

文章编号:1008-1534(2017)06-0421-07

# 基于变分自动编码器的动态主题模型

孙 凌,韩立新,勾智楠

(河海大学计算机与信息学院,江苏南京 211100)

**摘 要:**传统动态主题模型的后验分布推断需要复杂的推理过程,仅模型假设的细微变化就需要重新进行推断,时间成本较高,制约了模型的可变性和通用性。为了提高动态主题模型的性能,提出了基于变分自动编码器融合动态因子图进行推断的动态主题模型。该模型对变分下界进行再参数化,生成一个下界估计器,将隐变量转换为一组辅助参数,使得新的参数不依赖于变分参数,用标准随机梯度下降法直接优化变分目标,同时融合动态因子图对状态空间模型进行建模,弱化推断的概率特性,简化优化过程,实现有效的推断。结果表明,提出的模型不仅保证了准确性,而且其简化模型有效降低了推断的时间成本,从而为动态主题模型能有效应用于复杂的时间场景提供更多可能。

**关键词:**神经网络;动态主题模型;变分自动编码器;动态因子图;参数

**中图分类号:**TP389.1      **文献标志码:**A      **doi:** 10.7535/hbgykj.2017yx06006

## Dynamic topic model based on variational autoencoder

SUN Ling, HAN Lixin, GOU Zhinan

(College of Computer and Information, Hohai University, Nanjing, Jiangsu 211100, China)

**Abstract:** The posterior distribution of traditional dynamic topic model requires complex reasoning process, and a small change in model assume will require re-deduction, meanwhile with high time cost, which restricts the variability and generality of the model. A dynamic topic model based on variational autoencoder fusing with dynamic factor graph for inference is proposed in order to improve the performance of dynamic topic model. The model makes a reparameterization trick to evidence lower bound to generate a lower estimator, and converts the hidden parameters to a group of auxiliary parameters, which makes new parameters not depend on variational parameters; standard stochastic gradient descent method can be available to variational objective function directly. At the same time, integrating the dynamic factor graph on modeling the state space model weakens the probabilistic of the model, simplifies the optimization process, and makes effective inference. The experimental results show that this model guarantees the accuracy, and the simplified model reduces the time cost effectively, which will provide more possibilities for dynamic topic model to be applied to complex time scenarios effectively.

**Keywords:** neural network; dynamic topic model; variational autoencoder; dynamic factor graph; parameter

收稿日期:2017-08-28;修回日期:2017-10-17;责任编辑:陈书欣

基金项目:江苏省研究生科研与实践创新计划项目(KYCX17\_0486);中央高校基本科研业务费专项资金(2017B708X14);福建省信息处理与智能控制重点实验室(闽江学院)开放课题(MJUKF201740)

第一作者简介:孙 凌(1992—),女,江苏南京人,硕士研究生,主要从事机器学习、信息检索方面的研究。

通信作者:韩立新教授,博士生导师。E-mail:lhan@hhu.edu.cn

孙 凌,韩立新,勾智楠.基于变分自动编码器的动态主题模型[J].河北工业科技,2017,34(6):421-427.

SUN Ling, HAN Lixin, GOU Zhinan. Dynamic topic model based on variational autoencoder[J]. Hebei Journal of Industrial Science and Technology, 2017, 34(6): 421-427.

主题模型是一种非常流行的对文本隐含主题建模的方法,广泛应用于自然语言处理、信息检索等各个领域。针对具有时间属性的文本,BLEI等<sup>[1]</sup>首次提出了动态主题模型(dynamic topic model, DTM),在主题建模时引入时间维度,从而跟踪并分析主题随时间变化的状况。应用DTM的主要问题是参数估计的计算成本,传统的估计方法是利用EM算法迭代计算贝叶斯后验分布的变分推断方法<sup>[2]</sup>,将近似后验分布问题转换为解决KL散度的优化问题,对隐变量的推断转换为对变分变量的迭代更新,然而迭代模型的推导复杂且需高时间成本,尤其是新模型在建模假设有细微变化都需重新推断,严重制约了不同建模自由探索的实践可能性。由此激发了研究者对黑盒推断方法的探索,文献<sup>[3—6]</sup>提出了仅需有限的信息并给定生成过程,能够自动应用到新模型的一些方法。

