SIEMENS

Neural Topic Modeling with Continual Lifelong Learning

(TLDR: Improving Unsupervised document-topic modeling within transfer learning and lifelong learning paradigm)

Presenter: **Dr. Pankaj Gupta** | ICML 2020 July, 12-18

Affiliation(s): Lead Research Scientist (ML for NLP), Siemens AG Munich Germany

Co-authors: [1]Pankaj Gupta, [1,2]Yatin Chaudhary, [1]Thomas Runkler, [2]Hinrich Schütze

[1]Corporate Technology, Siemens AG Munich,

[2]CIS, University of Munich (LMU) Munich, Germany

Content

SIEMENS
Ingenuity for life

Motivation: Lifelong Neural Topic Modeling

Background: Neural Topic Model (DocNADE) for Text Documents

Methodology: Lifelong Learning in Neural Topic Modeling over Streaming Documents

- ❑ **Topic Regularization** with TopicPool (TR)
- ❑ **Transfer Learning** with WordPool (EmbTF)
- ❑ **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training

Experiments and Analysis on Future Tasks in Sparse settings

- ❑ Document Representation via Information Retrieval (IR)
- ❑ Topic Quality via Coherence (COH)

Unrestricted © Siemens AG 2020

Content

SIEMENS
Ingenuity for life

Motivation: Lifelong Neural Topic Modeling

Background: Neural Topic Model (DocNADE) for Text Documents

Methodology: Lifelong Learning in Neural Topic Modeling over Streaming Documents

- ❑ **Topic Regularization** with TopicPool (TR)
- ❑ **Transfer Learning** with WordPool (EmbTF)
- ❑ **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training

Experiments and Analysis on Future Tasks in Sparse settings

- ❑ Document Representation via Information Retrieval (IR)
- ❑ Topic Quality via Coherence (COH)

Unrestricted © Siemens AG 2020

Content

SIEMENS
Ingenuity for life

Motivation: Lifelong Neural Topic Modeling

Background: Neural Topic Model (DocNADE) for Text Documents

Methodology: Lifelong Learning in Neural Topic Modeling over Streaming Documents

- ❑ **Topic Regularization** with TopicPool (TR)
- ❑ **Transfer Learning** with WordPool (EmbTF)
- ❑ **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training

Experiments and Analysis on Future Tasks in Sparse settings

- ❑ Document Representation via Information Retrieval (IR)
- ❑ Topic Quality via Coherence (COH)

Unrestricted © Siemens AG 2020

Content

SIEMENS
Ingenuity for life

Motivation: Lifelong Neural Topic Modeling

Background: Neural Topic Model (DocNADE) for Text Documents

Methodology: Lifelong Learning in Neural Topic Modeling over Streaming Documents

- ❑ **Topic Regularization** with TopicPool (TR)
- ❑ **Transfer Learning** with WordPool (EmbTF)
- ❑ **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training

Experiments and Analysis on Future Tasks in Sparse settings

- ❑ Document Representation via Information Retrieval (IR)
- ❑ Topic Quality via Coherence (COH)

Unrestricted © Siemens AG 2020

Content

SIEMENS
Ingenuity for life

Motivation: Lifelong Neural Topic Modeling

Background: Neural Topic Model (DocNADE) for Text Documents

Methodology: Lifelong Learning in Neural Topic Modeling over Streaming Documents

- ❑ **Topic Regularization** with TopicPool (TR)
- ❑ **Transfer Learning** with WordPool (EmbTF)
- ❑ **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training

Experiments and Analysis on Future Tasks in Sparse settings

- ❑ Document Representation via Information Retrieval (IR)
- ❑ Topic Quality via Coherence (COH)

Unrestricted © Siemens AG 2020

Lifelong Machine Learning (LML)

SIEMENS
Ingenuity for life

Build computational learning systems that can:

- ❑ Continually **acquire** knowledge from many **tasks** from several **domains**
- ❑ **Retain** and **transfer** the acquired knowledge over **lifetime**

Processing continuous streams of information →

Key Components of LML

Knowledge Mining
Transfer Learning
Domain Adaptation
Knowledge Retention
Catastrophic Forgetting
Knowledge Accumulation

Lifelong Machine Learning (LML)

SIEMENS
Ingenuity for life

Build computational learning systems that can:

- ❑ Continually **acquire** knowledge from many **tasks** from several **domains**
- ❑ **Retain** and **transfer** the acquired knowledge over **lifetime**

Processing continuous streams of information →

Key Components of LML

Knowledge Mining
Transfer Learning
Domain Adaptation
Knowledge Retention
Catastrophic Forgetting
Knowledge Accumulation

