A REINFORCEd Variational Autoencoder Topic Model

Amit Kumar[1,2], Nazanin Esmaili[1], and Massimo Piccardi[1]

[1] University of Technology Sydney, Broadway, NSW, 2007, Australia
[2] Food Agility CRC Ltd, 81 Broadway, Ultimo, 2007, Australia
amit270980@gmail.com,{nazanin.esmaili,massimo.piccardi}@uts.edu.au ⋆

Abstract. Topic modeling is an unsupervised natural language processing approach for automatically extracting the main topics from a large collection of documents, and simultaneously assigning the individual documents to the extracted topics. While many algorithms for topic modelling have been proposed in the literature, to date there has been little use of the popular reinforcement learning framework for this task. For this reason, in this paper we leverage two pillars of reinforcement learning – the policy gradient theorem and the REINFORCE algorithm – to define a novel loss function for training topic models. In the paper, the loss function is applied to a state-of-the-art topic model based on a variational autoencoder. Experimental results on two social media datasets have shown that the proposed approach has been able to outperform the original variational autoencoder and other baselines in terms of evaluation measures such as model perplexity and topic coherence.

Keywords: Topic models · deep neural networks · variational autoencoders · reinforcement learning · REINFORCE.

1 Introduction and Related Work

The continued growth of digital data sources, and especially social media, has led to an unprecedented rise in the volume of available text documents. This presents a major challenge for the systematic analysis of their contents, together with their management and organisation. While until the recent past these tasks could be undertaken based on human annotation, nowadays there is a compelling need for computational tools that can automatically extract topics and patterns from document collections and organise them accordingly.

In recent years, topic models have emerged as a powerful, unsupervised tool for identifying useful structure in such vast amounts of unstructured text data. In technical terms, a topic model is an algorithm that can efficiently discover the main topics of a potentially large corpus of documents, and assign the individual documents to the topics. A "topic" is commonly intended as a characteristic

⋆ Supported by funding from Food Agility CRC Ltd, funded under the Commonwealth Government CRC Program. The CRC Program supports industry-led collaborations between industry, researchers and the community.

probability distribution over the words of a vocabulary. For example, a topic like "computers" can be described by a probability distribution where words such as "motherboard," "CPU", "monitor," "mouse and the like have the highest probabilities. In turn, individual documents can be assigned to multiple topics in specific proportions. Topic models have proved useful for the analysis of a variety of data, from scientific publications to user posts on social media [1].

Many topic models have been proposed over the years, primarily based on techniques such as non-negative matrix factorization and variational inference. Latent semantic indexing (LSI) is generally regarded as the first "proper" topic model [4]. However, the most widespread topic model is likely the latent Dirichlet allocation (LDA) [2]. LDA's basic components are: 1) the word distributions of each topic, and 2) the topic proportions of each document. Since both are modeled as multinomial distributions, LDA conveniently uses an eponymous Dirichlet distribution as their prior. The conjugacy between the multinomial and the Dirichlet makes it easy to derive the posteriors and support inference (more details are provided in the following section). In addition, many LDA derivatives have been proposed over time, including, among others, sparse [3], sequential [5], and hierarchical [12] versions.

Recently, neural topic models have started to appear in the literature, joining the benefits of traditional models such as LDA with those of *deep generative models* [6, 7, 13, 15, 18]. Some neural topic models have made use of generative adversarial networks (GANs) [6, 7] and convolutional neural networks (CNNs) [18]. However, the most effective neural topic models seem to be those based on *variational autoencoders* (VAEs) [13,15]. Miao et al. in [13]. have proposed a VAE based neural topic model using the logistic normal distribution and the stick-breaking construction to infer the topic proportions. More recently, Srivastava and Sutton in [15] have proposed a neural topic model integrating LDA with a variational autoencoder, establishing state-of-the-art performance on all the tested datasets.

