Supplementary: Neural Topic Modeling with Continual Lifelong Learning

Pankaj Gupta [1] **Yatin Chaudhary** [1 2] **Thomas Runkler** [1] **Hinrich Schütze** [2]

A. Data Description

Discussed in section 3, we perform lifelong topic learning over following three streams:
S1: `AGnews` → `TMN` → `R21578` → `20NS` → `20NSshort`
S2: `AGnews` → `TMN` → `R21578` → `20NS` → `TMNtitle`
S3: `AGnews` → `TMN` → `R21578` → `20NS` → `R21578title`

Each stream of document collections consisting of four long-text (high-resource) corpora in sequence: `AGnews`, `TMN`, `R21578` and `20NS` (20NewsGroups), and three short-text (low-resource, sparse) corpora $T+1$ as future (target) tasks $T+1$: `20NSshort`, `TMNtitle` and `R21578title`.

Following is the description of document collections used in this work:

1. `20NSshort`: We take documents from 20NewsGroups data, with document size (number of words) less than 20.
2. `TMNtitle`: Titles of the Tag My News (TMN) news dataset.
3. `R21578title`: Reuters corpus, a collection of new stories from `nltk.corpus`. We take titles of the documents.
4. `TMN`: The Tag My News (TMN) news dataset.
5. `R21578`: Reuters corpus, a collection of new stories from `nltk.corpus`.
6. `AGnews`: AGnews data sellection.
7. `20NS`: 20NewsGroups corpus, a collection of news stories from `nltk.corpus`.

See Table 1 for the description of each of the document collections used in our experiments. Observe that we employ sparse document collections as target datasets.

[1]Corporate Technology, Siemens AG Munich, Germany [2]CIS, University of Munich (LMU) Munich, Germany. Correspondence to: Pankaj Gupta <pankaj.gupta@drimco.net>.

Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119, 2020.

Table 1. Data statistics: Document collections used in lifelong topic modeling. Symbols- K: vocabulary size, L: average text length (#words), C: number of classes and k: thousand. For short-text, L<15. We use $\mathcal{T}^1$, $\mathcal{T}^2$ and $\mathcal{T}^3$ are treated as target corpora for future tasks $T+1$ of topic modeling and $\mathcal{S}^1$-$\mathcal{S}^4$ are used as historical corpora in the stream of document collections.

ID	Data	Train	Val	Test	K	L	C
$\mathcal{T}^1$	20NSshort	1.3k	0.1k	0.5k	1.4k	13.5	20
$\mathcal{T}^2$	TMNtitle	22.8k	2.0k	7.8k	2k	4.9	7
$\mathcal{T}^3$	R21578title	7.3k	0.5k	3.0k	2k	7.3	90
$\mathcal{S}^1$	AGNews	118k	2.0k	7.6k	5k	38	4
$\mathcal{S}^2$	TMN	22.8k	2.0k	7.8k	2k	19	7
$\mathcal{S}^3$	R21578	7.3k	0.5k	3.0k	2k	128	90
$\mathcal{S}^4$	20NS	7.9k	1.6k	5.2k	2k	107.5	20

Table 2. Illustration of Domain-overlap in pairs of corpora, when used in source-target settings. $\mathcal{I}$: Identical, $\mathcal{R}$: Related and $\mathcal{D}$: Distant domains determined based on overlap in labels

	$\mathcal{T}^1$	$\mathcal{T}^2$	$\mathcal{T}^3$
$\mathcal{S}^1$	$\mathcal{R}$	$\mathcal{R}$	$\mathcal{D}$
$\mathcal{S}^2$	$\mathcal{R}$	$\mathcal{I}$	$\mathcal{D}$
$\mathcal{S}^3$	$\mathcal{D}$	$\mathcal{D}$	$\mathcal{I}$
$\mathcal{S}^4$	$\mathcal{I}$	$\mathcal{R}$	$\mathcal{D}$

Table 2 suggests a domain overlap (in terms of labels) among the document collections used in transfer learning within neural topic modeling framework. The notations such as `I`, `R` and `D` represent domain overlap, where `I` (identical): identical-domain in terms of labels in pair of datasets, `R` (related): related-domain due to partial overlap in labels, and `D`: distant-domain due to no overlap in labels of pair of document collections.

See Table 3 for the label information for each of the document collections used in streams of information to model.

To reproduce the scores reported, we have also provided the **code** of the LNTM framework and **pre-processed datasets** used in our experiments.