变分贝叶斯自动编码(auto encoding variational bayes, AEVB)<sup>[6]</sup>是一种基于变分自动编码器(variational autoencoder, VAE)的深度生成模型,可对变分下界再参数化生成一个下界估计器,将隐变量转换为一组辅助参数,使得新的参数不依赖于变分参数,可用标准随机梯度下降法优化变分目标,从而简化优化过程,实现有效的后验推断。然而在训练过程中会造成组件塌陷<sup>[7]</sup>,即一种类似于先验信任的局部最优,在应用到主题模型时会造成所有主题都一样。SRIVASTAVA等<sup>[8]</sup>利用AEVB对主题模型进行推断,提出了新的主题模型ProdLDA,但是无法适用于DTM。基于以上研究,本文以神经网络模型进行概率推断为基础,提出一种基于变分自动编码器的动态主题模型推断学习方法。

## 1 动态主题模型

在DTM中,时间片 $t$ 上的序列语料的生成过程如下。

- 1) 根据 $\beta_t | \beta_{(t-1)} \sim N(\beta_{(t-1)}, \sigma^2 I)$ 生成时间片 $t$ 上的主题-词汇概率分布 $\beta_t$ ;
- 2) 根据 $\alpha_t | \alpha_{(t-1)} \sim N(\alpha_{(t-1)}, \delta^2 I)$ 生成时间片 $t$ 上的先验主题先验分布 $\alpha_t$ ;
- 3) 循环生成该时间片 $t$ 上的每一篇文章 $d$ :
  - ① 根据 $\theta \sim N(\alpha_t, a^2 I)$ 生成时间片 $t$ 上的文档主题概率分布 $\theta$ ;
  - ② 循环文档 $d$ 中的每一个词 $n$ ,根据 $Z \sim \text{Mult}(\pi(\theta))$ 生成词-主题分配标识向量 $Z$ ;根据 $W_{(t,d,n)} \sim \text{Mult}(\pi(\beta_{(t,z)}))$ 生成词 $W_{(t,d,n)}$ 。

DTM包含的隐变量包括文档和主题层面上的,分别为主题分布 $\theta$ 、词-主题分配标识向量 $Z$ ,时间片 $t$ 上的主题-词汇分布参数变量 $\beta_t$ 。

在DTM中,高斯模型被应用于时间序列上的自然参数的处理,从而对时间上的动态变化进行建模,然而,由于高斯分布和多元分布的非共轭性,直接对后验分布进行推断难以操作。另一方面,尽管Gibbs抽样<sup>[9]</sup>在静态的主题模型上得到有效的应用,但是非共轭性导致抽样方法难以应用于动态模型<sup>[10]</sup>。因此,DTM采用变分推断<sup>[11]</sup>来近似后验分布。变分推断首先假设一个较简单的包含隐变量的分布,这个分布包含相应的变分变量,通过不断更新变分变量,达到优化变分分布与真实后验分布之间的KL散度的目的,当KL散度小于某个阈值后,可将变分分布作为真实的后验分布的近似替代。在文档层面的隐变量有 $\theta_{(t,d)}$ 和 $z_{(t,d,n)}$ ,对应变分变量的更新采用共轭梯度法来优化,对主题层面的隐变量 $\beta_{(t,k)}$ 的变分变量的更新采用基于卡尔曼过滤算法的近似推断<sup>[11]</sup>。然而,尽管对目标下界的优化和变分变量的更新均有解析解,但由于灵活性受限,一旦应用和需求变化所带来的新模型则需要重新推导出有解析解的迭代更新模型,尤其对一些复杂的模型难以得到,甚至没有解析解。

DTM中的状态空间模型描述了变量 $\beta$ 随时间的变化,由此对生成文档产生影响。DTM中采用基于卡尔曼滤波的近似方法对主题参数 $\beta$ 进行推断,需要复杂的推导,本文用动态因子图(dynamic factor graph, DFG)<sup>[12]</sup>描述状态空间模型,简化推断。

状态变量和观测变量之间的依赖可表示为在序列数据上的因子图,并且在时间片上重复。动态因子图即因子图用于协调序列数据的扩展,相比现有的状态空间方法,DFG允许建模过程中使用复杂的函数来预测状态和观测值,弱化推断的概率特性,实现高维度和非线性模型的有效推断<sup>[12]</sup>。

## 2 基于AEVB的推断学习方法

针对传统变分推断方法的局限性,AEVB提供了很好的选择,通过训练推断网络编码器将文档映射到近似后验分布,并且利用再参数化将隐变量转换为一组辅助参数,使得新的参数不依赖于变分参数,无需推导出变分参数的迭代更新模型。

