Variational Recurrent Topic Model

Mehdi Rezaee

Francis Ferraro

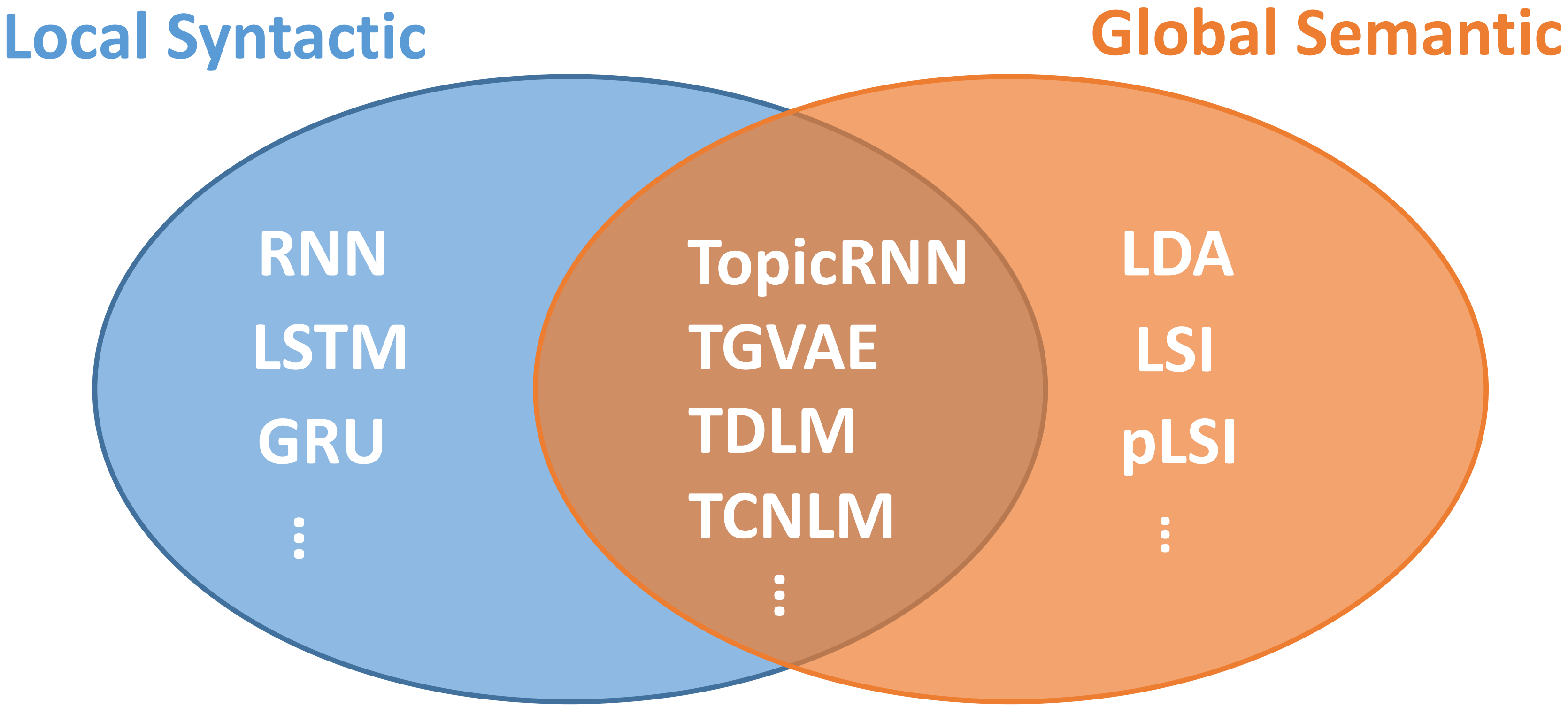
Natural Language Understanding

