**Abstract**

- Despite recent advances in learning and inference algorithms, evaluating the predictive performance of topic models is still painfully slow and unreliable.
- We propose a new strategy for computing relative log-likelihood (or perplexity) scores of topic models, based on annealed importance sampling.
- The proposed method has smaller Monte Carlo error than previous approaches, leading to marked improvements in both accuracy and computation time.

**Annealed Importance Sampling (Neal, 2001)**

- Draw from distribution \( p_0 \).
- Simulated annealing towards \( \text{Target distribution of interest } p_0 \).
- Use this as an importance sampling proposal distribution for:
  - Annealing in the reverse direction, from the target to the source.

The importance samples can be used to estimate the ratio of normalizing constants of \( f_0 \propto p_0 \) and \( f_0 \propto p_0 \), via

\[
\sum \frac{w^{(i)}}{N} = \int f_0(z|x)dz = \int f_0(z|x)dz
\]

Wallach et al. (2009) show how to employ AIS in the context of topic models to estimate \( P_z(w|\phi, \alpha) \).
- Perform AIS on the topic assignments \( z \).
- Anneal from the prior to the posterior.
- Estimate the likelihood by averaging the importance samples.

**The Proposed Method**

- Typically for evaluation we are interested in the relative performance of topic model 1 (e.g. a new model) and topic model 2 (e.g. vanilla LDA):

\[
\log Pr(w_{(d)}|\phi^{(1)}), \alpha^{(1)}_{(d)}) - \log Pr(w_{(d)}|\phi^{(2)}), \alpha^{(2)}_{(d)})
\]

\[
\begin{align*}
&= \log Pr(w_{(d)}|\phi^{(1)}), \alpha^{(1)}_{(d)}) \\
&\quad - \log Pr(w_{(d)}|\phi^{(2)}), \alpha^{(2)}_{(d)}) \\
&= \log Pr(w_{(d)}|\phi^{(1)}), \alpha^{(1)}_{(d)}) \\
&\quad \times \frac{Pr(w_{(d)}|\phi^{(1)}), \alpha^{(1)}_{(d)})}{Pr(w_{(d)}|\phi^{(2)}), \alpha^{(2)}_{(d)})}
\end{align*}
\]

- This could be estimated by running AIS once for each model.
- However, AIS is already capable of computing ratios. We therefore propose to use AIS to compute this ratio directly. The procedure is:

- Draw from posterior for Topic Model 2.
- Simulated annealing towards posterior for Topic Model 1.

**Overall Results**

<table>
<thead>
<tr>
<th>Method</th>
<th>Percent of Documents with Correct Evaluation (I.e., the Unperturbed Topics Win vs the Perturbed Topics)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expensive runs</td>
<td>AIS: 88%</td>
</tr>
<tr>
<td></td>
<td>AIS (difference): 95%</td>
</tr>
<tr>
<td></td>
<td>AIS (difference, reverse): 95%</td>
</tr>
<tr>
<td>Cheap runs</td>
<td>AIS: 52%</td>
</tr>
<tr>
<td></td>
<td>AIS (difference): 95%</td>
</tr>
<tr>
<td></td>
<td>AIS (difference, reverse): 96%</td>
</tr>
</tbody>
</table>

**Comparison to Ground Truth on Very Small Problems**

- In this graph, lower values are better.
- Note: in this regime (4 topics, 8 words per document), importance sampling is better than the naive AIS method. This does not hold in general.

**Varying the Number of Temperatures**

- The proposed method is much more stable. One importance sample gives essentially the same answer as 100 importance samples.
- The number of temperatures, which controls the amount of the space explored, is important for all methods.
- Recommendation: use the proposed method, with one importance sample, and as many temperatures as time permits.

**Mathematical Details**

The standard AIS method for topic models (Wallach et al., 2009):
- AIS on topic assignments \( z^{(d)} \), collapsing \( \theta^{(d)} \).
- Draw initial state from the prior over \( z \), \( \tau_0 = Pr(z^{(d)}|\theta^{(d)}) \).
- Anneal towards a distribution proportional to the posterior, \( \tau_0 = Pr(z^{(d)}|\theta^{(d)}) \).
- Estimate the likelihood via:

\[
\sum \frac{w^{(i)}}{N} = \sum \frac{Pr(w_{(d)}|z^{(d)}), \alpha^{(d)}_{(d)})}{Pr(w_{(d)}|z^{(d)}), \alpha^{(d)}_{(d)})} = \frac{Pr(w_{(d)}|\phi, \alpha^{(d)})}{1}
\]

The proposed AIS scheme:
- Set the initial and final distributions proportional to the posteriors for the two models

\[
\tau_i = Pr(z^{(d)}|\phi, \alpha^{(d)})
\]

A similar argument to the above gives us:

\[
\sum \frac{w^{(i)}}{N} = \sum \frac{Pr(w_{(d)}|z^{(d)}), \alpha^{(d)}_{(d)})}{Pr(w_{(d)}|z^{(d)}), \alpha^{(d)}_{(d)})} = \frac{Pr(w_{(d)}|\phi, \alpha^{(d)})}{1}
\]

**References**