Table 3. Label space of the document collections used

data	labels / classes
TMN	world, us, sport, business, sci_tech, entertainment, health
TMNtitle	world, us, sport, business, sci_tech, entertainment, health
AGnews	business, sci_tech, sports, world
20NS 20NSshort,	misc.forsale, comp.graphics, rec.autos, comp.windows.x, rec.sport.baseball, sci.space, rec.sport.hockey, soc.religion.christian, rec.motorcycles, comp.sys.mac.hardware, talk.religion.misc, sci.electronics, comp.os.ms-windows.misc, sci.med, comp.sys.ibm.pc.hardware, talk.politics.mideast, talk.politics.guns, talk.politics.misc, alt.atheism, sci.crypt
R21578title R21578	trade, grain, crude, corn, rice, rubber, sugar, palm-oil, veg-oil, ship, coffee, wheat, gold, acq, interest, money-fx, carcass, livestock, oilseed, soybean, earn, bop, gas, lead, zinc, gnp, soy-oil, dlr, yen, nickel, groundnut, heat, sorghum, sunseed, cocoa, rapeseed, cotton, money-supply, iron-steel, palladium, platinum, strategic-metal, reserves, groundnut-oil, lin-oil, meal-feed, sun-meal, sun-oil, hog, barley, potato, orange, soy-meal, cotton-oil, fuel, silver, income, wpi, tea, lei, coconut, coconut-oil, copra-cake, propane, instal-debt, nzdlr, housing, nkr, rye, castor-oil, palmkernel, tin, copper, cpi, pet-chem, rape-oil, oat, naphtha, cpu, rand, alum

Table 4. Hyper-parameters search space in the Generalization task

Hyperparameter	Search Space
retrieval fraction	[0.02]
learning rate	[0.001]
hidden units (#topics), H	[50, 200]
activation function (g)	sigmoid
iterations	[100]
λ_{TR}	[0.1, 0.01, 0.001]
λ_{EmbTF}	[1.0, 0.5, 0.1]
λ_{SAL}	[1.0, 0.5, 0.1]

Table 5. Hyper-parameters search space in the IR task

Hyperparameter	Search Space
retrieval fraction	[0.02]
learning rate	[0.001]
hidden units (#topics), H	[50, 200]
activation function (g)	tanh
iterations	[100]
λ_{TR}	[0.1, 0.01, 0.001]
λ_{EmbTF}	[1.0, 0.5, 0.1]
λ_{SAL}	[1.0, 0.5, 0.1]

B. Reproducibility: Hyper-parameter Settings

In the following sections, we provide the hyper-parameter settings (search space) used to build topic models based on development set.

B.1. Hyper-parameter settings for Generalization

Table 4 provides hyper-parameters search space used within lifelong topic modeling framework for generalization task over lifetime. The models built are used further in extracting topics and computing topic coherence.

B.2. Hyper-parameter settings for IR Task

Table 5 provides hyper-parameters search space used within lifelong topic modeling framework for information retrieval task over lifetime.

B.3. Optimal Configurations of λ^{TR}, λ^{EmbTF}, λ^{SAL}

Tables 6 and 7 provide the optimal (best) hyper-parameter setting for generalization and IR task, respectively for each of the three target datasets. The hyper-parameters corresponds to the scores reported in the paper content.

To reproduce the scores reported, we have also provided the **code** of the LNTM framework and **pre-processed datasets** used in our experiments.

To reproduce the scores reported, we have also provided the

Table 6. Generalization Task: Optimal settings of hyper-parameters (λ_{TR} / λ_{EmbTF} / λ_{SAL}) for each of the three streams where the datasets: `20NSshort`, `TMNtitle` and `R21578title` are treated as targets, respectively in each of the streams. The optimal hyper-parameters are obtained in joint training of three approaches: `TR`, `EmbTF` and `SAL` with the proposed LNTM framework.

Target	**Stream**	**Hyper-parameters (λ_{TR} / λ_{EmbTF} / λ_{SAL}) for Streams of Document Collections**			
		`AGnews`	`TMN`	`R21578`	`20NS`
`20NSshort`	**S1**	0.001 / 0.1 / 1.0	0.001 / 0.1 / 1.0	0.001 / 0.1 / 1.0	0.001 / 1.0 / 1.0
`TMNtitle`	**S2**	0.001 / 0.1 / 0.1	0.001 / 1.0 / 1.0	0.001 / 0.1 / 0.1	0.001 / 0.1 / 0.1
`R21578title`	**S3**	0.001 / 0.1 / 0.1	0.001 / 0.1 / 0.1	0.001 / 1.0 / 0.1	0.1 / 0.1 / 0.1

Table 7. IR Task: Optimal settings of hyper-parameters (λ_{TR} / λ_{EmbTF} / λ_{SAL}) for each of the three streams where the datasets: `20NSshort`, `TMNtitle` and `R21578title` are treated as targets, respectively in each of the streams. The optimal hyper-parameters are obtained in joint training of three approaches: `TR`, `EmbTF` and `SAL` with the proposed LNTM framework.

Target	**Stream**	**Hyper-parameters (λ_{TR} / λ_{EmbTF} / λ_{SAL}) for Streams of Document Collections**			
		`AGnews`	`TMN`	`R21578`	`20NS`
`20NSshort`	**S1**	0.001 / 0.1 / 1.0	0.001 / 0.1 / 1.0	0.001 / 0.1 / 1.0	0.001 / 1.0 / 1.0
`TMNtitle`	**S2**	0.001 / 0.1 / 0.1	0.01 / 1.0 / 1.0	0.001 / 0.1 / 0.1	0.001 / 0.1 / 0.1
`R21578title`	**S3**	0.001 / 0.1 / 1.0	0.001 / 0.1 / 1.0	0.001 / 1.0 / 1.0	0.1 / 0.1 / 1.0

code of the LNTM framework and **pre-processed datasets** used in our experiments.