AEVB是基于变分自动编码器VAE的一种深度生成模型,可为主题模型提供一种有效的变分推断和学习算法,其网络结构如图1所示。AEVB利用神经网络来分别建模主题模型的文档生成概率和

表示文档-主题的后验概率分布的变分近似分布。其中,用于建模变分近似分布的神经网络作为编码器,将输入文档映射为表示文档-主题的后验分布;用于建模文档生成概率的神经网络作为解码器,将后验分布映射为生成的文档概率分布。



图 1 基于 AEVB 推断学习的网络结构

Fig.1 Network structure of inference learning based on AEVB

基于 AEVB 的推断学习方法,即使在大规模文档数据集也可以实现有效的推断和学习,并在一些轻微的可微性条件下,能处理复杂的后验分布<sup>[7]</sup>,由此面对主题模型由应用和需求变化带来的复杂新模型,AEVB 只需对变分下界的再参数化产生一个下界估计器适配一个近似推断模型来处理复杂的后验,最后用标准随机梯度下降方法直接优化,由此实现有效的后验推断和学习。但是,针对动态主题模型中满足时序动态变化的状态变量,AEVB 无法建模该状态变量的时序依赖关系。

### 3 基于变分自动编码器的动态主题模型 VADTM

采用基于变分自动编码器的推断方法对 DTM 进行学习。利用神经网络来分别建模 DTM 的文档生成概率  $p(w'_t | \theta_t, \beta_t)$  和表示文档-主题后验的变分近似分布  $q_\phi(\theta_t | w_t)$ 。其中,用于建模变分近似分布的神经网络作为编码器,将输入文档  $w_t$  映射为表示文档-主题的后验分布  $q_\phi(\theta_t | w_t)$ ,并捕获动态先验,避免陷入局部最优;用于建模文档生成概率的神经网络作为解码器,将后验  $q_\phi(\theta_t | w_t)$  映射为新文档  $w'_t$  的生成概率  $p(w'_t | \theta_t, \beta_t)$ ,并将解码器融合动态因子图对满足时序依赖的状态变量进行建模。本文提出的基于变分自动编码器的动态主题模型 VADTM 的网络结构如图 2 所示。

首先,根据上述 DTM 生成过程得出模型的变分目标。一个时间片  $t$  内按照静态主题模型的思想建模,那么生成文档  $w'_t$  的边际似然函数为

$$p(w'_t | \alpha_t, \beta_t) =$$

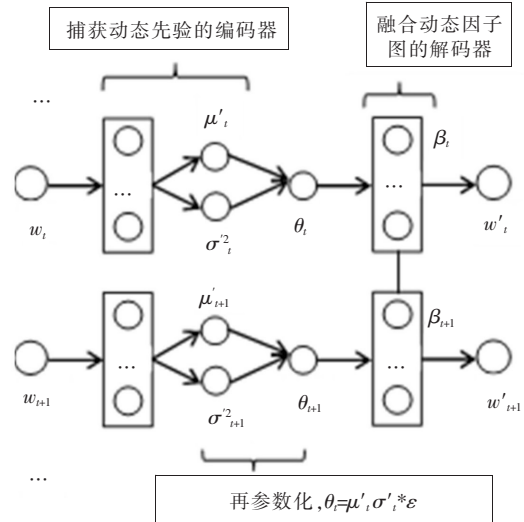


图 2 VADTM 的网络结构

Fig.2 Network structure of VADTM

$$\int \left( \prod_{n=1}^N \sum_{z_{t,n}} p(w'_{(t,n)} | z_{(t,n)}, \beta_t) p(z_{(t,n)} | \theta_t) \right) p(\theta_t | \alpha_t) d\theta_t. \quad (1)$$

考虑到需要利用再参数化对隐变量进行处理,而文档中每个单词分配的主题标识变量  $z_t$  难以进行再参数化,同时作为生成过程中的过渡变量,对表示最后生成的文档以及理解新的文档都没有用处,所以本文使用加法规则<sup>[9]</sup>将该变量去除,生成文档  $w'_t$  的边际似然函数则简化为式(2)所示:

$$p(w'_t | \alpha_t, \beta_t) = \int \prod_{n=1}^N p(w'_{(t,n)} | \theta_t, \beta_t) p(\theta_t | \alpha_t) d\theta_t. \quad (2)$$

