

A. GBN-RNN

GBN-RNN: $\{y_{1:T}, \mathbf{d}\}$ denotes a sentence-context pair, where $\mathbf{d} \in \mathbb{Z}_+^{V_c}$ represents the document-level context as a word frequency count vector, the v th element of which counts the number of times the v th word in the vocabulary appears in the document excluding sentence $y_{1:T}$. The hierarchical model of an L -hidden-layer GBN, from top to bottom, is expressed as

$$\begin{aligned} \boldsymbol{\theta}^L &\sim \text{Gam}(\mathbf{r}, c^{L+1}), \dots, \boldsymbol{\theta}^l \sim \text{Gam}(\boldsymbol{\Phi}^{l+1} \boldsymbol{\theta}^{l+1}, c^{l+1}), \dots, \\ \boldsymbol{\theta}^1 &\sim \text{Gam}(\boldsymbol{\Phi}^2 \boldsymbol{\theta}^2, c^2), \mathbf{d} \sim \text{Pois}(\boldsymbol{\Phi}^1 \boldsymbol{\theta}^1). \end{aligned} \quad (10)$$

The stacked-RNN based language model described in (3) is also used in GBN-RNN.

Statistical inference: To infer GBN-RNN, we consider a hybrid of stochastic gradient MCMC (Welling & Teh, 2011; Patterson & Teh, 2013; Li et al., 2015; Ma et al., 2015; Cong et al., 2017a), used for the GBN topics ϕ_k^l , and auto-encoding variational inference (Kingma & Welling, 2013; Rezende et al., 2014), used for the parameters of both the inference network (encoder) and RNN. More specifically, GBN-RNN generalizes Weibull hybrid auto-encoding inference (WHAI) of Zhang et al. (2018): it uses a deterministic-downward-stochastic-upward inference network to encode the bag-of-words representation of \mathbf{d} into the latent topic-weight variables $\boldsymbol{\theta}^l$ across all hidden layers, which are fed into not only GBN to reconstruct \mathbf{d} , but also a stacked RNN in language model, as shown in (3), to predict the word sequence in $y_{1:T}$. The topics ϕ_k^l can be sampled with topic-layer-adaptive stochastic gradient Riemannian (TLASGR) MCMC, whose details can be found in Cong et al. (2017a); Zhang et al. (2018), omitted here for brevity. Given the sampled topics ϕ_k^l , the joint marginal likelihood of $\{y_{1:T}, \mathbf{d}\}$ is defined as

$$p(y_{1:T}, \mathbf{d} | \{\boldsymbol{\Phi}^l\}_l) = \int p(\mathbf{d} | \boldsymbol{\Phi}^1 \boldsymbol{\theta}^1) \left[\prod_{t=1}^T p(y_t | y_{1:t-1}, \boldsymbol{\theta}^{1:L}) \right] \left[\prod_{l=1}^L p(\boldsymbol{\theta}^l | \boldsymbol{\Phi}^{l+1} \boldsymbol{\theta}^{l+1}) \right] d_{\boldsymbol{\theta}^{1:L}}. \quad (11)$$

For efficient inference, an inference network as $Q = \prod_{l=1}^L q(\boldsymbol{\theta}^l | \mathbf{d}, \boldsymbol{\Phi}^{l+1} \boldsymbol{\theta}^{l+1})$ is used to provide an ELBO of the log joint marginal likelihood as

$$L(y_{1:T}, \mathbf{d}) = \mathbb{E}_Q \left[\ln p(\mathbf{d} | \boldsymbol{\Phi}^1 \boldsymbol{\theta}^1) + \sum_{t=1}^T \ln p(y_t | y_{1:t-1}, \boldsymbol{\theta}^{1:L}) \right] - \sum_{l=1}^L \mathbb{E}_Q \left[\ln \frac{q(\boldsymbol{\theta}^l | \mathbf{d}, \boldsymbol{\Phi}^{l+1} \boldsymbol{\theta}^{l+1})}{p(\boldsymbol{\theta}^l | \boldsymbol{\Phi}^{l+1} \boldsymbol{\theta}^{l+1})} \right] \quad (12)$$

and the training is performed by maximizing $\mathbb{E}_{p_{\text{data}}(y_{1:T}, \mathbf{d})} [L(y_{1:T}, \mathbf{d})]$; following Zhang et al. (2018), we define $q(\boldsymbol{\theta}^l | \mathbf{d}, \boldsymbol{\Phi}^{l+1}, \boldsymbol{\theta}^{l+1}) = \text{Weibull}(\mathbf{k}^l + \boldsymbol{\Phi}^{l+1} \boldsymbol{\theta}^{l+1}, \boldsymbol{\lambda}^l)$, where both \mathbf{k}^l and $\boldsymbol{\lambda}^l$ are deterministically transformed from \mathbf{d} using neural networks. Distinct from a usual variational auto-encoder whose inference network has a pure bottom-up structure, the inference network here has a deterministic-upward–stochastic-downward ladder structure (Zhang et al., 2018).

B. TLASGR-MCMC for rGBN-RNN

To allow for scalable inference, we apply the TLASGR-MCMC algorithm (Cong et al., 2017a; Zhang et al., 2018; Guo et al., 2018), which can be used to sample simplex-constrained global parameters (Cong et al., 2017b) in a mini-batch based manner. It improves its sampling efficiency via the use of the Fisher information matrix (FIM) (Girolami & Calderhead, 2011), with adaptive step-sizes for the latent factors and transition matrices of different layers. In this section, we discuss how to update the global parameters $\{\boldsymbol{\Phi}^l, \boldsymbol{\Pi}^l\}_{l=1}^L$ of rGBN in detail and give a complete one in Algorithm in 1.