Challenging: Continuous non-stationary data distributions or domain shifts

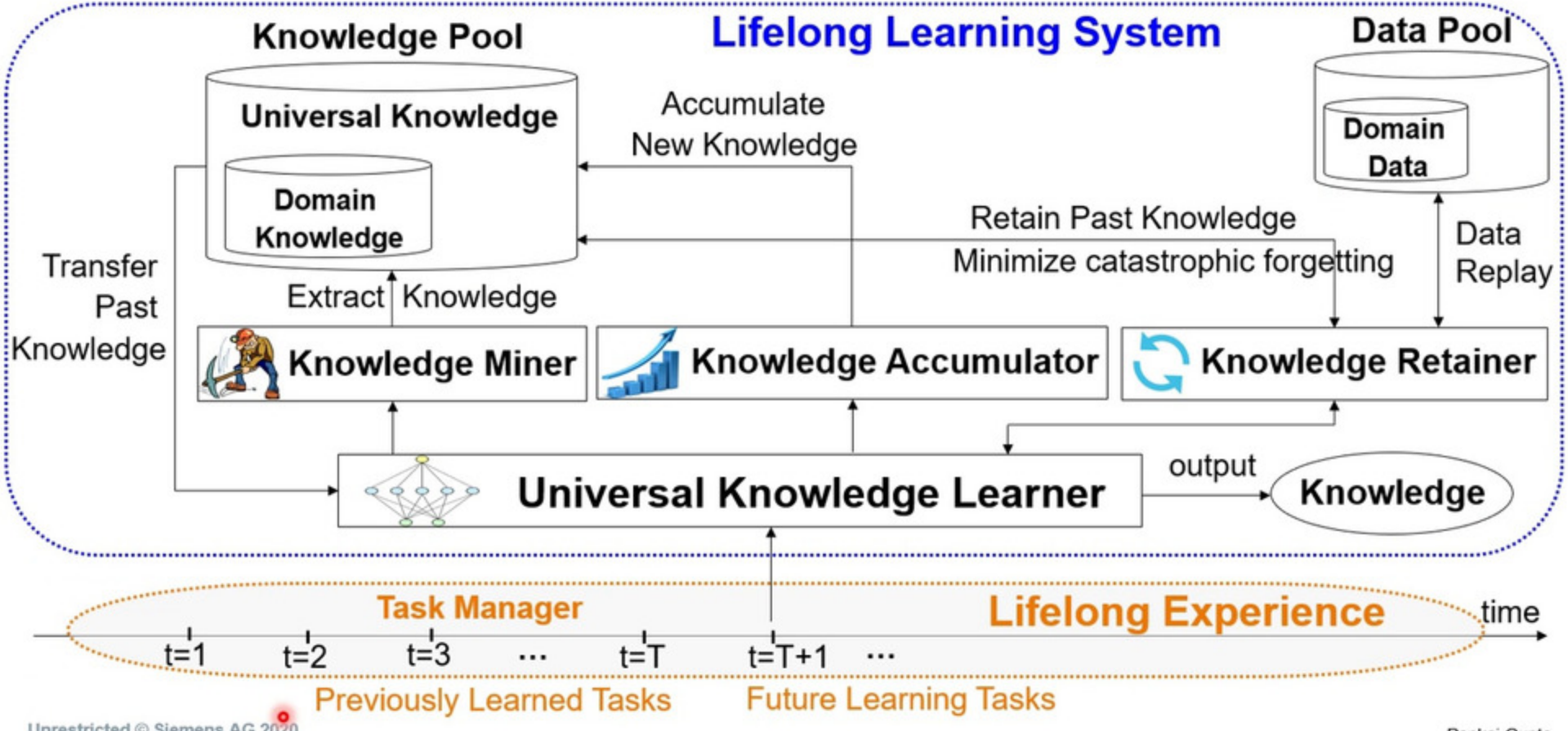
→ catastrophic forgetting or interference

This work: **Lifelong Learning in Unsupervised Document Topic Modeling**

Unrestricted © Siemens AG 2020

Lifelong Machine Learning (LML)