为了计算  $\log p(w'_t | \alpha_t, \beta_t)$ ,需引入一个额外的变分函数  $q_\phi(\theta_t | w_t)$ ,则生成文档  $w'_t$  的边际对数似然函数如式(3)所示:

$$l(w'_t, \alpha_t, \beta_t) = \log \int q_\phi(\theta_t | w_t) \times \frac{p(w'_t | \theta_t, \beta_t) * p(\theta_t | \alpha_t)}{q_\phi(\theta_t | w_t)} d\theta_t. \quad (3)$$

则边际对数似然函数  $l(w'_t, \alpha_t, \beta_t)$  的变分下界<sup>[7]</sup>  $\mathcal{L}(q, \beta_t, \alpha_t; w_t)$  如式(4)所示:

$$\mathcal{L}(q, \beta_t, \alpha_t; w_t) \triangleq E_{q_\phi(\theta_t | w_t)} [\log p(w_t | \theta_t, \beta_t)] - DL[q_\phi(\theta_t | w_t) || p(\theta_t | \alpha_t)]. \quad (4)$$

仅当变分函数  $q_\phi(\theta_t | w_t) = p(\theta_t | w_t; \alpha_t)$  时,  $l(w'_t, \alpha_t, \beta_t) = \mathcal{L}(q, \beta_t, \alpha_t; w_t)$ 。然而由于涉及积分计算,无法计算真实的后验分布  $p(\theta_t | w_t; \alpha_t)$ ,

因此找到最接近的近似分布  $q$ , 通过 KL 散度来衡量 2 个分布之间的距离来实现  $l(w'_t, \alpha_t, \beta_t)$  的近似计算, 即求  $\phi^* = \operatorname{argmin} DL(q_\phi(\theta_t | w_t) || p(\theta_t | w_t; \alpha_t))$ 。由文献[7]可知, 近似分布  $q_\phi(\theta_t | w_t)$  与真实后验分布  $p(\theta_t | w_t; \alpha_t)$  的 KL 散度等于对数边际似然与其下界的差, 由此  $t$  时刻变分目标即最大化变分下界, 如式(5)所示:

$$\begin{aligned} \phi^* &= \operatorname{argmin} DL(q_\phi(\theta_t | w_t) || p(\theta_t | w_t; \alpha_t)) = \\ &= \operatorname{argmin}(l(w'_t, \alpha_t, \beta_t) - \mathcal{L}(q, \beta_t, \alpha_t; w_t)) = \\ &= \operatorname{argmax} \mathcal{L}(q, \beta_t, \alpha_t; w_t) = \\ &= \operatorname{argmax}(E_{q_\phi(\theta_t | w_t)}[\log p(w'_t | \theta_t, \beta_t)] - \\ &DL[q_\phi(\theta_t | w_t) || p(\theta_t | \alpha_t)]), \end{aligned} \quad (5)$$

式中: 第 1 项为负的平均重构误差; 第 2 项为近似分布  $q$  和先验分布  $\theta$  之间的 KL 距离。由此变分目标的两项可由解码器和编码器来分别建模。

### 3.1 捕获动态先验的编码器

在 VAE 中, 用神经网络来估计  $p(\theta_t | w_t; \alpha_t)$  的近似分布  $q_\phi(\theta_t | w_t)$ , 称为推断网络, 即编码器。假设  $q_\phi(\theta_t | w_t)$  服从对角化协方差的正态分布, 则  $\log q_\phi(\theta_t | w_t) = \log \mathcal{N}(w_t; \mu'_t, \sigma'^2_t)$ , 其中均值  $\mu'_t$  和协方差  $\sigma'^2_t$  用神经网络来计算。

同样, DTM 中  $\theta_t$  服从以  $\alpha_t$  为均值的对数高斯分布, 即  $\log p(\theta_t | w_t) = \log \mathcal{N}(w_t; \mu_t, \sigma_t)$ , 其中  $\mu_t = \alpha_t, \sigma_t = a$ 。那么式(5)中第 2 项记作  $L_{kl}$ , 则

$$\begin{aligned} L_{kl} &= DL[q_\phi(\theta_t | w_t) || p(\theta_t | \alpha_t)] = \\ &= \frac{1}{2}(\operatorname{tr}(\sigma_t^{-1} \sigma'^2_t) + (\mu_t - \mu'_t)^T \\ &\sigma_t^{-1}(\mu_t - \mu'_t) - K + \log \frac{|\sigma_t|}{|\sigma'^2_t|})。 \end{aligned} \quad (6)$$