Sample the auxiliary counts: This step is about the “backward” and “upward” pass. Let us denote $Z_{\cdot, kj}^l = \sum_{k_l=1}^{K_l} Z_{k_l, kj}^l$, $Z_{\cdot, k, J+1}^l = 0$, and $x_{kj}^{(1,1)} = d_{vj}$, where $\mathbf{d}_j = \{d_{1j}, \dots, d_{vj}, \dots, d_{V_c j}\}$ is the same as in (2). Working backward for $j = J, \dots, 1$ and upward for $l = 1, \dots, L$, we draw

$$(A_{k_1 j}^l, \dots, A_{k_{K_l} j}^l) \sim \text{Multi} \left(x_{kj}^{(l,l)}; \frac{\phi_{k_1}^l \theta_{1j}^l}{\sum_{k_l=1}^{K_l} \phi_{k_l}^l \theta_{k_l j}^l}, \dots, \frac{\phi_{k_{K_l}}^l \theta_{K_l j}^l}{\sum_{k_l=1}^{K_l} \phi_{k_l}^l \theta_{k_l j}^l} \right), \quad (13)$$

$$x_{kj}^{l+1} \sim \text{CRT} \left[A_{k_j}^l + Z_{\cdot, k, j+1}^l, \tau_0 \left(\sum_{k_{l+1}=1}^{K_{l+1}} \phi_{k_{l+1}}^{l+1} \theta_{k_{l+1} j}^{l+1} + \sum_{k_l=1}^{K_l} \pi_{k k_l}^l \theta_{k_l, j-1}^l \right) \right]. \quad (14)$$

Note that via the deep structure, the latent counts x_{kj}^{l+1} will be influenced by the effects from both time $j-1$ at layer l and time j at layer $l+1$. With $p_1 := \sum_{k_l=1}^{K_l} \pi_{k k_l}^l \theta_{k_l, j-1}^l$ and $p_2 := \sum_{k_{l+1}=1}^{K_{l+1}} \phi_{k_{l+1}}^{l+1} \theta_{k_{l+1} j}^{l+1}$, we can sample the latent counts at

layer l and $l + 1$ by

$$(x_{kj}^{l+1,l}, x_{kj}^{l+1,l+1}) \sim \text{Multi} \left(x_{kj}^{l+1}, p_1/(p_1 + p_2), p_2/(p_1 + p_2) \right), \quad (15)$$

and then draw

$$(Z_{k1j}^l, \dots, Z_{kK_l j}^l) \sim \text{Multi} \left(x_{kj}^{l+1,l}, \frac{\pi_{k1}^l \theta_{1,j-1}^l}{\sum_{k_l=1}^{K_l} \pi_{kk_l}^l \theta_{k_l,j-1}^l}, \dots, \frac{\pi_{kK_l}^l \theta_{K_l,j-1}^l}{\sum_{k_l=1}^{K_l} \pi_{kk_l}^l \theta_{k_l,j-1}^l} \right). \quad (16)$$

In rGBN, the prior and the likelihood of $\{\Phi^l\}_{l=1}^L$ is very similar to $\{\Pi^l\}_{l=1}^L$, so we also apply the TLASGR-MCMC sampling algorithm on both of them conditioned on the auxiliary counts.

Sample the hierarchical components $\{\Phi^l\}_{l=1}^L$: For ϕ_k^l , the k th column of the loading matrix Φ^l of layer l , its sampling can be efficiently realized as

$$\begin{aligned} (\phi_k^l)_{n+1} = & \left[(\phi_k^l)_n + \frac{\varepsilon_n}{P_k^l} \left[\left(\rho \tilde{A}_{:k}^l + \eta_0^l \right) - \left(\rho \tilde{A}_{:k}^l + K_{l-1} \eta_0^l \right) (\phi_k^l)_n \right] \right. \\ & \left. + \mathcal{N} \left(0, \frac{2\varepsilon_n}{P_k^l} \left[\text{diag}(\phi_k^l)_n - (\phi_k^l)_n (\phi_k^l)_n^T \right] \right) \right]_{\angle}, \end{aligned} \quad (17)$$

where P_k^l is calculated using the estimated FIM, $\tilde{A}_{k_l j}^l = \sum_{j=1}^J A_{k_l k_l j}^l$, $\tilde{A}_{:k}^l = \{\tilde{A}_{1j}^l, \dots, \tilde{A}_{K_l j}^l\}^T$ and $\tilde{A}_{:k}^l = \sum_{k_l=1}^{K_l} \tilde{A}_{k_l j}^l$. $A_{k_l k_l j}^l$ comes from the augmented latent counts A^l in (13), η_0^l denote the prior of ϕ_k^l , and $[\cdot]_{\angle}$ denotes the simplex constraint.

