1 Training Algorithm

In this section, we summarize the learning process into the training algorithm at Algo. 1.

2 Additional Experiments

2.1 Different Number of Topics for Document Classification

In the main paper, we set 64 topics and conduct document classification. Here, we vary the number of topics $K$ from 16 to 256, and summarize classification accuracy on the overall test set at Fig. 1.

Overall, our models perform stably across different number of topics. GATON presents the best results among baseline models in plain text and word embedding category, since it can incorporate higher-order connectivity on the graph by multi-layer convolutions. By comparing to it, our models further improve the performance, due to the advantage of semantic complement meta-learning. When incorporating auxiliary document networks, our models improve GATON more, since both textual and structural semantic complement helps short text topic modeling.

2.2 Complete Results of Effect of Semantic Complement and Meta-Learning

Due to space limit, we present the effect of semantic complement and meta-learning for the plain text version of our models in the main paper. Here we show the complete results at Table 1, including the word embedding version and document network version. Similarly, for all three versions, our models without semantic complement or meta-learning deteriorate the results, which verifies the usefulness of semantic complement and meta-learning to improve topic quality. Furthermore, the results decline more on the short subset than on the overall test set, which demonstrates that semantic complement and meta-learning indeed help improve short text topic modeling, and disregarding them leads to worse performance on short documents.
Algorithm 1 Training Process of MCTM

Input: Corpus $\mathcal{D}$. (Word embeddings $\mathcal{H}$ and document network $\mathcal{E}$, if available.)
Output: Topic model $f_\theta$ and semantics complement function $g_\mu$, with parameters $\Phi = \{\theta, \mu\}$.  

1: Initialize model parameters $\Phi = \{\theta, \mu\}$.  
2: \textbf{while} not converged \textbf{do}  
3: Sample a minibatch of documents.  
\hspace{1em} ////Encoding  
4: \hspace{1em} for each document $d$ in the minibatch \textbf{do}  
5: \hspace{2em} Encoding documents and words by $z_d, z_w = f_\theta(z^{(l=0)}_d, z^{(l=0)}_w)$.  
6: \hspace{2em} if word embeddings are available then  
7: \hspace{3em} $z_{k,w} := \frac{1}{2} (z_{k,w} + \cos(t_k, h_w))$.  
8: \hspace{3em} Obtain new topic-word distribution $Z_Y = [z_{k,w}]$.  
9: \hspace{2em} end if  
10: \hspace{2em} if document networks are available then  
11: \hspace{3em} $\kappa^{(l+1)}_d = \text{AGG}(W^{(l+1)}_3 \kappa^{(l)}_d, W^{(l+1)}_3 \kappa^{(l)}_{d'} | d' \in \mathcal{S}_{N_d}(d))$ for $l = 0, 1, ..., L - 1$.  
12: \hspace{3em} $z_d := \frac{1}{2} (z_d + \kappa_d)$.  
13: \hspace{2em} end if  
14: \hspace{1em} end for  
\hspace{1em} ////Semantic complement  
15: \hspace{1em} for each document $d$ in the minibatch \textbf{do}  
16: \hspace{2em} if $d \in \mathcal{D}_{\text{long}}$ then  
17: \hspace{3em} $d' \leftarrow \text{WordDropout}(d)$, such that $l_{d'} < \mathcal{L}$.  
\hspace{3em} //Missing semantics prediction  
18: \hspace{3em} Infer missing semantics for pruned short document $d'$ by GNN function or semantics clustering function. $m_{d'} = g_\mu(z_{d'}, z_w | w \in S_d)$.  
19: \hspace{3em} Semantic complement $z_{d'}^{*} = z_{d'} + m_{d'}$.  
\hspace{3em} //Constraint loss  
20: \hspace{3em} Evaluate constraint loss $L_{\text{con}}(d)$.  
21: \hspace{2em} end if  
22: \hspace{2em} if $d \in \mathcal{D}_{\text{short}}$ then  
23: \hspace{3em} //Missing semantics prediction  
24: \hspace{4em} Repeat line 18–19 and obtain complemented $z_{d'}^{*}$.  
25: \hspace{2em} end if  
26: \hspace{1em} Evaluate generative loss $L_{\text{gen}}(d)$ and complete loss $L(d)$.  
27: \hspace{1em} end for  
\hspace{1em} //Meta-learning optimization  
28: \hspace{1em} for each document $d$ (including pruned short version $d'$) \textbf{do}  
29: \hspace{2em} Local update for each document $d$ and obtain “personalized” $d$-specific parameter $\theta_d$.  
30: \hspace{2em} end for  
31: Global update w.r.t. loss for all documents and obtain $\Phi^* = \{\theta^*, \mu^*\}$.  
32: $\Phi \leftarrow \Phi^*$.  
33: end while
Figure 1: Document classification accuracy with different number of topics on ML and PL.

Table 1: Effect of semantic complement and meta-learning on document classification on ML.

<table>
<thead>
<tr>
<th>Model</th>
<th>Effect of Semantic Complement</th>
<th>Effect of Meta-Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall test set</td>
<td>Short subset</td>
</tr>
<tr>
<td></td>
<td>with</td>
<td>without</td>
</tr>
<tr>
<td>MCTM (G) (decline)</td>
<td>67.1±2.1</td>
<td>58.2±2.4</td>
</tr>
<tr>
<td>MCTM (C) (decline)</td>
<td>66.9±1.4</td>
<td>56.7±3.1</td>
</tr>
<tr>
<td>MCTM+WE (G) (decline)</td>
<td>66.8±2.0</td>
<td>62.6±2.3</td>
</tr>
<tr>
<td>MCTM+WE (C) (decline)</td>
<td>66.0±1.4</td>
<td>62.1±2.2</td>
</tr>
<tr>
<td>MCTM+DN (G) (decline)</td>
<td>83.3±1.7</td>
<td>71.6±1.2</td>
</tr>
<tr>
<td>MCTM+DN (C) (decline)</td>
<td>83.0±1.2</td>
<td>72.5±1.4</td>
</tr>
</tbody>
</table>