其中超参数  $\alpha_t$  满足高斯随机游走模型, 作为常数, 本文在初始化时采用满足高斯分布的随机采样进行参数微调, 而非全设为 1<sup>[9]</sup>, 在一定程度上缓解组件塌陷, 从而获得解释性更优的主题词。

为了计算式(5)中第 1 项对应  $q_\phi$  的期望, 用再参数化即一个蒙特卡洛估计器<sup>[1]</sup>来计算  $q_\phi$ : 添加辅助参数  $\epsilon$ , 使得  $\theta_t = \mu'_t + \sigma'_t * \epsilon$ , 通过对  $\epsilon \sim \mathcal{N}(0, I)$  采样实现对  $\theta_t \sim \mathcal{N}(\mu'_t, \sigma'^2_t)$  的采样, 此时从编码器输出到  $\theta_t$  只涉及线性操作, 因此容易实现使用随机梯度下降法进行优化。

### 3.2 基于动态因子图的解码器

解码器根据已知  $q_\phi(\theta_t | w_t)$  的情况下, 估计更好的生成模型  $p(w'_t | \theta_t, \beta_t)$ , 使得输入文档  $w_t$  与生成的新文档  $w'_t$  尽量接近, 即  $w_t$  经过编码再解码尽可能多的恢复出原来的信息, 使得式(5)第 1 项平

均重构误差最小。在时间片  $t$  给定训练集  $D$ , 其中包含  $N$  个独立同分布样本  $w_t^1, w_t^2, \dots, w_t^N$ , 由 3.1 可知, 其中式(5)中第 1 项  $E_{q_\phi(\theta_t | w_t^i)}[\log p(w_t^i | \theta_t, \beta_t)]$  通过采样计算, 每次采样 1 个点, 那么  $E_{q_\phi(\theta_t | w_t^i)}[\log p(w_t^i | \theta_t, \beta_t)] \approx \log p(w_t^i | \theta_t, \beta_t)$ , 且该模型服从多变量伯努利分布, 则该对数似然即负的交叉熵。

上述是在  $t$  时刻的生成模型, 在考虑相邻时间片状态变量的依赖关系后, 本文采用动态因子图来改进解码器。变量  $\beta_t$  满足如下状态空间模型:

$$\begin{aligned} \beta_t | \beta_{t-1} &\sim \mathcal{N}(\beta_{t-1}, \sigma^2 I), \\ w'_t &\sim \operatorname{Mult}(\pi(\beta_t)). \end{aligned} \quad (7)$$

将其表示为动态因子图的形式, 如图 3 所示。

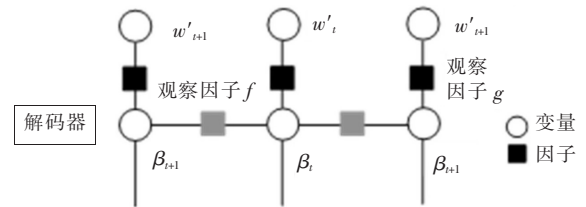


图 3 状态空间模型转换的动态因子图

Fig.3 Dynamic factor graph converted from state space model

其中观察因子  $g$  连接状态变量  $\beta_t$  和观察变量  $w_t$ , 即  $w'_t = g(\beta_t; \theta_t)$ ; 动态因子  $f$  连接相邻时间片之间的状态变量  $\beta$ , 即  $\beta_t = f(\beta_{t-1})$ , 并建模变量之间的条件依赖。因子可以用负对数域表示: 每个因子计算一个能量值, 即相互连接的变量配置的负对数似然。系统的总能量是因子能量的总和, 所以变量配置的最大化似然通过最小化总能量来获得。

模型的似然概率是每个因子发生概率的乘积, 在对数空间中, 乘积变求和。观察因子  $g(\beta_t; \theta_t)$  发生的概率为  $p(w'_t | \theta_t, \beta_t)$ , 对数似然记作  $L_g$ , 由前述可知, 即负的交叉熵,  $L_g = -w_t * \log p(w'_t | \theta_t, \beta_t)$ 。