Sample the transmission matrix $\{\Pi^l\}_{l=1}^L$: For π_k^l , the k th column of the transition matrix Π^l of layer l , its sampling can be efficiently realized as

$$\begin{aligned} (\pi_k^l)_{n+1} = & \left[(\pi_k^l)_n + \frac{\varepsilon_n}{M_k^l} \left[\left(\rho \tilde{Z}_{:k}^l + \eta_{:k}^l \right) - \left(\rho \tilde{Z}_{:k}^l + \eta_{:k}^l \right) (\pi_k^l)_n \right] \right. \\ & \left. + \mathcal{N} \left(0, \frac{2\varepsilon_n}{M_k^l} \left[\text{diag}(\pi_k^l)_n - (\pi_k^l)_n (\pi_k^l)_n^T \right] \right) \right]_{\angle}, \end{aligned} \quad (18)$$

where M_k^l is calculated using the estimated FIM, $\tilde{Z}_{k_l j}^l = \sum_{j=1}^J Z_{k_l k_l j}^l$, $\tilde{Z}_{:k}^l = \{\tilde{Z}_{1j}^l, \dots, \tilde{Z}_{K_l j}^l\}^T$ and $\tilde{Z}_{:k}^l = \sum_{k_l=1}^{K_l} \tilde{Z}_{k_l j}^l$. $Z_{k_l k_l j}^l$ comes from the augmented latent counts Z^l in (16), and $[\cdot]_{\angle}$ denotes a simplex constraint, and $\eta_{:k}^l$ denotes the prior of π_k^l , more details about TLASGR-MCMC for our proposed model can be found in Cong et al. (2017a).

C. Datasets

We consider three publicly available corpora¹. APNEWS is a collection of Associated Press news articles from 2009 to 2016, IMDB is a set of movie reviews collected by Maas et al. (2011), and BNC is the written portion of the British National Corpus (Consortium, 2007). Following the preprocessing steps in Lau et al. (2017), we tokenize words and sentences using Stanford CoreNLP (Klein & Manning, 2003), lowercase all word tokens, and filter out word tokens that occur less than 10 times. For the topic model, we additionally exclude stopwords² and the top 0.1% most frequent words. All these corpora are partitioned into training, validation, and testing sets, whose summary statistics are provided in Table 2 of the Appendix.

D. Complexity of rGBN-RNN

The proposed rGBN-RNN consists of both language model and topic model components. For the topic model component, there are the global parameters of rGBN (decoder), including $\{\Phi^l, \Pi^l\}_{l=1}^L$ in (2), and the parameters of the recurrent variational inference network (encoder), consisting of $\text{RNN}_{\text{sent}}^l$, f_k^l , and f_{λ}^l in (9). The language model component is parameterized by $\text{LSTM}_{\text{word}}^l$ in (3) and the coupling vectors g^l described in (4). We summarize in Table 3 the complexity of rGBN-RNN (ignoring all bias terms), where V denotes the vocabulary size of the language model, E the dimension

¹<https://ibm.ent.box.com/s/1s61p8ovc1y87w45oa02zink2z1716z4>

²We use Mallet's stopword list: <https://github.com/mimno/Mallet/tree/master/stoplists>

Table 2. Summary statistics for the datasets.

Dataset	Vocabulary		Training			Validation			Testing		
	LM	TM	Docs	Sents	Tokens	Docs	Sents	Tokens	Docs	Sents	Tokens
APNEWS	34231	32169	50K	0.8M	15M	2K	33K	0.6M	2K	32K	0.6M
IMDB	36009	34925	75K	1.1M	20M	12.5K	0.18M	0.3M	12.5K	0.18M	0.3M
BNC	43703	41552	15K	1M	18M	1K	57K	1M	1K	66K	1M

Table 3. Complexity of the three-layer rGBN-RNN.

Component	Language Model		Topic Model				
	$LSTM_{word}^l$ in (3)	g^l in (4)	Φ^l in (2)	Π^l in (2)	RNN_{sent}^l in (9)	f_k^l in (9)	f_λ^l in (9)
Layer1	$O(4 \times (E + H_1^w) \times H_1^w)$	$O(3 \times (K_1 + H_1^w) \times H_1^w)$	$O(V_c \times K_1)$	$O(K_1 \times K_1)$	$O((V_c + H_1^s) \times H_1^s)$	$O(H_1^s)$	$O(K_1 \times H_1^s)$
Layer2	$O(4 \times (H_1^w + H_2^w) \times H_2^w)$	$O(3 \times (K_2 + H_2^w) \times H_2^w)$	$O(K_1 \times K_2)$	$O(K_2 \times K_2)$	$O((H_1^s + H_2^s) \times H_2^s)$	$O(H_2^s)$	$O(K_2 \times H_2^s)$
Layer3	$O(4 \times (H_2^w + H_3^w) \times H_3^w)$	$O(3 \times (K_3 + H_3^w) \times H_3^w)$	$O(K_2 \times K_3)$	$O(K_3 \times K_3)$	$O((H_2^s + H_3^s) \times H_3^s)$	$O(H_3^s)$	$O(K_3 \times H_3^s)$

of word embedding vectors, V_c the size of the vocabulary of the topic model that excludes stop words, H_l^w the number of hidden units of the word-level LSTM at layer l (stacked-RNN language model), H_l^s the number of hidden units of the sentence-level RNN at layer l (recurrent variational inference network), and K_l the number of topics at layer l . Comparison of the number of parameters between various language models is provided in Table 1.

E. BLEU scores for IMDB

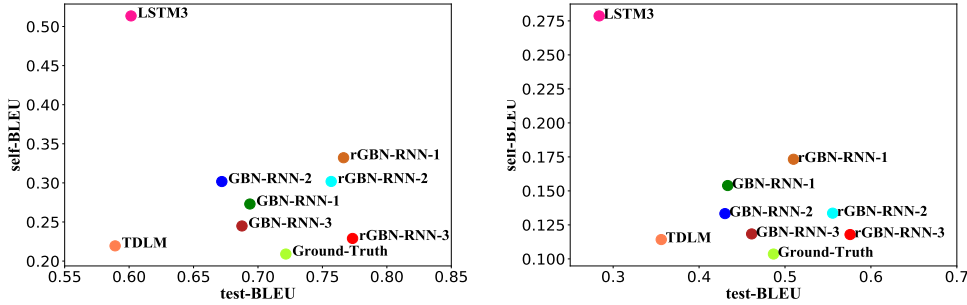


Figure 7. BLEU scores of different methods for IMDB. x-axis denotes test-BLEU, and y-axis self-BLEU. Left panel is BLEU-3 and right is BLEU-4, and a better BLEU score would fall within the lower right corner, where black point represents mean value and circles with different colors denote the elliptical surface of probability of BLEU in a two-dimensional space.