Despite the many available models, to date topic modeling has made limited use of the popular *reinforcement learning* framework [16]. Reinforcement learning offers the potential to leverage both differentiable and non-differentiable "rewards" to guide the extraction of the topics. An example of topic modeling with reinforcement learning has been presented in [8], leveraging word-reweighting rewards to encourage within-topic coherence and between-topic separation. However, we are not aware of any model that has used reinforcement learning to learn an effective *policy* over the topics. For this reason, in this paper we propose a topic model that uses the policy gradient theorem and the REINFORCE algorithm [17] to improve learning of an effective topic model. Experiments performed over two challenging datasets (20 Newsgroups and Amazon Fine Food Reviews, both collected from social media) have shown that the proposed approach has achieved a better performance than all the compared approaches in terms of topic coherence and model perplexity in a large majority of cases.

2 Topic Modeling with Variational Autoencoders

In recent years, deep generative models have gained widespread adoption in the deep learning community, thanks to their effective integration of features of generative models, Bayesian inference and deep neural networks. In particular, variational autoencoders (VAEs) have proven specially effective at learning representations for latent variables [10], making them appealing for topic modeling.

A VAE is basically a generalized version of an autoencoder, which is a neural network subdivided into an encoder and a decoder. The encoder takes in input a multidimensional measurement, and produces a latent representation in output. In turn, the decoder takes in input the latent representation and produces a "reconstruction" of the original measurement. In the case of a VAE, the reconstruction is simply meant as the probability of the measurement in the parametrized decoder. When VAEs are used for topic modeling, the measurement in input is a document representation, w (typically, a bag-of-words or a TF-IDF vector), while the latent variable is its topic vector, θ. In turn, the likelihood of the document representation, w, can be obtained by marginalizing the topic vector, θ, as in:

$$p(w|\alpha, \beta) = \int_{\theta} p(w, \theta|\alpha, \beta) d\theta \tag{1}$$

where α is the parameter of the prior probability over the topics, β is the matrix of the word distributions for all the topics, and $p(w, \theta|\alpha, \beta)$ is the joint probability of the document representation and the topic vector.

The training of a VAE aims to maximize (1) over the given document collection. However, this is typically impossible to perform directly. Therefore, the VAE sets to maximize a tractable lower bound (the evidence lower bound, or ELBO) [10]:

$$\mathcal{L}(w|\alpha, \beta) = \mathbb{E}_{q(\theta|w)}\big[\log p(w|\theta, \beta)\big] - D_{\mathrm{KL}}(q(\theta|w) \| p(\theta|\alpha)) \tag{2}$$

Hereafter, we briefly describe the meaning of the terms in (2); further details can be found in [10]. Term $q(\theta|w)$ (the "encoder") estimates the probability of the topic vector for the given document. Term $\log p(w|\theta, \beta)$ (the "decoder") is the log-probability of the document given its topic vector and the word distributions; its expectation over $q(\theta|w)$, $\mathbb{E}_{q(\theta|w)}\big[\log p(w|\theta, \beta)\big]$, is the "reconstruction term". Finally, term $p(\theta|\alpha)$ is a trainable prior over the topic vectors. During training, (2) trades off increasing the reconstruction term against reducing the Kullback-Leibler divergence (D_{KL}) between the encoder and the prior.

To facilitate the reparametrization of the encoder and the prior, Srivastava and Sutton in [15] have proposed replacing the usual Dirichlet distribution with a logistic normal distribution. Samples of a logistic normal distribution, $\mathcal{LN}(\mu, \Sigma)$, can be conveniently obtained by applying the softmax operator to samples of a Gaussian distribution of equal parameters, $\mathcal{N}(\mu, \Sigma)$. In turn, the Gaussian distribution can be reparametrized with the common inverse transform approach. Srivastava and Sutton's model, called *AVITM* (from autoencoding variational inference for topic models), models the prior as $p(\theta|\alpha) = \mathcal{LN}(\theta|\mu(\alpha), \Sigma(\alpha))$, where $\mu(\alpha)$ and $\Sigma(\alpha)$ are closed-form expressions for the mean and the variance obtained with a Laplace approximation [9]. In turn, the encoder is modeled as $q(\theta|w) = \mathcal{LN}(\theta|\mu(w, \phi_1), \Sigma(w, \phi_2))$, where ϕ_1 and ϕ_2 are the parameters of two feed-forward neural networks that infer, respectively, the mean and covariance of the encoder. Finally, the decoder is given by:

$$p(w|\theta, \beta) = \mathrm{Mult}(w \mid \mathrm{softmax}(\beta)\,\theta) \tag{3}$$

where Mult() denotes the multinomial distribution, and the word distributions are parametrized as logits rather than probabilities to bypass the simplex constraint during gradient descent. A second version of the decoder, inspired by products-of-experts and nicknamed *ProdLDA*, first computes the product, and then the softmax:

$$p(w|\theta, \beta) = \mathrm{Mult}(w \mid \mathrm{softmax}(\beta\,\theta)). \tag{4}$$

3 The Proposed Approach: a VAE Topic Model with REINFORCE

Reinforcement learning has become increasingly popular in recent years thanks to its ability to train models beyond conventional maximum-likelihood approaches. The main advantages of reinforcement learning are its ability to minimize non-differentiable training objectives and its use of sampling, which permits a certain degree of *exploration* in the parameter space. In the case of our model, the loss function in (2) is an expectation over θ, the topic vector for the document, and should therefore not depend on it. However, since the expectation is empirical and based on typically only one sample per document, some dependence on θ persists, and we emphasize it by noting the loss as $\mathcal{L}(\theta)$ in the following. To improve the estimate of the encoder distribution, $q(\theta|w)$, we choose to minimize the *predictive risk*:

$$\mathcal{R} = \mathbb{E}_{q(\theta|w)}\big[\mathcal{L}(\theta)\big] = \int_{\theta} \mathcal{L}(\theta) q(\theta|w) d\theta \tag{5}$$

which is the expectation of the loss function, $\mathcal{L}(\theta)$, over the probability of variable θ, the document's topic vector. In order to minimize (5), training will attempt to assign high probability to values of θ that cause low values of the loss, and the vice versa, thus promoting an effective encoder. The minimization of (5) can be performed using the policy gradient theorem [17], which ignores the indirect dependence of the loss on the model's parameters and only differentiates the probability distribution in its own parameters, ϕ:

$$\begin{aligned} \frac{\partial}{\partial \phi}\mathcal{R} &= \int_\theta \mathcal{L}(\theta)\frac{\partial}{\partial \phi} q(\theta|w)d\theta \\ &= \int_\theta \mathcal{L}(\theta)\frac{\partial}{\partial \phi}\log q(\theta|w) q(\theta|w)d\theta \\ &= \mathbb{E}_{q(\theta|w)}\big[\mathcal{L}(\theta)\frac{\partial}{\partial \phi}\log q(\theta|w)\big] \end{aligned} \tag{6}$$

As common in practice, we compute the resulting expectation empirically from a single sample:

$$\frac{\partial}{\partial \phi}\mathcal{R} \approx \mathcal{L}(\theta)\frac{\partial}{\partial \phi}\log q(\theta|w), \quad \theta \sim q(\theta|w) \tag{7}$$

The above estimator of the gradient of the predictive risk is the popular REINFORCE, a fundamental approach of reinforcement learning which has been applied successfully in many fields [17]. However, the REINFORCE estimator typically suffers from high variance, often affecting the stability of training. This issue can be mollified by subtracting a baseline, b, from the loss (an approach known as REINFORCE *with baseline*):

$$\frac{\partial}{\partial \phi}\mathcal{R} \approx \big(\mathcal{L}(\theta) - b\big)\frac{\partial}{\partial \phi}\log q(\theta|w), \theta \sim q(\theta|w) \tag{8}$$

With this modification, a training iteration will decrease $q(\theta|w)$ only if the loss, $\mathcal{L}(\theta)$, is greater than b (i.e., a remarkably bad value). Otherwise, it will increase it or leave it unchanged. In addition, from the gradient estimator we can derive an expression for a loss that can be automatically differentiated by common autodiff tools[3]:

$$\mathcal{L}_{REINF} = \big(\mathcal{L}(\theta) - b\big)_{nograd}\log q(\theta|w) \tag{9}$$

where subscript *nograd* prevents differentiating the subscripted term.