对于动态因子, 由于从  $(t-1)$  时刻的主题到  $t$  时刻的主题满足高斯随机游走模型,  $f = p(\beta_t | \beta_{t-1}, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(\beta_t - \beta_{t-1})^2}{2\sigma^2})$ , 则对数似然  $L_f = -\|\beta_t - \beta_{t-1}\|_2$  (L2 正则)。

综上所述, 变分目标(5)最终可以表示为如下的目标损失函数:

$$\begin{aligned}
& \operatorname{argmax} \mathcal{L}(\Theta) = \operatorname{argmax}(L_g + L_f - L_{kl}) = \\
& \operatorname{argmin} \sum_t^T (\omega_t * \log p(w'_t | \theta_t, \beta_t) + \\
& \quad \|\beta_t - \beta_{(t-1)}\|_2 + L_{kl}) = \\
& \operatorname{argmin} \sum_t^T \sum_i^D (\omega_{t,i} * \log(\pi(\beta_t) \\
& \quad \pi(\mu'_t + \sigma'_t * \epsilon))) + \|\beta_t - \beta_{(t-1)}\|_2 + \\
& \quad \frac{1}{2} (\operatorname{tr}(\sigma_t^{-1} \sigma'_t) + (\mu_t - \mu'_t)^T \\
& \quad \sigma_t^{-1} (\mu_t - \mu'_t) - K + \log \frac{|\sigma_t|}{|\sigma'_t|})). \quad (8)
\end{aligned}$$

基于此优化目标,本文采用随机梯度下降法训练整个网络。

### 3.3 相关工作比较

动态主题模型性能优化的关键是如何获取更高效的训练算法<sup>[10]</sup>。常用的包括利用EM算法迭代计算后验分布的变分推断方法<sup>[2]</sup>和利用马尔可夫链蒙特卡洛方法的Gibbs抽样方法<sup>[9]</sup>,然而变分推断方法需要对变分变量的迭代更新公式进行复杂的推导过程,难以应用到扩展的模型,而本文结合VAE网络架构,利用再参数化将编码器输出的2个网络与辅助参数进行线性操作,此时对变分变量的推断简化为对2个网络参数的优化,简化了迭代更新模型;相较于Gibbs抽样方法,每次迭代采样都需遍历所有数据点直到收敛,本文的VADTM只需对正态分布进行一次采样,优化了推断时间。SRIVASTAVA等<sup>[8]</sup>同样利用AEVB进行变分推断,并基于该推断方法提出新的主题模型ProdLDA,但不适用动态主题模型,本文提出的VADTM融合了动态特性,对先验进行调优,同时将状态空间模型转换为动态因子图进行时序建模并适配解码器。由此,VADTM保证传统方法的准确性同时,还节约了时间成本,实现有效的性能优化。

## 4 实验及结果分析

### 4.1 实验环境

硬件上,采用英特尔酷睿 i7 四核处理器,处理器主频为 2.2 GHz,内存大小为 16 GB。软件上,采用 macOS Sierra 操作系统,深度学习框架 Tensorflow0.10.0,Python2.7 作为开发环境。

### 4.2 实验数据

采用一个时间跨度为 25 年(1985—2009),以自然语言处理领域中期刊论文(DBLP)为主导的语料库<sup>[13]</sup>进行实验。选取该语料库 13 600 篇论文的

英文摘要作为文档。语料库的词汇数目为 10 984。本文将语料库分为 10 个时间片,每个时间片上的论文数大致相当,对每篇论文进行预处理,包括去除非 UTF-8 字符、标点、长度小于 3 的单词、英文停止词等,并进行分词以及单词词干化等预处理。

### 4.3 评估方法

主题模型所面临的另一个主要问题是如何客观地衡量模型的优劣度,由于其非监督的特性,不同的应用中需实现不同的任务,使得模型的选择变得困难,因此很难直接评估一个模型的好坏<sup>[11]</sup>。传统方法使用 perplexity 对主题进行定性评估,文献<sup>[14]</sup>指出该方法不能很好地表示主题词的语义特性,LAU等<sup>[15]</sup>提出主题相关性的评估标准更接近人类判断,并使用归一化点互信息(normalized pointwise mutual information,NPMI)对主题定性评估。本文采用 NPMI 方法,NPMI( $t$ )表示主题  $t$  的主题相关性,取主题  $t$  的 top- $N$  主题词集进行计算,最后取  $K$  个主题的平均主题相关性作为评估指标,平均主题相关性越大,表示模型越优。

$$\text{NPMI}(t) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}}{-\log p(w_i, w_j)}.$$