F. BLEU scores for APNEWS

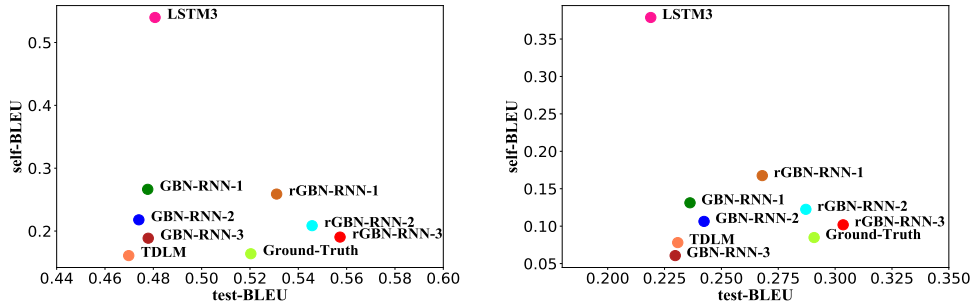


Figure 8. BLEU scores of different methods for APNEWS. x-axis denotes test-BLEU, and y-axis self-BLEU. Left panel is BLEU-3 and right is BLEU-4, and a better BLEU score would fall within the lower right corner, where black point represents mean value and circles with different colors denote the elliptical surface of probability of BLEU in a two-dimensional space.

G. Additional experimental results on IMDB and BNC

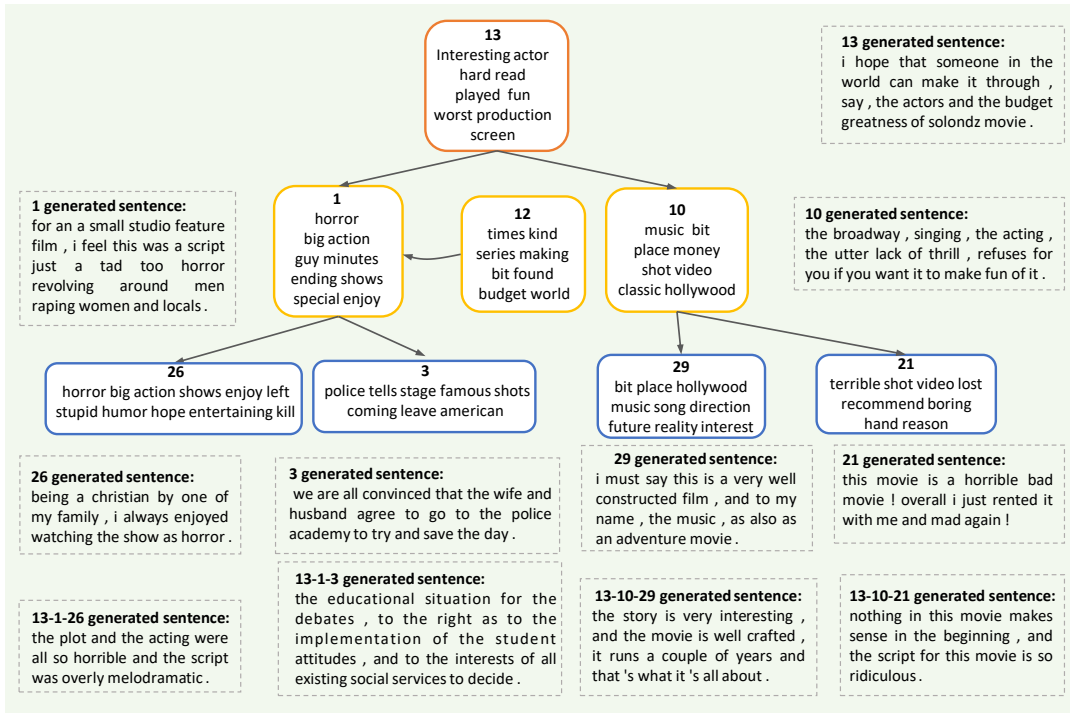


Figure 9. Analogous plot to Fig. 5 for the IMDB corpus.