The VAE loss (2) and the REINFORCE loss (9) can also be conveniently mixed, to explore trade-offs between the two. We therefore define the overall loss as:

$$\mathcal{L}_{overall} = \mathcal{L}(w|\alpha, \beta) + \epsilon\,\mathcal{L}_{REINF} \tag{10}$$

4 Experiments and Results

The experiments have been carried out over two probing datasets, *20 Newsgroups* (a benchmark for the field) and *Amazon Fine Food Reviews*. The 20 Newsgroups dataset comprises 18,846 documents from news shared on social media, while Amazon Fine Food Reviews consists of 568,454 user-posted food reviews. These datasets are very challenging because of their great variety of

[3] http://www.autodiff.org/, https://www.tensorflow.org/guide/autodiff.

topics and their utmost diversity of authors. As models, we have compared the proposed approach against two strong baselines (LDA and LSI) and the state-of-the-art topic model of Srivastava and Sutton, in its two versions AVITM and ProdLDA. For this reason, we present the results for the corresponding versions of our model, AVITM-REINF and ProdLDA-REINF. As hyperparameters, for those shared with the model of Srivastava and Sutton we have used the same values. For the loss balance parameter, ϵ, we have carried out a preliminary evaluation and chosen $\epsilon = 10^{-15}$ since the scale of $\mathcal{L}_{REINF}$ is much larger. To set the baseline, b, we have first trained the models without the REINFORCE loss and recorded the value of their loss at convergence, noted as l; then, we have set b in the range $[l, l \pm 25, l \pm 50]$, using only the training set for the selection. As a number of topics to explore, we have used the oft-used values of 20 and 50. For performance evaluation, we have adopted two popular measures, the *perplexity* and the *topic coherence*. The perplexity measures how poorly the model fits a given set of data (NB: lower values are better); to assess the models' ability to generalize, we have measured it over the test sets. The topic coherence measures the internal "coherence" of the extracted topics (NB: higher values are better). Since coherence can be quantified in different ways, we report both the *normalized pointwise mutual information* (`coher-NMPI`) [11] and the C_V *coherence* (`coher-Cv`) [14]. Unlike the perplexity, the coherence is computed over the training set itself to ensure that all of the topics' M most-frequent words are present in the set. In all the experiments, M has been set to 10. Given the significantly different nature of the perplexity and the topic coherence, some disagreement in their ranking of the models is to be expected.

4.1 Results

Tables 1 and 2 show the experimental results for the 20 Newsgroups dataset for 20 and 50 topics, respectively. Due to the different architecture and amount of degrees of freedom, the perplexity values for LDA cannot be directly compared to those of the autoencoder models; for this reason, we display them in italics. At its turn, LSI is not a probabilistic model and the perplexity values are not defined. When compared to the variational autoencoder approaches in terms of coherence, both LDA and LSI have reported significantly worse results and cannot be considered competitive. AVITM has achieved better perplexity values than ProdLDA, but ProdLDA has achieved higher coherence values in most cases, so there is no clear winner between them. However, both our proposed variants have been able to gain improvements over AVITM and ProdLDA, respectively: compared to AVITM, AVITM-REINF has achieved better perplexity and coherence in the case of 20 topics, and coherence in the case of 50 topics; compared to ProdLDA, ProdLDA-REINF has achieved better perplexity as well as coherence in the case of 20 topics, and coherence in the case of 50 topics. Overall, AVITM-REINF has achieved the best perplexity of all compared models, and ProdLDA-REINF the best coherence.

Tables 3 and 4 show the results for the Amazon Fine Food Reviews dataset with 20 and 50 topics, respectively. Again, LDA and LSI have reported sig-

Table 1. Results on the 20 Newsgroups dataset with 20 topics.