Thematic word relations

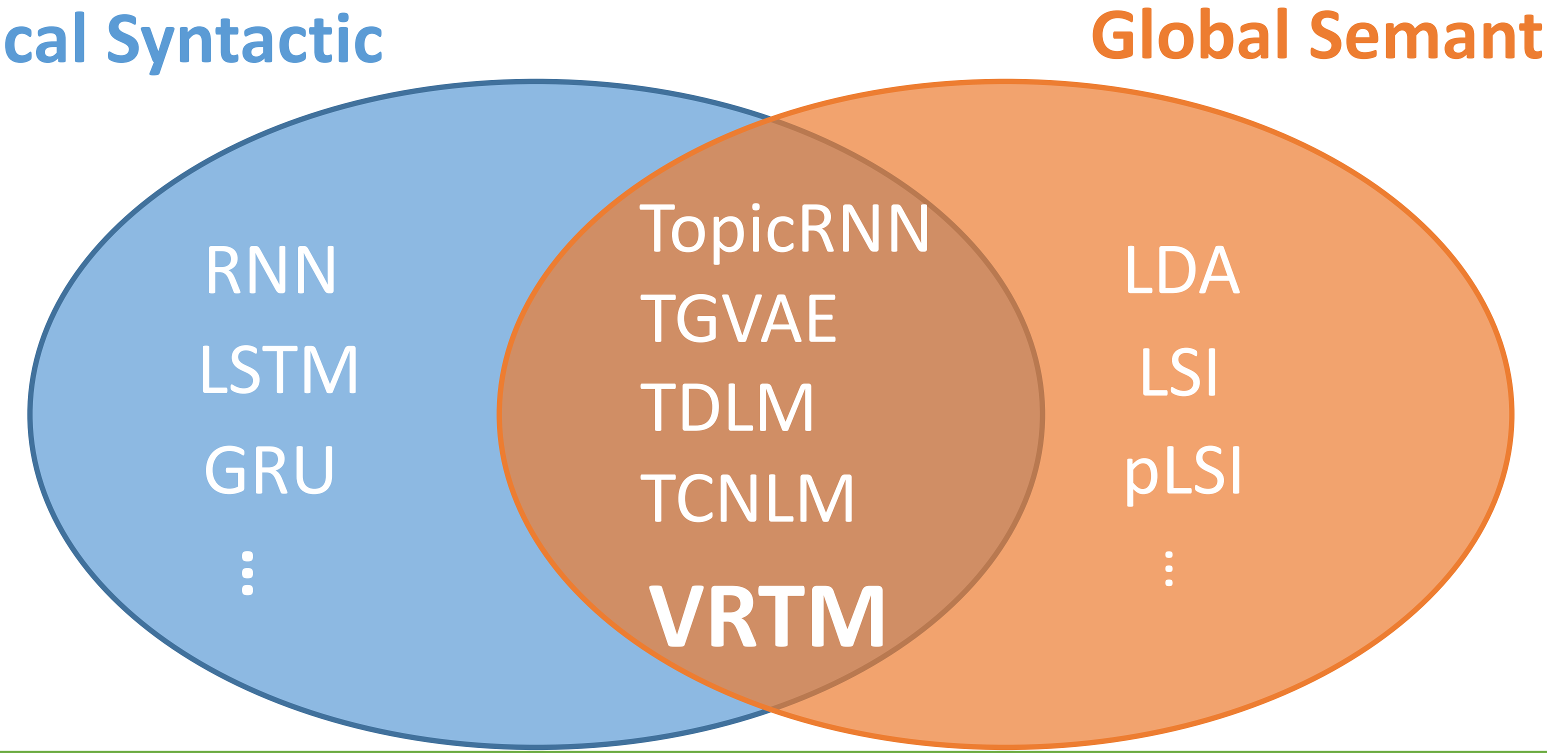
*"She received a **master's degree** in **EE** from Anytown **University**."*

