecture

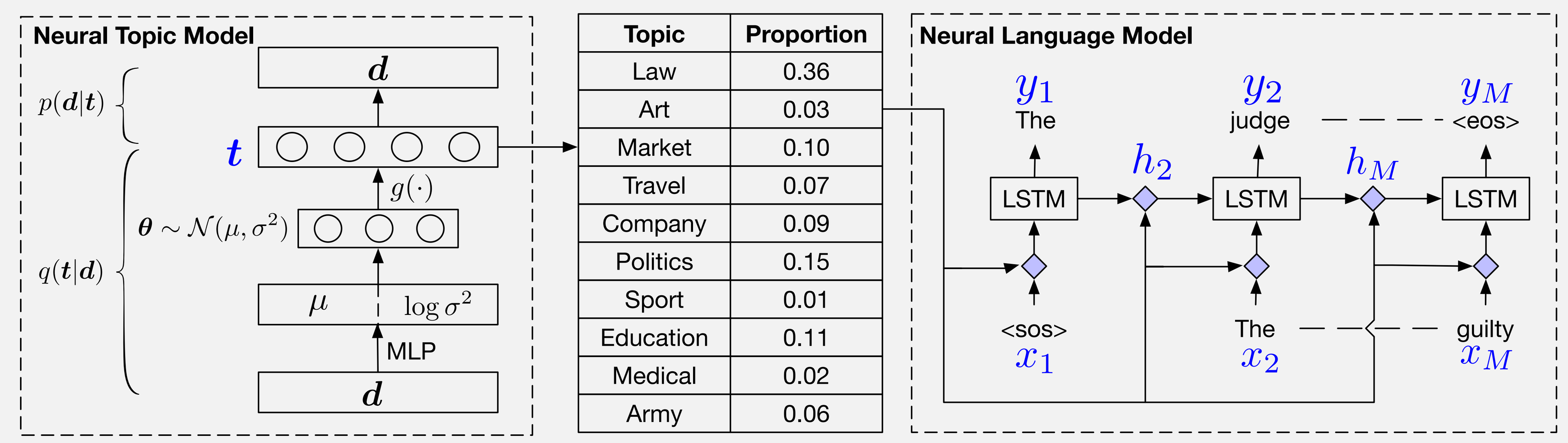

Figure: The overall architecture of the proposed model.

- We summarize out topic compositional LSTM cell as:

$$\begin{aligned} \boldsymbol{i}_m &= \sigma(\mathbf{W}_{ia}\tilde{\boldsymbol{x}}_{i,m-1} + \mathbf{U}_{ia}\tilde{\boldsymbol{h}}_{i,m-1}) \\ \boldsymbol{f}_m &= \sigma(\mathbf{W}_{fa}\tilde{\boldsymbol{x}}_{f,m-1} + \mathbf{U}_{fa}\tilde{\boldsymbol{h}}_{f,m-1}) \\ \boldsymbol{o}_m &= \sigma(\mathbf{W}_{oa}\tilde{\boldsymbol{x}}_{o,m-1} + \mathbf{U}_{oa}\tilde{\boldsymbol{h}}_{o,m-1}) \\ \tilde{\boldsymbol{c}}_m &= \sigma(\mathbf{W}_{ca}\tilde{\boldsymbol{x}}_{c,m-1} + \mathbf{U}_{ca}\tilde{\boldsymbol{h}}_{c,m-1}) \\ \boldsymbol{c}_m &= \boldsymbol{i}_m \odot \tilde{\boldsymbol{c}}_m + \boldsymbol{f}_m \cdot \boldsymbol{c}_{m-1} \\ \boldsymbol{h}_m &= \boldsymbol{o}_m \odot \tanh(\boldsymbol{c}_m)\,. \end{aligned} \tag{8}$$

For $* = i, f, o, c$, we define

$$\tilde{\boldsymbol{x}}_{*,m-1} = \mathbf{W}_{*b}\boldsymbol{t} \odot \mathbf{W}_{*c}\boldsymbol{x}_{m-1} \tag{9}$$
$$\tilde{\boldsymbol{h}}_{*,m-1} = \mathbf{U}_{*b}\boldsymbol{t} \odot \mathbf{U}_{*c}\boldsymbol{h}_{m-1}\,. \tag{10}$$

Compared with a standard LSTM cell, our LSTM unit has a total number of parameters in size of $4n_f \cdot (n_x + 2T + 3n_h)$