Metrics	LDA	LSI	AVITM	ProdLDA	AVITM-REINF	ProdLDA-REINF
Perplexity	*1480.3*	—	1140.2	1173.3	**1137.8**	1167.8
Coher-NPMI	-0.033	-0.053	0.094	0.141	0.131	**0.153**
Coher-Cv	0.309	0.371	0.671	0.779	0.734	**0.786**

Table 2. Results on the 20 Newsgroups dataset with 50 topics.

Metrics	LDA	LSI	AVITM	ProdLDA	AVITM-REINF	ProdLDA-REINF
Perplexity	*2389.6*	—	1133.1	1159.9	**1132.1**	1162.8
Coher-NPMI	-2.346	-0.062	0.117	0.111	0.115	**0.141**
Coher-Cv	-0.053	0.294	0.704	0.751	0.699	**0.763**

Table 3. Results on the Amazon Fine Food Reviews dataset with 20 topics.

Metrics	LDA	LSI	AVITM	ProdLDA	AVITM-REINF	ProdLDA-REINF
Perplexity	*1480.3*	—	**1000.9**	1099.7	1137.8	1091.4
Coher-NPMI	0.047	0.004	**0.144**	0.066	0.131	0.105
Coher-Cv	0.493	0.395	0.707	0.651	**0.734**	0.676

Table 4. Results on the Amazon Fine Food Reviews dataset with 50 topics.

Metrics	LDA	LSI	AVITM	ProdLDA	AVITM-REINF	ProdLDA-REINF
Perplexity	*2697.3*	—	1008.6	1012.5	**1008.3**	1009.0
Coher-NPMI	0.033	-0.008	0.144	-0.048	**0.155**	0.036
Coher-Cv	0.470	0.359	0.682	0.430	**0.699**	0.588

nificantly lower coherence values than all the autoencoder models and cannot be regarded as competitive. For this dataset, AVITM has neatly outperformed ProdLDA in both perplexity and coherence. At its turn, our proposed AVITM-REINF has outperformed AVITM in 4 out of 6 measures across 20 and 50 topics, and should be deemed as the best performing model for this dataset. In addition, ProdLDA-REINF has improved in all measures compared to the original ProdLDA. Overall, we can conclude that our REINFORCE-based models have led to marked improvements over both datasets.

As further analysis, we have explored the sensitivity of the topic coherence to the value of the baseline, b, using the test set to simultaneously probe generalization. To this aim, Figure 1 plots the values of the C_V coherence for ProdLDA-REINF (20 Newsgroups, 50 topics) over the range of the baseline values. The coherence value for ProdLDA is also displayed for comparison. In this experiment, the loss at convergence without REINFORCE has been $l = 630$, and the best coherence value over the training set has been obtained for $b = l - 25 = 605$. Figure 1 shows that this has also been the best value for the test set, showing excellent generalization. In addition, ProdLDA-REINF has achieved better coherence values than ProdLDA for all values of the baseline.

Finally, for a qualitative analysis of the results, Table 5 displays a few examples of topics extracted from the 20 Newsgroups dataset. The first topic extracted by LDA is clearly meaningful, but the other two (highlighted in red) seem incoherent. The third topic extracted by AVITM also seems, at least, un-

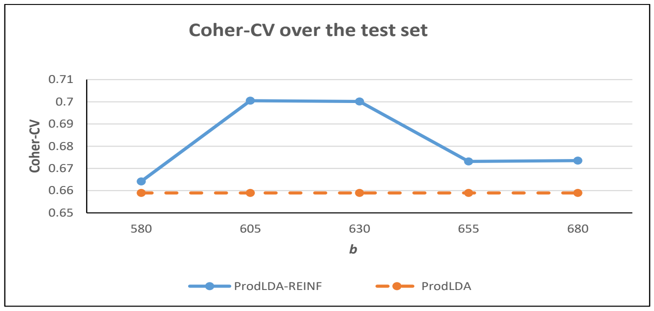

Fig. 1. Comparison of `coher-CV` on the test data for ProdLDA and ProdLDA-REINF (20 Newsgroups, 50 topics) by varying the baseline, b.

informative. Conversely, all the examples of topics extracted by AVITM-REINF seem consistent and properly descriptive.