Syntactic word relations

Previous Work



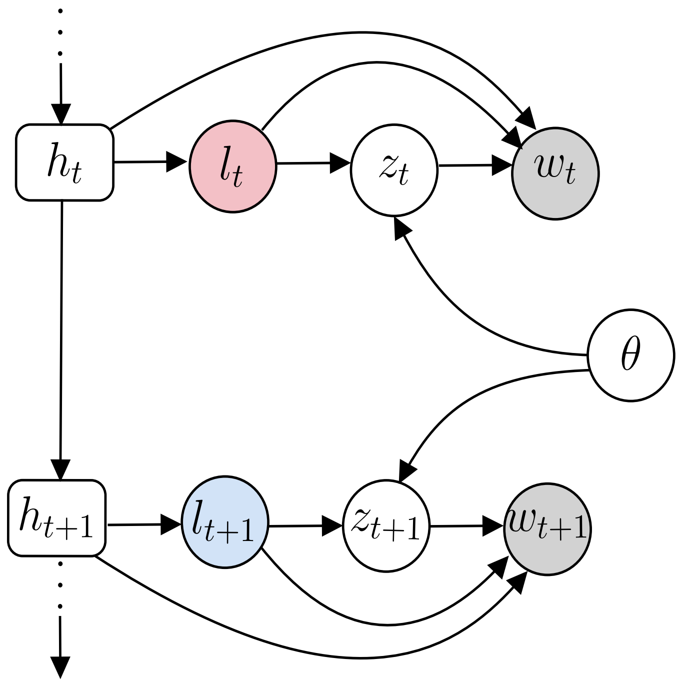
Previous Work



Previously: marginalize out *word-level* topic assignments
Our work: learn to model them within a neural framework