Model Inference

The joint marginal likelihood for the language model together with the topic model can be written as

$$p(y_{1:M}, \boldsymbol{d}|\mu_0, \sigma_0^2, \boldsymbol{\beta}) = \int_{\boldsymbol{t}} p(\boldsymbol{t}|\mu_0, \sigma_0^2) p(\boldsymbol{d}|\boldsymbol{\beta}, \boldsymbol{t}) \prod_{m=1}^{M} p(y_m|y_{1:m-1}, \boldsymbol{t}) d\boldsymbol{t}\,. \tag{11}$$

We employ variational inference and construct the evidence lower bound (ELBO) as

$$\mathcal{L} = \underbrace{\mathbb{E}_{q(\boldsymbol{t}|\boldsymbol{d})}\left(\log p(\boldsymbol{d}|\boldsymbol{t})\right) - \text{KL}\left(q(\boldsymbol{t}|\boldsymbol{d})||p(\boldsymbol{t}|\mu_0, \sigma_0^2)\right)}_{\text{neural topic model}} + \underbrace{\mathbb{E}_{q(\boldsymbol{t}|\boldsymbol{d})}\left(\sum_{m=1}^{M} \log p(y_m|y_{1:m-1}, \boldsymbol{t})\right)}_{\text{neural language model}} \tag{12}$$

EXPERIMENTS

Datasets:

Dataset	Vocabulary		Training			Development			Testing		
	LM	TM	# Docs	# Sents	# Tokens	# Docs	# Sents	# Tokens	# Docs	# Sents	# Tokens
APNEWS	32,400	7,790	$50K$	$0.7M$	$15M$	$2K$	$27.4K$	$0.6M$	$2K$	$26.3K$	$0.6M$
IMDB	34,256	8,713	$75K$	$0.9M$	$20M$	$12.5K$	$0.2M$	$0.3M$	$12.5K$	$0.2M$	$0.3M$
BNC	41,370	9,741	$15K$	$0.8M$	$18M$	$1K$	$44K$	$1M$	$1K$	$52K$	$1M$

Table: Summary statistics for the datasets used in the experiments.

Quantitative Evaluation

Dataset	LSTM type	basic-LSTM	LDA+LSTM			LCLM	Topic-RNN			TDLM			TCNLM		
			50	100	150		50	100	150	50	100	150	50	100	150
APNEWS	small	64.13	57.05	55.52	54.83	54.18	56.77	54.54	54.12	53.00	52.75	52.65	52.75	52.63	**52.59**
	large	58.89	52.72	50.75	50.17	50.63	53.19	50.24	50.01	48.96	48.97	48.21	48.07	47.81	**47.74**
IMDB	small	72.14	69.58	69.64	69.62	67.78	68.74	67.83	66.45	63.67	63.45	63.82	63.98	62.64	**62.59**
	large	66.47	63.48	63.04	62.78	67.86	63.02	61.59	60.14	58.99	59.04	58.59	57.06	56.38	**56.12**
BNC	small	102.89	96.42	96.50	96.38	87.47	94.66	93.57	93.55	87.42	85.99	86.43	87.98	86.44	**86.21**
	large	94.23	88.42	87.77	87.28	80.68	85.90	84.62	84.12	82.62	81.83	80.58	80.29	80.14	**80.12**

Table: Test perplexities of different models on APNEWS, IMDB and BNC.

Model	Coherence		
# topic=50	APNEWS	IMDB	BNC
LDA	0.125	0.084	0.106
NTM	0.075	0.064	0.081
TDLM(s)	0.149	0.104	0.102
TDLM(l)	0.130	0.088	0.095
Topic-RNN(s)	0.134	0.103	0.102
Topic-RNN(l)	0.127	0.096	0.100
TCNLM(s)	**0.159**	**0.106**	**0.114**
TCNLM(l)	0.152	0.100	0.101

Model	Coherence		
# topic=100	APNEWS	IMDB	BNC
LDA	0.136	0.092	**0.119**
NTM	0.085	0.071	0.070
TDLM(s)	0.152	0.087	0.106
TDLM(l)	0.142	0.097	0.101
Topic-RNN(s)	0.158	0.096	0.108
Topic-RNN(l)	0.143	0.093	0.105
TCNLM(s)	**0.160**	**0.101**	0.111
TCNLM(l)	0.152	0.098	0.104

Model	Coherence		
# topic=150	APNEWS	IMDB	BNC
LDA	0.134	0.094	**0.119**
NTM	0.078	0.075	0.072
TDLM(s)	0.147	0.085	0.100
TDLM(l)	0.145	0.091	0.104
Topic-RNN(s)	0.146	0.089	0.102
Topic-RNN(l)	0.137	0.092	0.097
TCNLM(s)	0.153	**0.096**	0.107
TCNLM(l)	**0.155**	0.093	0.102

Figure: Topic coherence scores of different models on APNEWS, IMDB and BNC. (s) and (l) indicate small and large model, respectively.

