Supplementary Material: A Neural Autoregressive Topic Model

Hugo Larochelle
Département d'informatique
Université de Sherbrooke
hugo.larochelle@usherbrooke.ca

Stanislas Lauly
Département d'informatique
Université de Sherbrooke
stanislas.lauly@usherbrooke.ca

Abstract

We provide additional details on our submission.

1 Pseudocode for computation of $p(\mathbf{v})$ and learning gradients for DocNADE

Input: training document vector $\mathbf{v}$
Output: $p(\mathbf{v})$ and gradients of $-\log p(\mathbf{v})$ on parameters

Computing $p(\mathbf{v})$
$\mathbf{a} \leftarrow \mathbf{c}$
$p(\mathbf{v}) \leftarrow 1$
for i from 1 to D **do**
  $\mathbf{h}_i \leftarrow \mathrm{sigm}(\mathbf{a})$ # for tanh activation, change to $\mathbf{h}_i \leftarrow \tanh(\mathbf{a})$
  $p(v_i = k|\mathbf{v}_{<i}) = 1$
  for m from 1 to $|\boldsymbol{\pi}(v_i)|$ **do**
    $p(v_i = k|\mathbf{v}_{<i}) \leftarrow p(v_i = k|\mathbf{v}_{<i}) p(\pi(v_i)_m|\mathbf{v}_{<i})$
  end for
  $p(\mathbf{v}) \leftarrow p(\mathbf{v}) p(v_i = k|\mathbf{v}_{<i})$
  $\mathbf{a} \leftarrow \mathbf{a} + \mathbf{W}_{:,v_i}$
end for

Computing gradients of $-\log p(\mathbf{v})$
$\delta\mathbf{a} \leftarrow 0$
$\delta\mathbf{c} \leftarrow 0$
$\delta\mathbf{b} \leftarrow 0$
for i from D to 1 **do**
  $\delta\mathbf{h}_i \leftarrow 0$
  for m from 1 to $|\boldsymbol{\pi}(v_i)|$ **do**
    $\delta b_{l(v_i)_m} \leftarrow \delta b_{l(v_i)_m} + (p(\pi(v_i)_m|\mathbf{v}_{<i}) - \pi(v_i)_m)$
    $\delta\mathbf{V}_{l(v_i)_m,:} \leftarrow \delta\mathbf{V}_{l(v_i)_m,:} + (p(\pi(v_i)_m|\mathbf{v}_{<i}) - \pi(v_i)_m)\,\mathbf{h}_i^\top$
    $\delta\mathbf{h}_i \leftarrow \delta\mathbf{h}_i + (p(\pi(v_i)_m|\mathbf{v}_{<i}) - \pi(v_i)_m)\,\mathbf{V}_{l(v_i)_m,:}^\top$
  end for
  $\delta\mathbf{g} \leftarrow \mathbf{h}_i \circ (1 - \mathbf{h}_i)$ # for tanh activation, change to $\delta\mathbf{g} \leftarrow (1 - \mathbf{h}_i^2)$
  $\delta\mathbf{c} \leftarrow \delta\mathbf{c} + \delta\mathbf{h}_i \circ \delta\mathbf{g}$
  $\delta\mathbf{W}_{:,v_i} \leftarrow \delta\mathbf{W}_{:,v_i} + \delta\mathbf{a}$
  $\delta\mathbf{a} \leftarrow \delta\mathbf{a} + \delta\mathbf{h}_i \circ \delta\mathbf{g}$
end for

return $p(\mathbf{v}), \delta\mathbf{b}, \delta\mathbf{V}, \delta\mathbf{c}, \delta\mathbf{W}$