UMBC

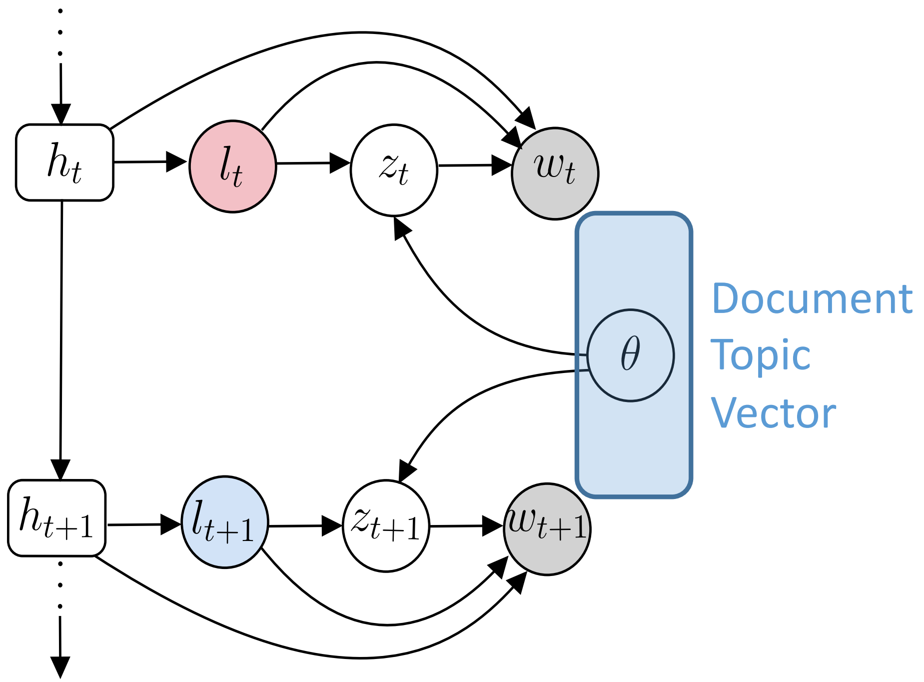
Model Architecture



UMBC

Model Architecture

1- Draw a document topic vector $\theta \sim \mathrm{Dir}(\alpha)$

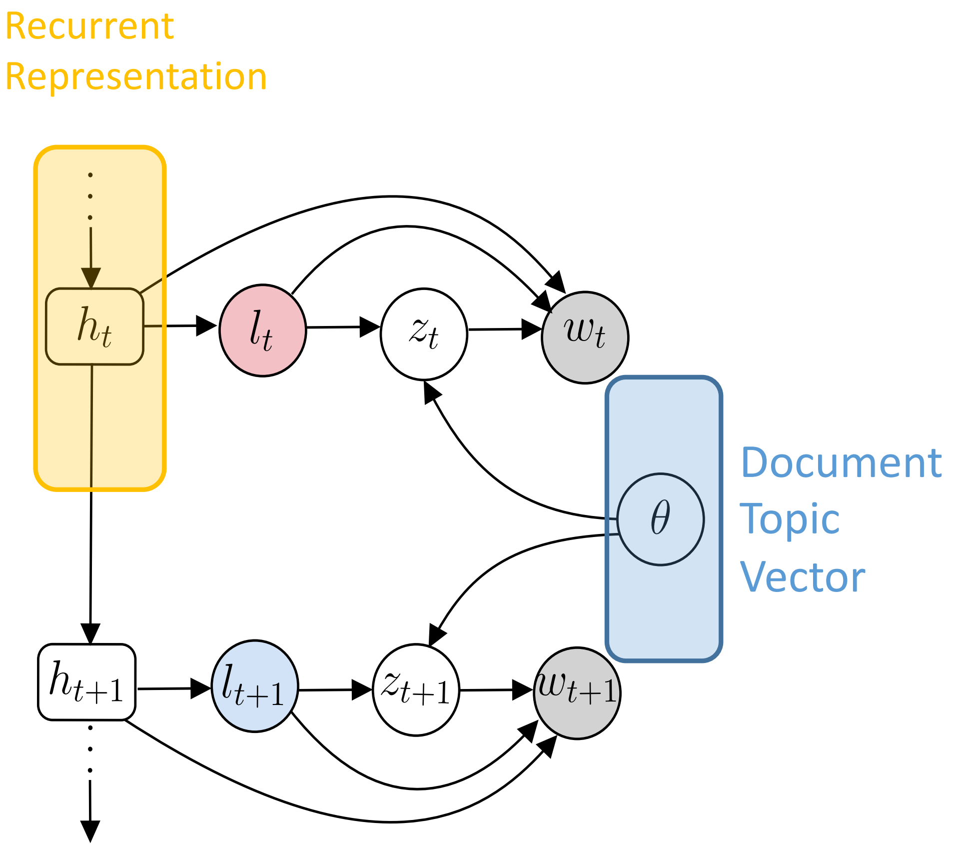


UMBC

Model Architecture

1- Draw a document topic vector $\theta \sim \mathrm{Dir}(\alpha)$

2- Compute the recurrent representation
$h_t = f(w_{t-1}, h_{t-1})$

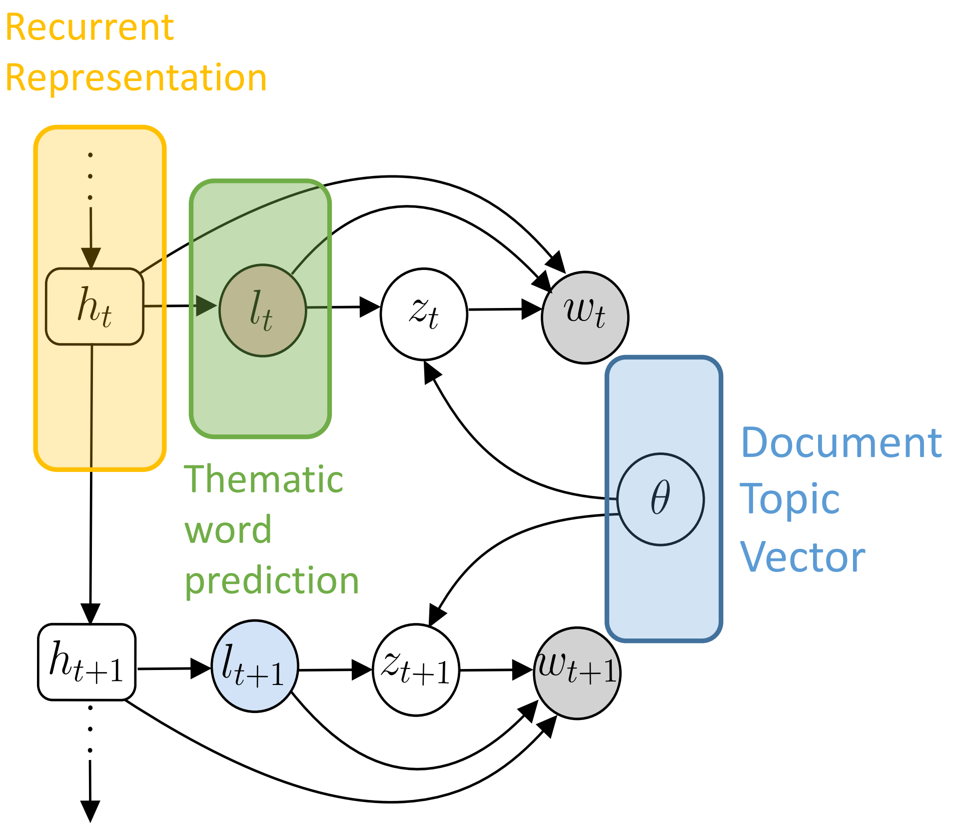


Model Architecture