Qualitative Evaluation

Dataset										
	army	**animal**	**medical**	**market**	**lottory**	**terrorism**	**law**	**art**	**transportation**	**education**
APNEWS	afghanistan	animals	patients	zacks	casino	syria	lawsuit	album	airlines	students
	veterans	dogs	drug	cents	mega	iran	damages	music	fraud	math
	soldiers	zoo	fda	earnings	lottery	militants	plaintiffs	film	scheme	schools
	brigade	bear	disease	keywords	gambling	al-qaida	filed	songs	conspiracy	education
	infantry	wildlife	virus	share	jackpot	korea	suit	comedy	flights	teachers
	horror	**action**	**family**	**children**	**war**	**detective**	**sci-fi**	**negative**	**ethic**	**epsiode**
IMDB	zombie	martial	rampling	kids	war	eyre	alien	awful	gay	season
	slasher	kung	relationship	snoopy	che	rochester	godzilla	unfunny	school	episodes
	massacre	li	binoche	santa	documentary	book	tarzan	sex	girls	series
	chainsaw	chan	marie	cartoon	muslims	austen	planet	poor	women	columbo
	gore	fu	mother	parents	jews	holmes	aliens	worst	sex	batman
	environment	**education**	**politics**	**business**	**facilities**	**sports**	**art**	**award**	**expression**	**crime**
BNC	pollution	courses	elections	corp	bedrooms	goal	album	john	eye	police
	emissions	training	economic	turnover	hotel	score	band	award	looked	murder
	nuclear	students	minister	unix	garden	cup	guitar	research	hair	killed
	waste	medau	political	net	situated	ball	music	darlington	lips	jury
	environmental	education	democratic	profits	rooms	season	film	speaker	stared	trail

Table: 10 topics learned from our TCNLM on APNEWS, IMDB and BNC.

Data	Topic	Generated Sentences
APNEWS	army	• a female sergeant, serving in the fort worth, has served as she served in the military in iraq .
	animal	• most of the bear will have stumbled to the lake .
	medical	• physicians seeking help in utah and the nih has had any solutions to using the policy and uses offline to be fitted with a testing or body .
	market	• the company said it expects revenue of $ <unk> million to $ <unk> million in the third quarter .
	lottory	• where the winning numbers drawn up for a mega ball was sold .
	army+terrorism	• the taliban 's presence has earned a degree from the 1950-53 korean war in pakistan 's historic life since 1964 , with two example of <unk> soldiers from wounded iraqi army shootings and bahrain in the eastern army .
	animal+lottory	• she told the newspaper that she was concerned that the buyer was in a neighborhood last year and had a gray wolf .
IMDB	horror	• the killer is a guy who is n't even a zombie .
	action	• the action is a bit too much , but the action is n't very good .
	family	• the film is also the story of a young woman whose <unk> and <unk> and very yet ultimately sympathetic , <unk> relationship , <unk> , and palestine being equal , and the old man , a <unk> .
	children	• i consider this movie to be a children 's film for kids .
	war	• the documentary is a documentary about the war and the <unk> of the war .
	horror+negative	• if this movie was indeed a horrible movie i think i will be better off the film .
	sci-fi+children	• paul thinks him has to make up when the <unk> eugene discovers defeat in order to take too much time without resorting to mortal bugs , and then finds his wife and boys .
BNC	environment	• environmentalists immediate base calls to defend the world .
	education	• the school has recently been founded by a <unk> of the next generation for two years .
	politics	• a new economy in which privatization was announced on july 4 .
	business	• net earnings per share rose <unk> % to $ <unk> in the quarter , and $ <unk> m , on turnover that rose <unk> % to $ <unk> m.
	facilities	• all rooms have excellent amenities .
	environment+politics	• the commission 's report on oct. 2 , 1990 , on jan. 7 denied the government 's grant to " the national level of water " .
	art+crime	• as well as 36, he is returning freelance into the red army of drama where he has finally been struck for their premiere .

Table: Generated sentences from given topics.