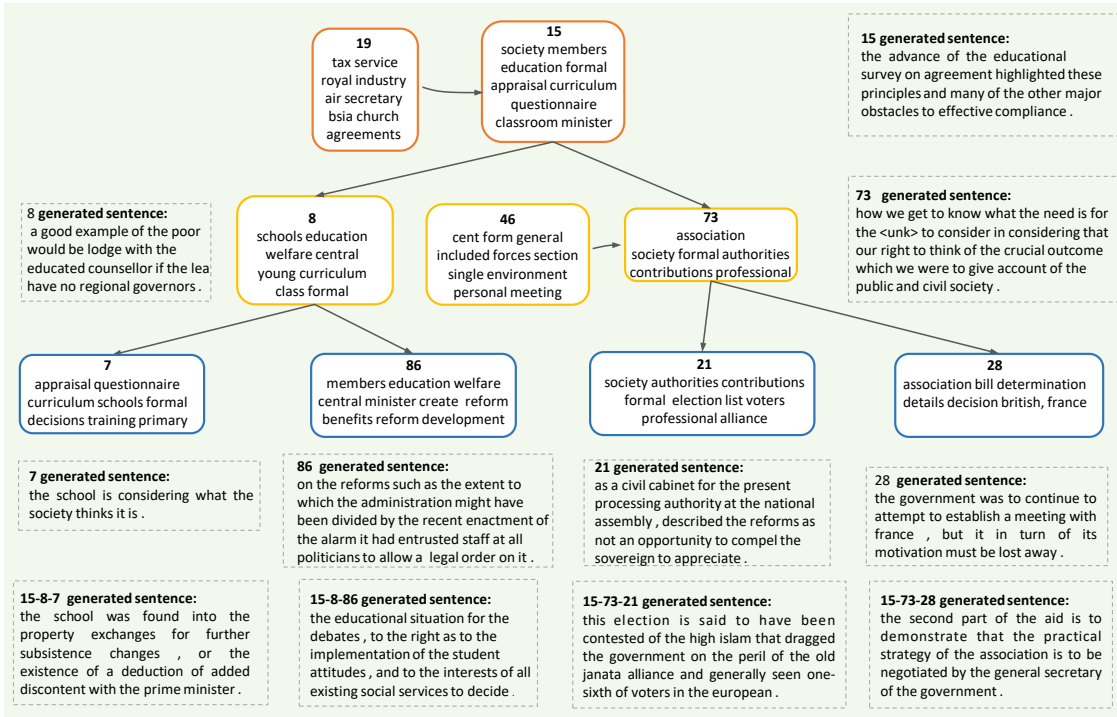


Figure 10. Analogous plot to Fig. 5 for the BNC corpus.

H. Additional example topic hierarchies and generated sentences.

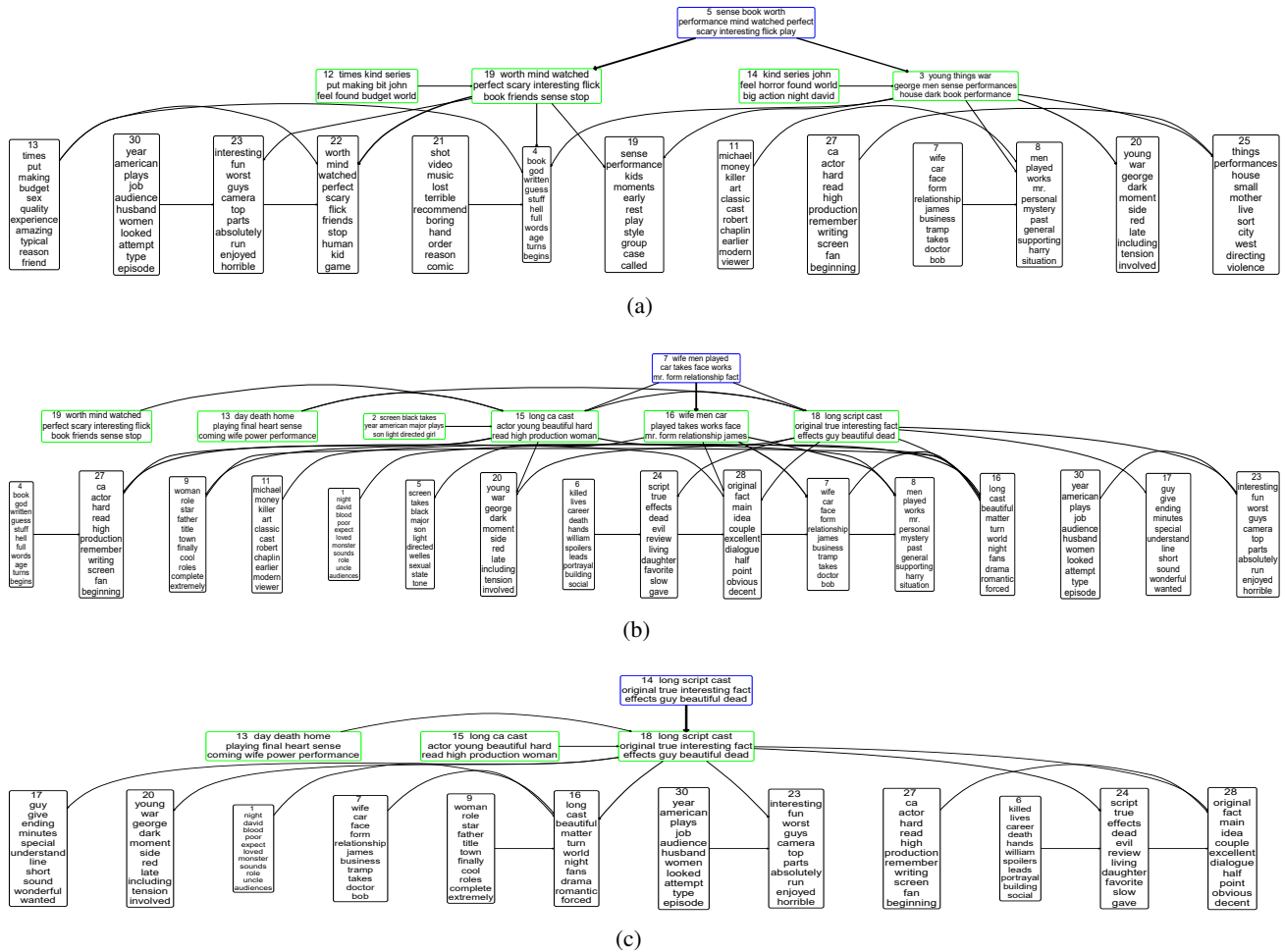


Figure 11. Example topics and their hierarchical and temporal connections inferred by a three-hidden-layer rGBN-RNN from the IMDB corpus. Top words of each topic at layers 3, 2, and 1 are shown in blue, green, and black boxes respectively. Shown in (a)-(c) are the 5th, 7th and 14th nodes of the top layer, respectively.