1- Draw a document topic vector $\theta \sim \mathrm{Dir}(\alpha)$.

2- Compute the recurrent representation
$h_t = f(w_{t-1}, h_{t-1})$.

3- Draw $l_t \sim \mathrm{Bern}(\rho)$.



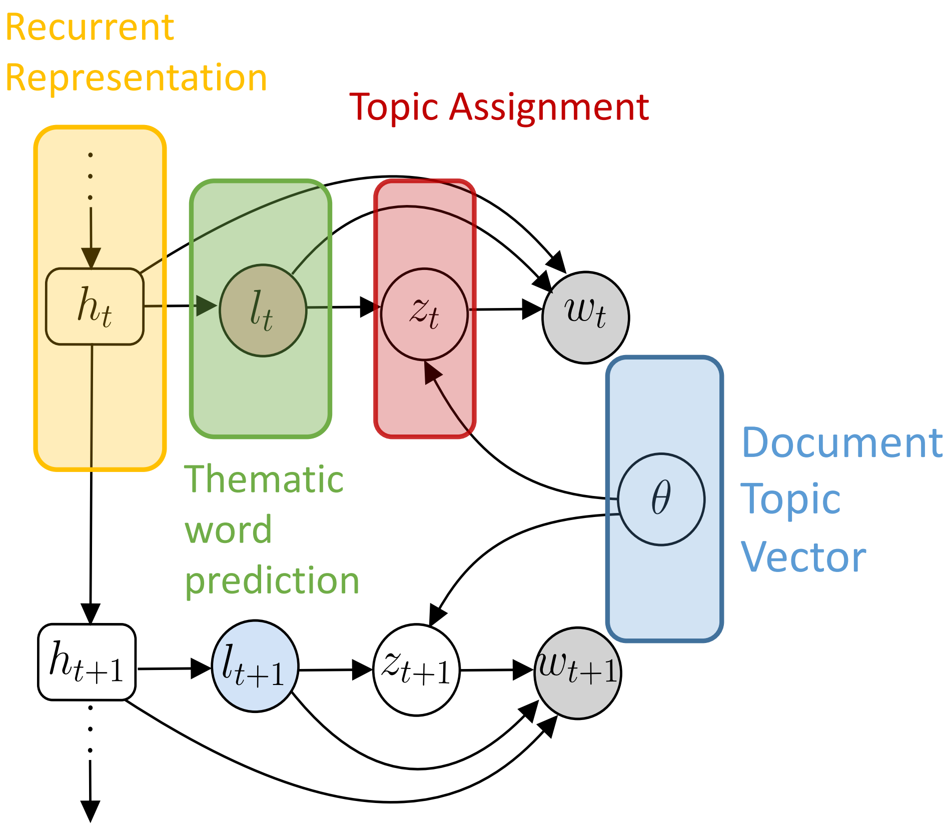
Model Architecture

1- Draw a document topic vector $\theta \sim \mathrm{Dir}(\alpha)$.

2- Compute the recurrent representation $h_t = f(w_{t-1}, h_{t-1})$.

3- Draw $l_t \sim \mathrm{Bern}(\rho)$.

4- Draw a topic $z_t \sim p(z_t | l_t, \theta)$.