### 4.4 对比实验结果分析

采用文献<sup>[9]</sup>提出的使用 Gibbs 抽样进行推断的 LDA、文献<sup>[2]</sup>提出的使用平均场变分推断的 DTM、文献<sup>[8]</sup>提出的基于变分自动编码器推断的 ProdLDA 以及本文提出的 VADTM 进行比较。实验中,选取主题数量为 50 进行实验。超参数  $\alpha_t$  是文档-主题分布即高斯分布的均值,它的初始化值需反映真实语料的分布特征,实验通过设置不同的  $\alpha_t$  取值来描述真实语料的概率分布,通过比较主题相关性来找到最优的  $\alpha_t$ ,图 4 给出了 VADTM 在超参数  $\alpha_t$  设置不同高斯分布方差的条件下平均主题相关性的对比结果,结果显示,方差  $\alpha_t$  取 0.8 时,模型达到最优,即 0.8 最能反映真实语料的概率分布特征。表 1 给出了当超参数  $\alpha_t$  全部设置为 1 时模型学到的主题词,结果显示在不同时间片中都存在相近的且没有意义的主题词,说明模型出现一定程度的组件塌陷问题,而当超参数  $\alpha_t$  的取值为高斯分布的随机采样结果时,能避免产生相近的且无意义的主题词,如表 2 所示。表 2 选取了同表 1 相同时间片下的部分主题词,可见不同主题之间不再出现相似的主题词且主题词解释性更优。说明一定程度上缓解了 AEVB 的组件塌陷问题,主题的解释性

更强。

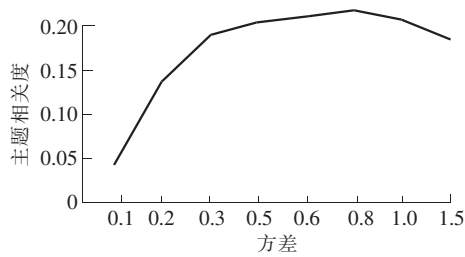


图4 不同方差设置下主题相关性对比

Fig.4 Contrast of topic coherence by different variances

表1 超参数  $\alpha_i$  全部设置为 1 时部分主题词

Tab.1 Partial topic words when  $\alpha_i$  is set as 1

1.fuzzy controller neural self network learning tuning rule predictive control
2.fuzzy controller control optimal mark tracking robot redundancy end coding
3.control controller fuzzy self tracking scheme internal tuning predictive structure
4. fuzzy fermentation repetitive tune controller chattering rule yeast yield timing

表2 超参数  $\alpha_i$  取值为随机采样结果时部分主题词

Tab.2 Partial topic words when  $\alpha_i$  is set as the random sampling results

1.control satisfactory fuzzy reheat industrial optimization quality network process neural
2.neural repetitive systematic chattering baker yeast fed net acquiring network
3.camera retrieval image face redundancy spectral valley recognition contour active
4.control fuzzy controller predictive privilege section payment electronic cash audit

表3显示了所有模型的平均主题相关性的比较。结果显示 ProdLDA 和 VADTM 产生了相近的主题相关性且优于两种传统方法。同时 VADTM 耗时 50 min 左右训练模型,然而基于平均场变分推断的 DTM 则耗时超过 10 h。这说明了本文提出的基于变分自动编码器并融合动态因

子图进行推断的动态主题模型 VADTM,保证一定准确性的同时,有效优化了推断时间,节约了时间成本,相较于 ProdLDA 也实现了动态模型的应用。

表3 不同方法的平均主题相关性比较

Tab.3 Contrast of topic coherence by different method

VADTM	LDA(Gibbs 抽样)	DTM(变分推断)	ProdLDA
0.218	0.153	0.112	0.221

## 5 结 语

基于变分自动编码器神经网络结构训练 DTM,使用再参数化将 DTM 中文档层面的隐变量文档-主题  $\theta$  转换为一组辅助参数,使得新的参数不依赖于变分参数,可用标准随机梯度下降法直接优化变分目标,同时将主题  $\beta$  的状态空间模型转换为动态因子图进行时序建模,将复杂的变分推导转换为对状态  $\beta$  的正则化约束,简化了变分目标。实验表明,本文提出的 VADTM 保证准确性,优化推断时间,实现了有效的性能优化。未来,将考虑根据具体的应用例如图像识别<sup>[16]</sup>来调整模型,并结合微博平台利用改进的 DTM 来建模用户信任关系<sup>[17]</sup>。