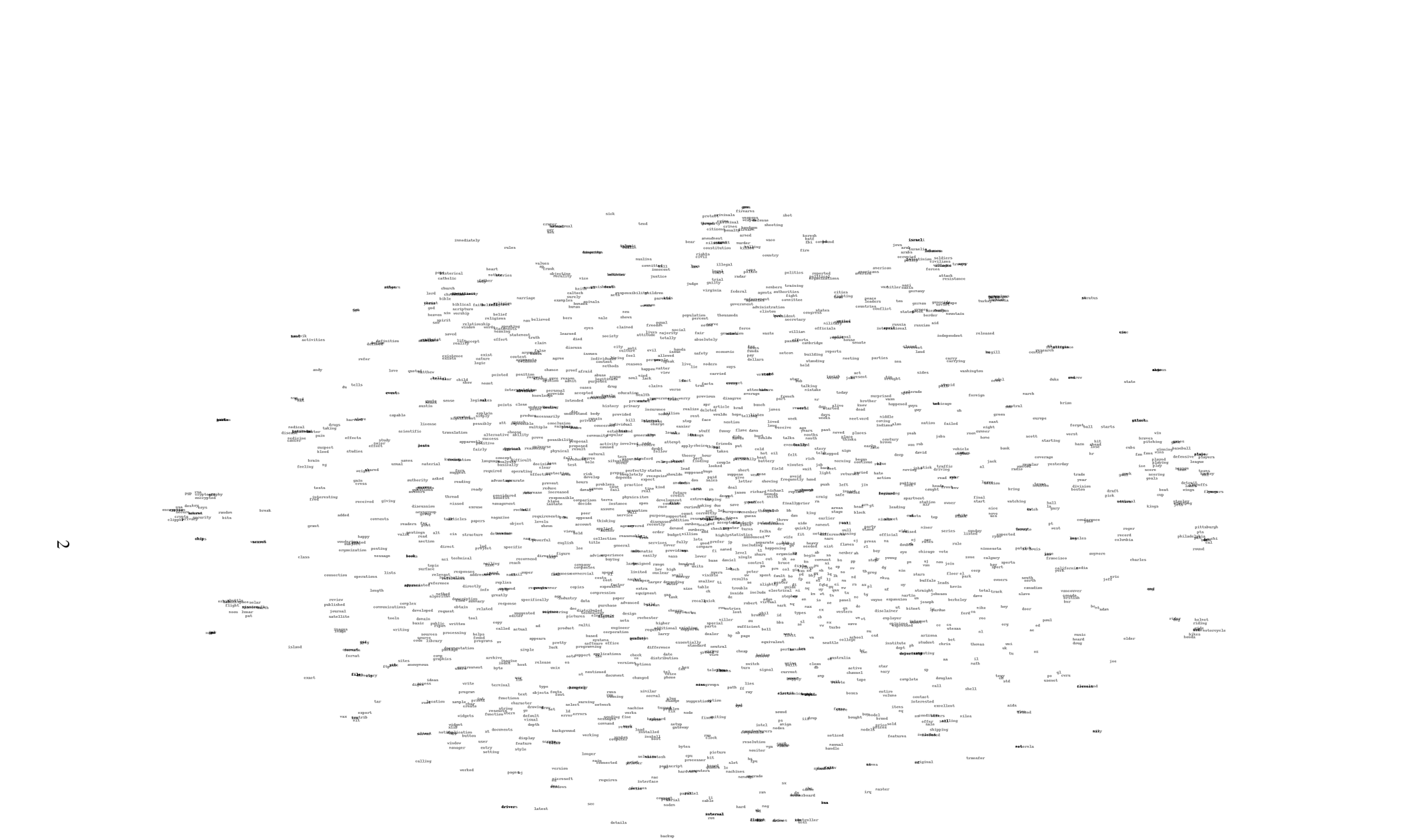

Figure 1: t-SNE visualization of learned word representations from 20 Newsgroups.

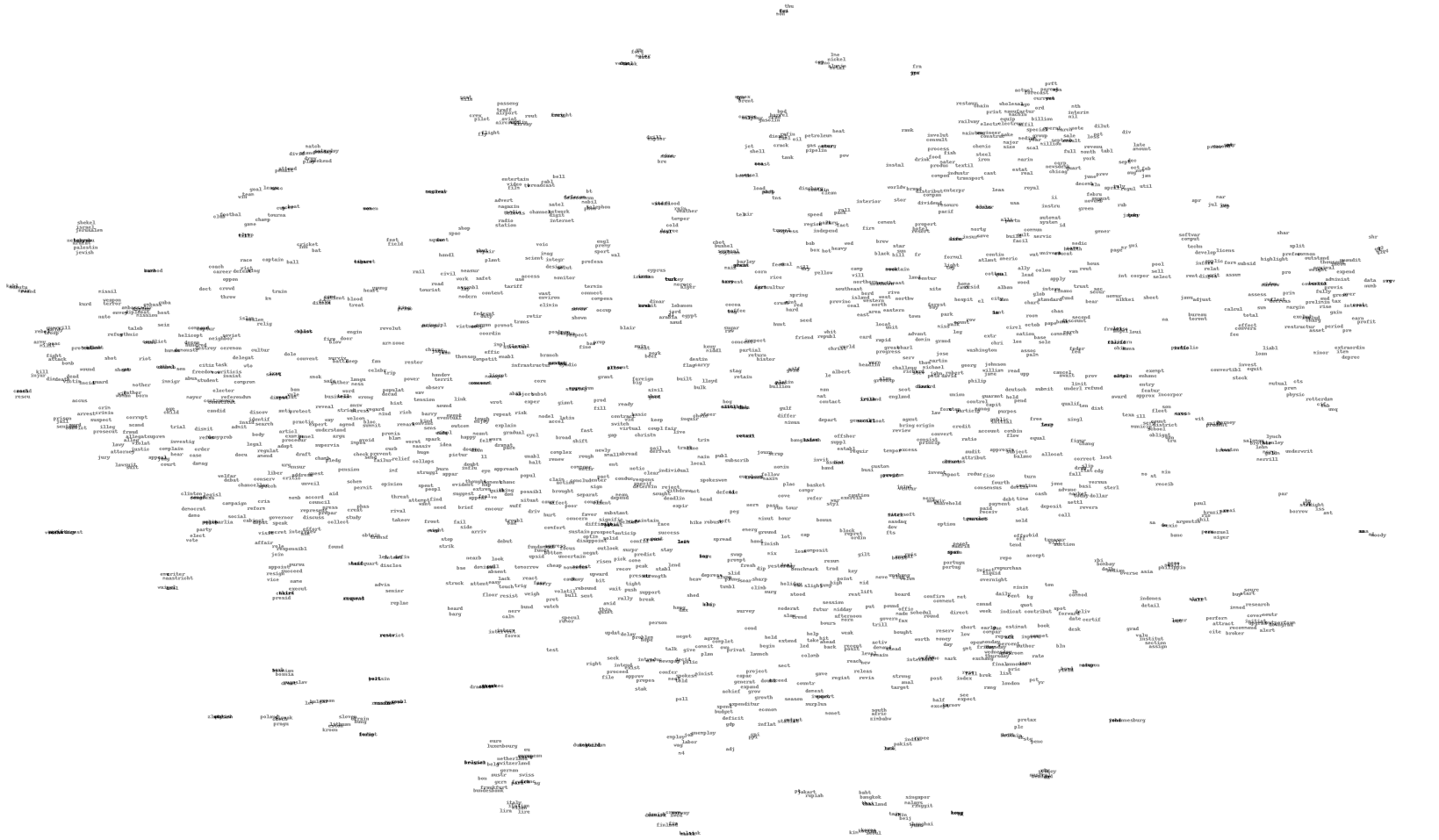

Figure 2: t-SNE visualization of learned word representations from RCV1-v2.