UMBC

Model Architecture

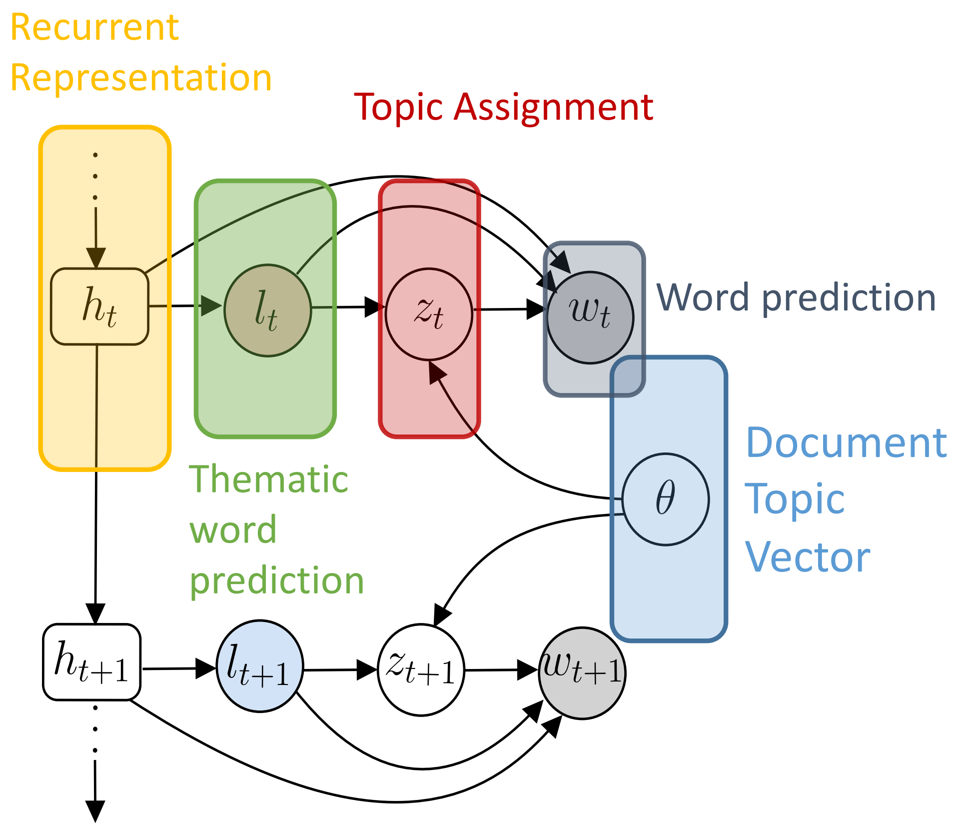
1- Draw a document topic vector $\theta \sim \mathrm{Dir}(\alpha)$.

2- Compute the recurrent representation $h_t = f(w_{t-1}, h_{t-1})$.

3- Draw $l_t \sim \mathrm{Bern}(\rho)$.

4- Draw a topic $z_t \sim p(z_t | l_t, \theta)$.

5- Draw the word $w_t \sim p(w_t | z_t; l_t, h_t, \beta)$.