Generated sentences conditioned on topic 5 at layer 3: (a) *i love this movie , i strongly recommend it , just watch it with friends and laugh .* (b) *i was seriously shocked with hogan 's performance.* (c) *the performances are terrible , the storyline is non-existing , the directing (if you can call it that) is horrible , and the acting is horrible .*

Generated sentences conditioned on topic 3 at layer 2: (a) *her performance is a bit afro looking and they have a very neat southern accent and the movie was well cast as in previous movies , was so much more beautiful especially when woody allen made this film .* (b) *without their personal history or the fact that he would like to kill his american adoptive parents in their own homes , it all wo n't make sense .* (c) *this is one of those romantic comedies where we have some good action and good acting .*

Generated sentences conditioned on topic 20 at layer 1: (a) *the new story was very touching , real , a new perspective of what it is like to be a back to war .* (b) *the movie is ok in my eyes , i know you will find it too scary ... but i kind of got a good movie from a somewhat dark sense .* (c) *at the same time i believe that the film was made just a couple of years earlier , the members of the u.s. government were almost always bad .*

Generated sentences conditioned on a combination of topics 5, 19 and 22 at layer 3, 2, 1: (a) *the show is well worth watching , i feel it is the best movie i have ever seen .* (b) *the movie is amazing , i thought the acting was great , and the movie was well worth the rental .* (c) *it would have been much better if it was a science fiction movie with a little more humor and some more action .*

Recurrent Hierarchical Topic-Guided RNN for Language Generation

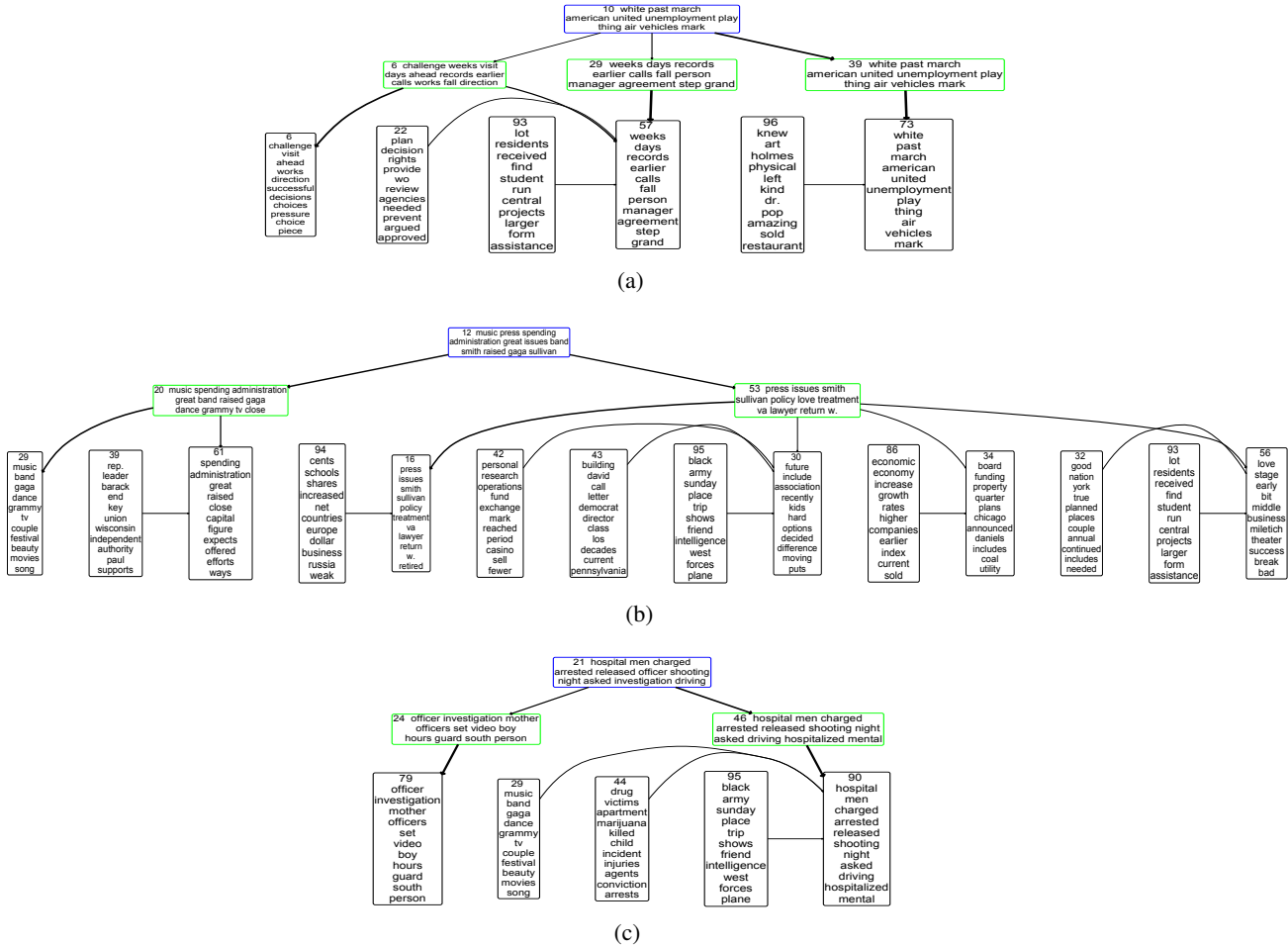


Figure 12. Example topics and their hierarchical and temporal connections inferred by a three-hidden-layer rGBN-RNN from the APNEWS corpus. Top words of each topic at layers 3, 2, and 1 are shown in blue, green, and black boxes respectively. Shown in (a)-(c) are the 10th, 12th and 21st nodes of the top layer, respectively.