Table 5. Topics discovered from the 20 Newsgroups dataset (50 topics). Seemingly incoherent topics are highlighted in red.

LDA:
monitor keyboard event appl mac usa ibm adapt use multi
date paper star robert confer divis surface mean june present
know say dont week white go your think year that
AVITM:
car bike ride honda bmw gear motorcycle rear dod ford
game team baseball player pitcher braves hitter score pitch fan
sea newspaper mountain april ii times angeles york francisco cambridge
AVITM-REINF:
windows microsoft memory setup mode modem nt port video vga
clinton congress economic government bush country administration economy american billion
laboratory nasa shuttle lab space engineering flight institute solar spacecraft

5 Conclusion

This paper has presented a novel training loss function for VAE topic models based on the reinforcement learning framework. In the proposed approach, we leverage the predictive risk and the REINFORCE algorithm to learn an effective policy over the topic vectors. The experimental results over two social media datasets have shown that the proposed approach has been able to attain a strong performance as measured by perplexity and topic coherence, with improvements of up to 2.4 percentage points in NPMI coherence and 2.7 percentage points in C_V coherence compared to the runner-up. In addition, the model has given evidence of good generalization over new documents. In the near future, we plan to explore other architectures for the implementation of the model's neural networks, possibly including transformers and document embeddings.

References

1. Blei, D.M.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
3. Cheng, X., Yan, X., Lan, Y., Guo, J.: BTM: topic modeling over short texts. IEEE Trans. Knowl. Data Eng. **26**(12), 2928–2941 (2014)
4. Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. J. Am. Soc. Inf. Sci. **41**(6), 391–407 (1990)
5. Du, L., Buntine, W., Jin, H., Chen, C.: Sequential latent Dirichlet allocation. Knowledge and Information Systems **31**, 475–503 (2012)
6. Glover, J.: Modeling documents with Generative Adversarial Networks. In: NIPS 2016 Workshop on Adversarial Training. pp. 1–7 (2016)
7. Grover, A., Dhar, M., Ermon, S.: Flow-GAN: Combining maximum likelihood and adversarial learning in generative models. In: The Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18). pp. 3069–3076. AAAI Press (2018)
8. Gui, L., Leng, J., Pergola, G., Zhou, Y., Xu, R., He, Y.: Neural topic model with reinforcement learning. In: Inui, K., Jiang, J., Ng, V., Wan, X. (eds.) The 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 3476–3481. Association for Computational Linguistics (2019)
9. Hennig, P., Stern, D.H., Herbrich, R., Graepel, T.: Kernel topic models. In: The Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2012. JMLR Proceedings, vol. 22, pp. 511–519 (2012)
10. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. In: The 2nd International Conference on Learning Representations (ICLR 2014). pp. 1–14 (2014)
11. Lau, J.H., Newman, D., Baldwin, T.: Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality. In: The 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014). pp. 530–539 (2014)
12. Li, W., Yin, J., Chen, H.: Supervised topic modeling using hierarchical Dirichlet process-based inverse regression: Experiments on e-commerce applications. IEEE Trans. Knowl. Data Eng. **30**(6), 1192–1205 (2018)
13. Miao, Y., Grefenstette, E., Blunsom, P.: Discovering discrete latent topics with neural variational inference. In: The 34th International Conference on Machine Learning (ICML 2017). pp. 2410–2419 (2017)
14. Röder, M., Both, A., Hinneburg, A.: Exploring the space of topic coherence measures. In: The Eighth ACM International Conference on Web Search and Data Mining (WSDM '15). pp. 399–408 (2015)
15. Srivastava, A., Sutton, C.A.: Autoencoding variational inference for topic models. In: The 5th International Conference on Learning Representations, (ICLR 2017). pp. 1–12 (2017)
16. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, second edn. (2018)
17. Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Mach. Learn. **8**, 229–256 (1992)
18. Zhang, Y., Xu, B., Zhao, T.: Convolutional multi-head self-attention on memory for aspect sentiment classification. IEEE/CAA Journal of Automatica Sinica **7**, 1038–1044 (2020)