UMBC

Model Architecture

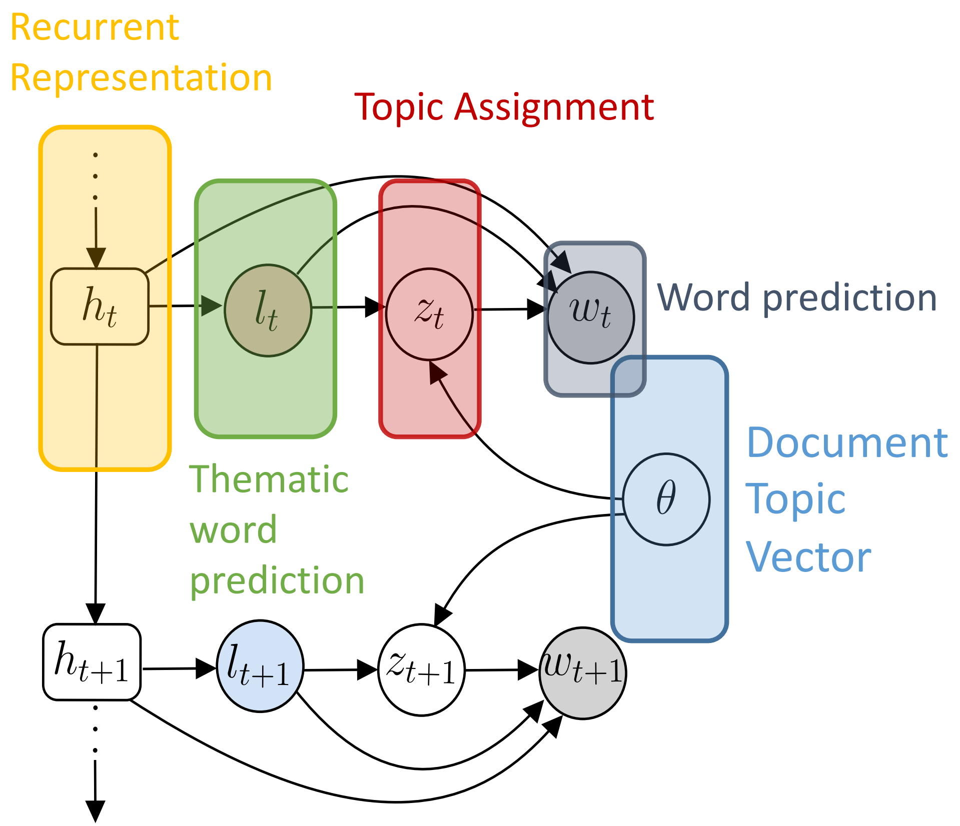
1- Draw a document topic vector $\theta \sim \mathrm{Dir}(\alpha)$.

2- Compute the recurrent representation $h_t = f(w_{t-1}, h_{t-1})$.

3- Draw $l_t \sim \mathrm{Bern}(\rho)$.

4- Draw a topic $z_t \sim p(z_t | l_t, \theta)$.

5- Draw the word $w_t \sim p(w_t | z_t; l_t, h_t, \beta)$.



$$p(w_t = v | z_t = k; h_t, l_t, \beta) \propto \exp(p_v^T h_t + (1 - l_t)\beta_{k,v})$$

Model and Inference

Joint Model:

$$p(\boldsymbol{w}, \boldsymbol{l}, \boldsymbol{z}, \theta; \beta, \boldsymbol{h}) = p(\theta) \prod_t p(w_t | z_t, l_t) p(z_t | l_t, \theta) p(l_t; h_t)$$

$\boldsymbol{w}$: *tokens.* $\boldsymbol{l}$: *word labels.* $\boldsymbol{z}$: *topics.* θ: *document vector*

UMBC

Model and Inference

Joint Model:

$$p(\boldsymbol{w}, \boldsymbol{l}, \boldsymbol{z}, \theta; \beta, \boldsymbol{h}) = p(\theta) \prod_t p(w_t | z_t, l_t) p(z_t | l_t, \theta) p(l_t; h_t)$$

Inference:

$$q(\theta, \boldsymbol{z} | \boldsymbol{w}, \boldsymbol{l}) = q(\theta | \boldsymbol{w}, \boldsymbol{l}) \prod_t q(z_t | w_t, l_t)$$

***w*: tokens.** ***l*: word labels.** ***z*: topics.** θ: *document vector*

UMBC

Stop-words topic assignment

We assume that syntactic words do not belong to any specific topic:

Model:

$$p(z_t = k | l_t, \theta) = \begin{cases} \theta_k\,, & if\ l_t = 0 \\ \frac{1}{K}, & if\ l_t = 1 \end{cases}$$

Stop-words topic assignment

We assume that syntactic words do not belong to any specific topic:

Inference:

$$q(z_t = k | w_t, \phi_t) = \begin{cases} \phi_t^k, & if\ l_t = 0 \\ \frac{1}{K}, & if\ l_t = 1 \end{cases}$$