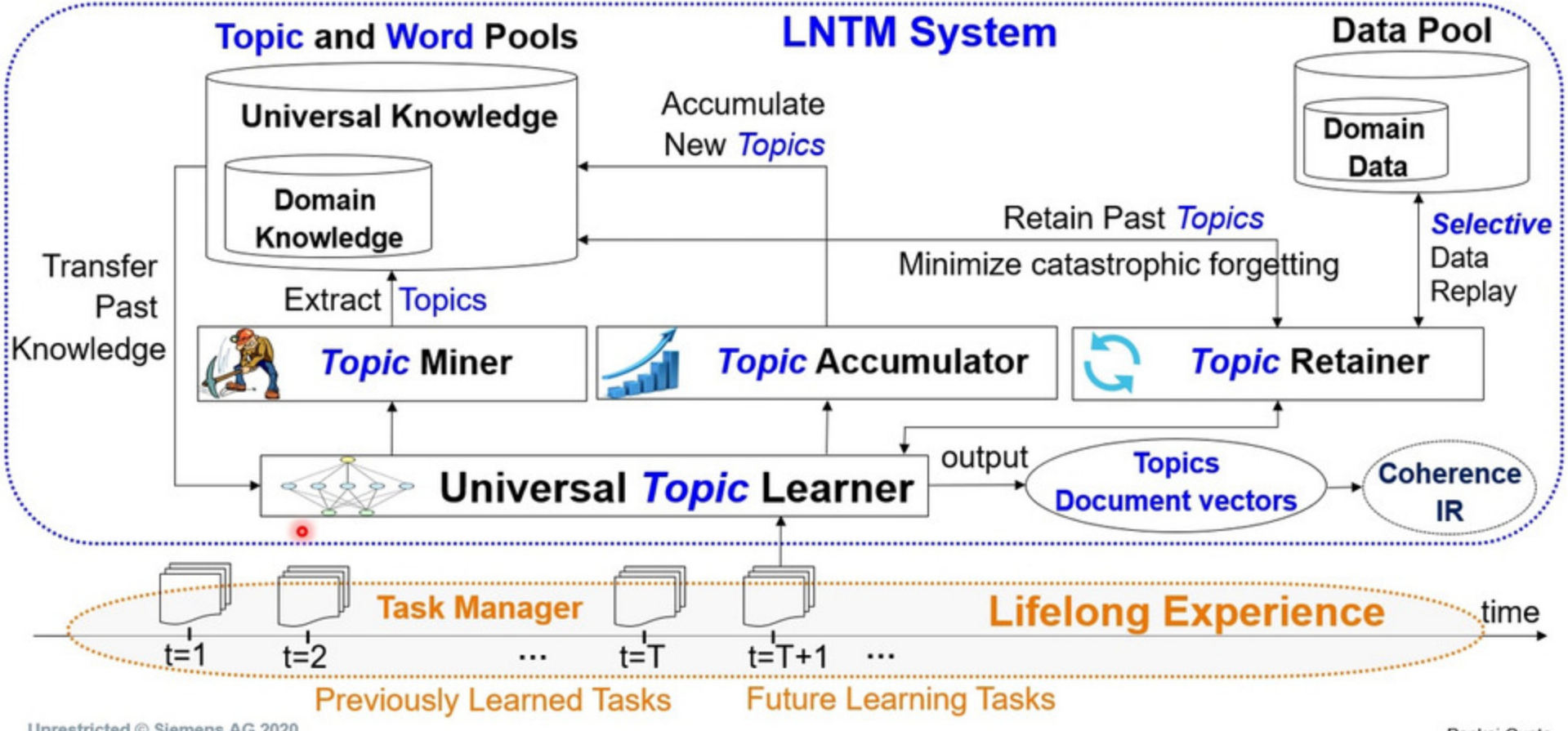
SIEMENS
Ingenuity for life



Unrestricted © Siemens AG 2020

Lifelong Neural Topic Modeling (LNTM)