Generated sentences conditioned on topic 12 at layer 3: (a) *they 're planning to attend a concert hall held by the rev. jesse.* (b) *approved by the standard free press , it will generate water for their own offices .* (c) *the christie administration will not give him an opinion if the of the state has issued its name .*

Generated sentences conditioned on topic 53 at layer 2: (a) *. the national park service said the maine department of law will hold a agreement on the <unk> law for the first time .* (b) *the detroit free press reports the city asked former winston-salem public schools commission chairman <unk> <unk> to take the seat .* (c) *earlier this month , the state police issued several orders to the fbi and send a <unk> team to the sheriff 's office .*

Generated sentences conditioned on topic 29 at layer 1: (a) *but it was and a few months later , the music had performed at the <unk> theatre in its <unk> .* (b) *the festival draws hundreds of thousands of viewers , a tourist year by a member of the oxford state team .* (c) *the university made the first " the most exciting , very beautiful " album followed by the 1996 " the sky " includes a <unk> version .*

Generated sentences conditioned on a combination of topics 21, 46 and 44 at layer 3, 2, 1: (a) *police say the suspect was taken to a hospital for treatment .* (b) *the man was arrested after police say he was driving in a car in north mississippi , which was the first <unk> to be used in the shootout .* (c) *police said wednesday that the victims ' deaths are not believed to be gang affiliation .*

Recurrent Hierarchical Topic-Guided RNN for Language Generation

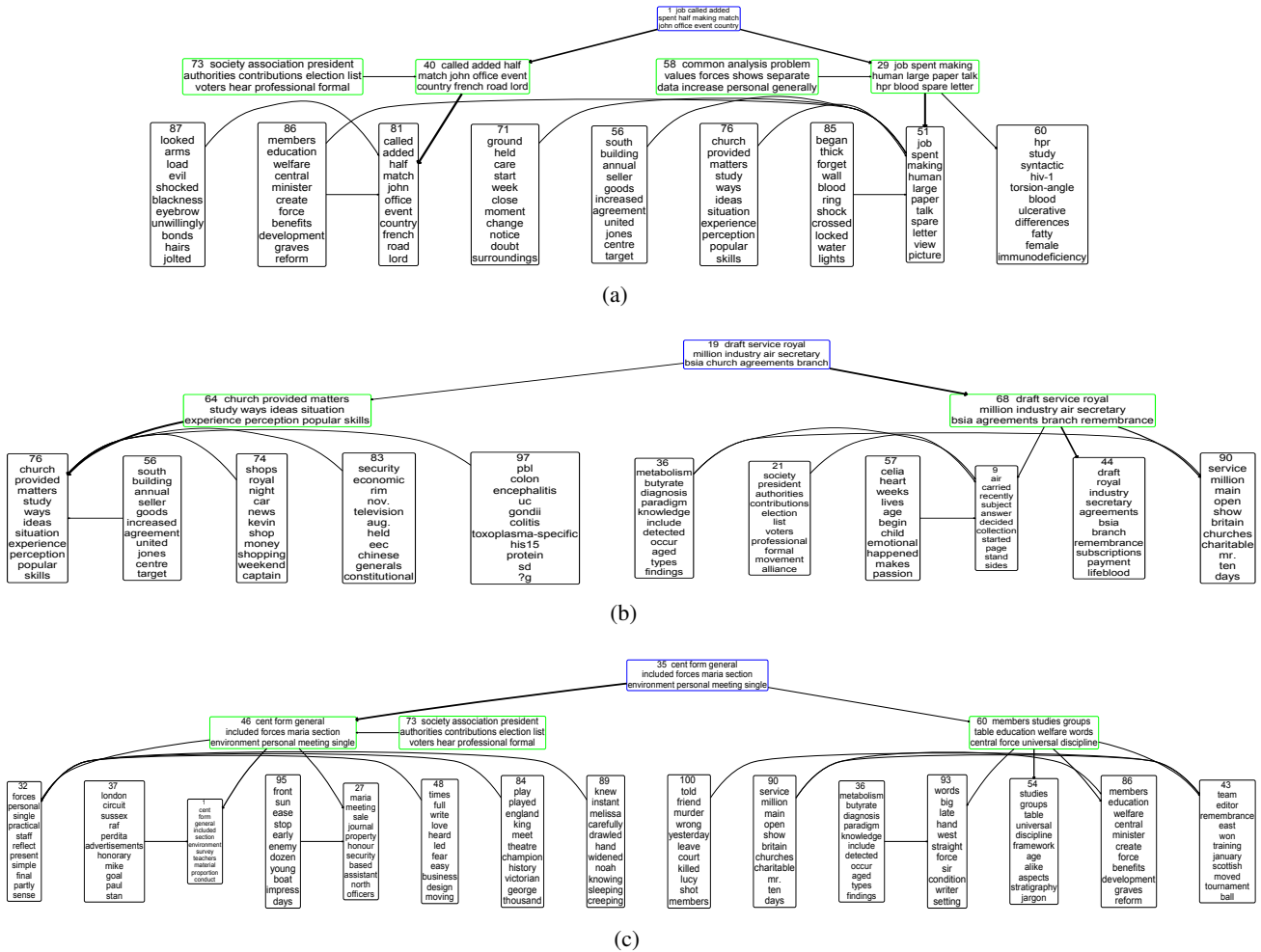


Figure 13. Example topics and their hierarchical and temporal connections inferred by a three-hidden-layer rGBN-RNN from the BNC corpus. Top words of each topic at layers 3, 2, and 1 are shown in blue, green, and black boxes respectively. Shown in (a)-(c) are the 15th, 19th and 35th nodes of the top layer, respectively.