## 参考文献/References:

- [1] BLEI D M, LAFFERTY J D. Dynamic topic models[C]// Proceedings of the 23rd International Conference on Machine Learning. New York:ACM, 2006: 113-120.
- [2] WAINWRIGHT M J, JORDAN M I. Graphical models, exponential families, and variational inference[J]. Foundations and Trends in Machine Learning, 2008, 1(1/2): 1-305.
- [3] RANGANATH R, GERRISH S, BLEI D. Black box variational inference [C]//Artificial Intelligence and Statistics. [S.l.]:[s.n.], 2014: 814-822.
- [4] MNH A, GREGOR K. Neural variational inference and learning in belief networks [J]. International Conference on Multimedia Retrieval, 2014, 32:1791-1799.
- [5] KUCUKELBIR A, TRAN D, RANGANATH R, et al. Automatic differentiation variational inference [J]. Journal of Machine Learning Research, 2017, 18(1):430-474.
- [6] KINGMA D P, WELING M. Auto-encoding variational bayes [EB/OL]. [http://dpkingma.com/wordpress/wp-content/uploads/2014/05/2014-03\\_talk\\_iclr.pdf](http://dpkingma.com/wordpress/wp-content/uploads/2014/05/2014-03_talk_iclr.pdf), 2014-05-01.
- [7] DINH L, DUMOULIN V. Training neural Bayesian nets[EB/OL]. [http://www. iro. umontreal. ca/~ bengioy/cifar/NCAP2014-summer-sch-ool/slides/Laurent\\_dinh\\_cifar\\_presentation.pdf](http://www. iro. umontreal. ca/~ bengioy/cifar/NCAP2014-summer-sch-ool/slides/Laurent_dinh_cifar_presentation.pdf), 2016-02-15.

- [8] SRIVASTAVA A, SUTTON C. Autoencoding variational inference for topic models[EB/OL].[https://www.researchgate.net/publication/314261246\\_Autoencoding\\_Variational\\_Inference\\_For\\_Topic\\_Models](https://www.researchgate.net/publication/314261246_Autoencoding_Variational_Inference_For_Topic_Models),2017-03-29.
- [9] GRIFFITHS T. Gibbs sampling in the generative model of latent dirichlet allocation [ R ]. Stanford: Stanford University,2002.
- [10] 桂小庆,张俊,张晓民,等.时态主题模型方法及应用研究综述[J].计算机科学,2017,44(2):46-55.  
GUI Xiaoqing, ZHANG Jun, ZHANG Xiaomin, et al. Survey on temporal topic model methods and application [ J ]. Computer Science, 2017,44(2):46-55.
- [11] WINN J M. Variational Message Passing and its Applications [D]. London:University of Cambridge, 2004.
- [12] MIROWSKI P, LECUN Y. Dynamic factor graphs for time series modeling [ J ]. Machine Learning and Knowledge Discovery in Databases, 2009,5782: 128-143.
- [13] 蒋卓人,陈燕,高良才,等.一种结合有监督学习的动态主题模型[J].北京大学学报(自然科学版),2015,51(2):367-376.  
JIANG Zhuoren, CHEN Yan, GAO Liangcai, et al. A supervised dynamic topic model [ J ]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2015, 51(2):367-376.
- [14] NEWMAN D, LAU J H, GRIESER K, et al. Automatic evaluation of topic coherence [ C ]// Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association of Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2010:100-108.
- [15] LAU J H, NEWMAN D, BALDWIN T. Machine reading tea leaves: Automatically evaluating topic coherence and topic model quality [ C ]// Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. [ S.l. ]: [ s.n. ], 2014: 530-539.
- [16] 郑琨,张杨,赖杰,等.基于支持向量机和特征向量提取的人脸识别框架[J].河北工业科技,2016,33(1):58-62.  
ZHENG Kun, ZHANG Yang, LAI Jie, et al. Face recognition framework based on support vector machine and feature vector extraction [ J ]. Hebei Journal of Industrial Science and Technology, 2016, 33(1):58-62.
- [17] 孙博,陈通.微博用户之间信任形成的因素研究[J].河北工业科技,2015,32(5):384-389.  
SUN Bo, CHEN Tong. Factors on formation of trust between microblog users [ J ]. Hebei Journal of Industrial Science and Technology, 2015, 32(5):384-389.