UMBC

Performance Guarantee

Theorem 1. *If in the generative story $\rho = 1$, and we have a uniform distribution for the prior and variational topic posterior approximation $\big(p(\theta) = q(\theta|\mathbf{w}, \mathbf{l}) = 1/K\big)$, then VRTM reduces to a Recurrent Neural Network to just reconstruct the words given previous words.*

UMBC

Performance Guarantee

Theorem 1. *If in the generative story* $\rho = 1$*, and we have a uniform distribution for the prior and variational topic posterior approximation* $\left(p(\theta) = q(\theta|\mathbf{w}, \mathbf{l}) = 1/K\right)$*, then VRTM reduces to a Recurrent Neural Network to just reconstruct the words given previous words.*

VRTM outperforms an RNN cell.

UMBC

Datasets

Dataset	Vocabulary	Training Docs
APNEWS	7788	50 K
IMDB	8734	75 K
BNC	9769	15 K

VRTM as a Language Model

how well a probability model predicts a sample

$$\text{perplexity} = \exp\left(\frac{-\sum_{d=1}^{N_d} \log p(w_{1:N_d})}{N_D}\right)$$

Test Perplexity (50 Topics)

Methods	APNEWS	IMDB	BNC
Basic-LSTM	62.79	70.38	100.07
LDA+ LSTM	57.05	69.58	96.42
Topic-RNN	56.77	68.74	94.66
TDLM	53.00	63.67	91.42
TCNLM	52.75	63.98	87.98
TGVAE	48.73	57.11	87.86
VRTM (Ours)	**47.78**	**51.08**	**86.33**

Lower is Better

UMBC

Test Perplexity

Methods	APNEWS	IMDB	BNC
TGVAE (T=10)	55.77	62.22	91.19
TGVAE (T=30)	51.27	59.45	88.34
TGVAE (T=50)	48.73	57.11	87.86
VRTM (T=10)	**54.31**	**59.82**	**92.89**
VRTM (T=30)	**51.47**	**54.36**	**89.26**
VRTM (T=50)	**47.78**	**51.08**	**86.33**

Topic Switch Percent (Lund et al., 2019)

Tokens near each other should switch infrequently, and thus be consistent in expressing a single idea.

$$\text{SwitchP} = \frac{1}{T_d - 1} \sum_{t=1}^{T_d - 1} \delta(z_t, z_{t+1})$$

Topic Switch Percent (Lund et al., 2019)

Topic	APNEWS		IMDB		BNC	
	LDA	**VRTM**	LDA	**VRTM**	LDA	**VRTM**
5	0.26	**0.59**	0.24	**0.52**	0.24	**0.51**
10	0.18	**0.43**	0.14	**0.35**	0.15	**0.40**
15	0.14	**0.33**	0.12	**0.31**	0.13	**0.35**
30	0.10	**0.31**	0.09	**0.28**	0.10	**0.23**
50	0.08	**0.20**	0.07	**0.26**	0.07	**0.20**

Summary

- Integrate both topic modeling and sequential neural networks.
- Relatively consistent word-level topic assignments.
- The model offers promising word prediction and generated sentences.

VRTM as a Topic Model

Dataset	#1	#2	#3
APNEWS	dead	washington	fund
	killed	american	million
	hunting	california	bill
	deaths	texas	finance
IMDB	films	horror	friends
	directed	murder	series
	story	strange	dvd
	imdb	killing	channel
BNC	king	house	today
	londen	st	ago
	northern	street	life
	conservative	town	years

UMBC

Generated Sentences

Dataset	Generated Sentence
APNEWS	a damaged car and body <UNK> were taken to the county medical center from dinner with one driver
	another agency will investigate possible abuse of violations to the police facility .
	not even if it represents everyone under control . we are getting working with other items .
IMDB	the film is very funny and entertaining . while just not cool and all ; the worst one can be expected
	if you must view this movie , then i ’d watch it again again and enjoy it .this movie surprised me .
	they definitely are living with characters and can be described as vast in their parts .
BNC	she drew into her eyes . she stared at me . molly thought of the young lady , there was lack of same feelings of herself.
	these conditions are needed for understanding better performance and ability and entire response .
	not a conservative leading male of his life under waste worth many a few months to conform with how it was available .

UMBC