Generated sentences conditioned on topic 1 at layer 3: (a) *the fourth should be in the obligation of the enforcement officer for proof where a supplementary liability of pension funds has not been advocated.* (b) *approved by the standard free press , it will generate water for their own offices .* (c) *the court , has recently agreed to participate in the investigation to allow the justices to succeed to be responsible for the full remit of the submissions .*

Generated sentences conditioned on topic 73 at layer 2: (a) *another , which was the period of message from a federal of the new presidential opposition to the new constitution , was to change his autonomy rather than through the different strategies .* (b) *the president is to bring out the best of all and the most widely understood state of affairs in the country .* (c) *the icrc has announced that the <unk> could not be accepted on the factors outlined in the societies ' choice .*

Generated sentences conditioned on topic 76 at layer 1: (a) *the church of st clement danes – the great continent , the southern of the realm , a treaty give such assistance to brother.* (b) *the legality of the political instability that followed by the profession has been criticised for the necessity for the study of individuals and friends .* (c) *the character of the monarchy is dominated by a panoramic style which includes and attempts to limit the genre to his/her ideas . '*

Generated sentences conditioned on a combination of topics 19, 68 and 44 at layer 3, 2, 1: (a) *again the royal air force in the middle of the war was now part of the struggle by the indian resistance .* (b) *the mailing list for another example of the british aerospace industry shows a, exclusive catalogue to enable object to be changed to steam .* (c) *the great britain will continue to submit to the thinking and nature of the reciprocal international economic agreement.*

I. Additional examples of generated sentences / paragraph conditioning on a paragraph.

<p>Document</p> <ul style="list-style-type: none"> ● splitting up the disintegration of any relationship is painful . but the problems which follow may not only be emotional ones , and this is where a solicitor can help . divorce or separation mean thinking about arrangements for children ; maintenance ; and the division of property .these are all areas in which a solicitor offers professional expertise and practical advice -- as well as support and comfort . an unmarried couple splitting up can also face legal problems . if they did not take legal advice before buying property together , its division could be complicated . there may also be difficulties about parental responsibility for their children . a solicitor will be able to advise on all these matters .
<p>Generated Sentences with GBN-RNN</p> <ul style="list-style-type: none"> ● divorce which can not be completion of couple, or on behalf of the disaster . ● despite divorced , and also to the right to rely on anything at the magistrates ' courts .
<p>Generated Sentences with rGBN-RNN</p> <ul style="list-style-type: none"> ● children 's divorced parents were to be allowed to provide the best for that to happen . ● parents to protect themselves from the existing community and to a mutual trust in the home in the child 's mind.
<p>Generated temporal Sentences with rGBN-RNN (Paragraph)</p> <ul style="list-style-type: none"> ● parents or fathers were denied to care for and respect for both the self-employed and his wife . for the first time , since the fact that many parents and children were earning their lives during the war , the health and safety act in the <unk> household . for this reason , a mutual relationship to voluntary arrangements for which the rights of its members are subject to : compensation ! parents in the <unk> should be allowed to test the impact of liability on the community life fund . Both the parents and housing , should be able to provide good teaching and care for any elderly parent community . the children live in australia , organised by social services and housing associations . divorce and remarriage are just as important as was due to the ill and the whole of the damage or the reservation . it is to be hoped that , more than that , it must be necessary to discuss the analyses of family needs .

Figure 14. An example of generated sentences and paragraph conditioned on a document from BNC (green denotes novel words, blue the key words in document and generated sentences.)

<p>Document</p> <ul style="list-style-type: none"> ● this is one of the best and moodiest vampire tales ever ! i love this movie really . the character are great , even the locations and the story . indeed the picture is n't a big budget production , but it is absolutely worth seeing . ok there are some faults (especially the names of the castle and the locations) in this movie , but such mistakes are typically and are almost in every horror movie . the scenery fits perfect to the story and is close to reality , i can say that honest , because i visit them once when i was in romania in my vacations . in my opinion this is the best part of the subspecies series .
<p>Generated Sentences with GBN-RNN</p> <ul style="list-style-type: none"> ● it was one of the most powerful films i 've ever seen . ● okay , i would say special up about it , but if you were a true horror fanatic , you wo n't be disappointed .
<p>Generated Sentences with rGBN-RNN</p> <ul style="list-style-type: none"> ● this is one of the more refreshingly horrible films you will treasure , great music and even the scenes. ● this is a wonderful movie, worth to look in and the plot is very gold and some of the characters are true to their characters .
<p>Generated temporal Sentences with rGBN-RNN (Paragraph)</p> <ul style="list-style-type: none"> ● i 'm a fan of vampire or zombie movies and i was never bored. this film is well worth a look for years over. the effect and scale are enough to refresh the horror movie. the story between vampire and werewolf runs through the story, and vampire wants to become the werewolf . i am very glad to see that the vampires in this movie have the hunger to kill the zombies. what i like about this film is the cast , their performances , the plausibility , and its relationship. the ending of about vampire is very beautiful and i guess that is all that would make the movie so great for me .

Figure 15. An example of generated sentences and paragraph conditioned on a document from IMDB (green denotes novel words, blue the key words in document and generated sentences.)