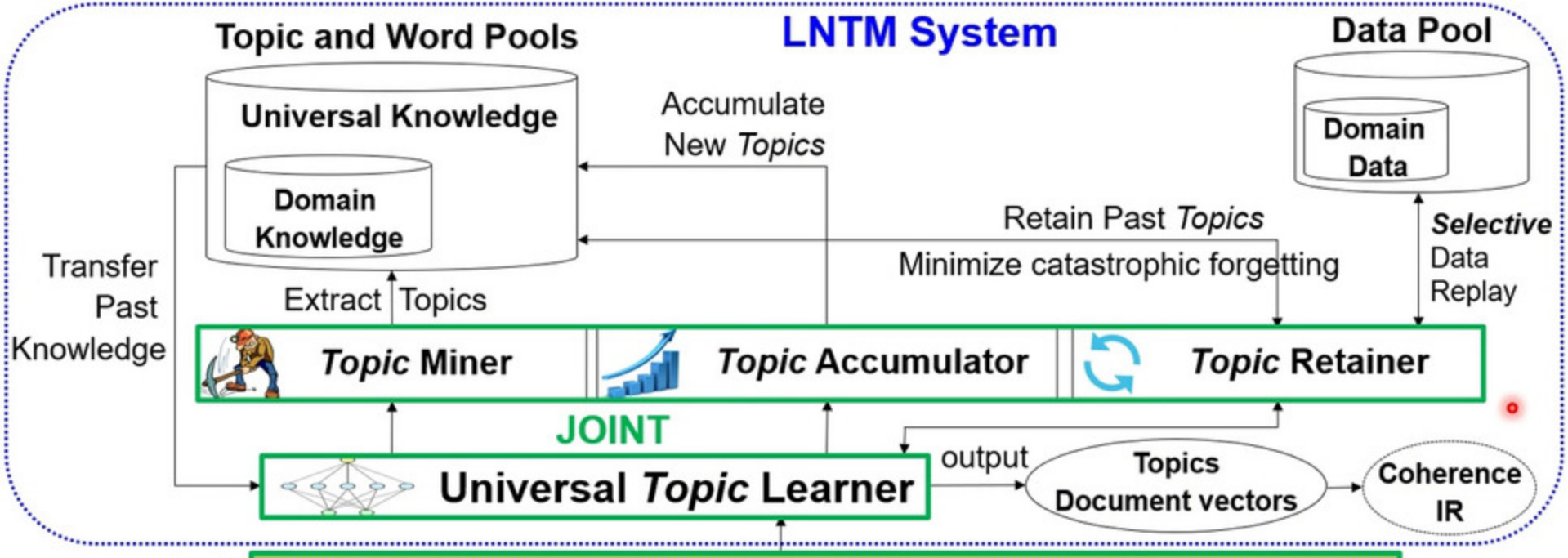
SIEMENS
Ingenuity for life

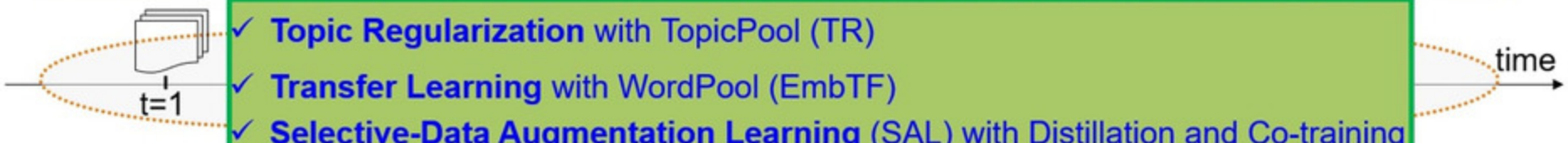


Unrestricted © Siemens AG 2020

Lifelong Neural Topic Modeling (LNTM)

SIEMENS
Ingenuity for life





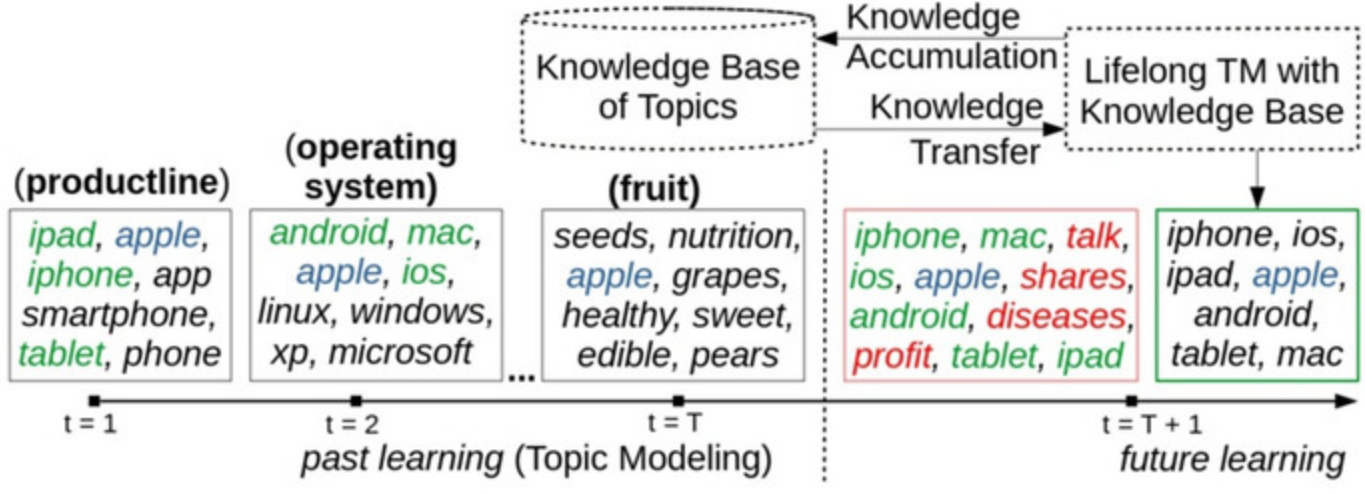
- ✓ **Topic Regularization** with TopicPool (TR)
- ✓ **Transfer Learning** with WordPool (EmbTF)
- ✓ **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training

Unrestricted © Siemens AG 2020

Why Lifelong Neural Topic Modeling?

SIEMENS
Ingenuity for life

- **Continuously** analyze plethora of **unsupervised streaming texts**
- Accumulate and Build **concepts/semantics** over time
- Improve **sparse data modeling** with **knowledge-augmented** topic learning over lifetime
- Improve **topic quality** and **document representations (IR)** in **sparsity** settings



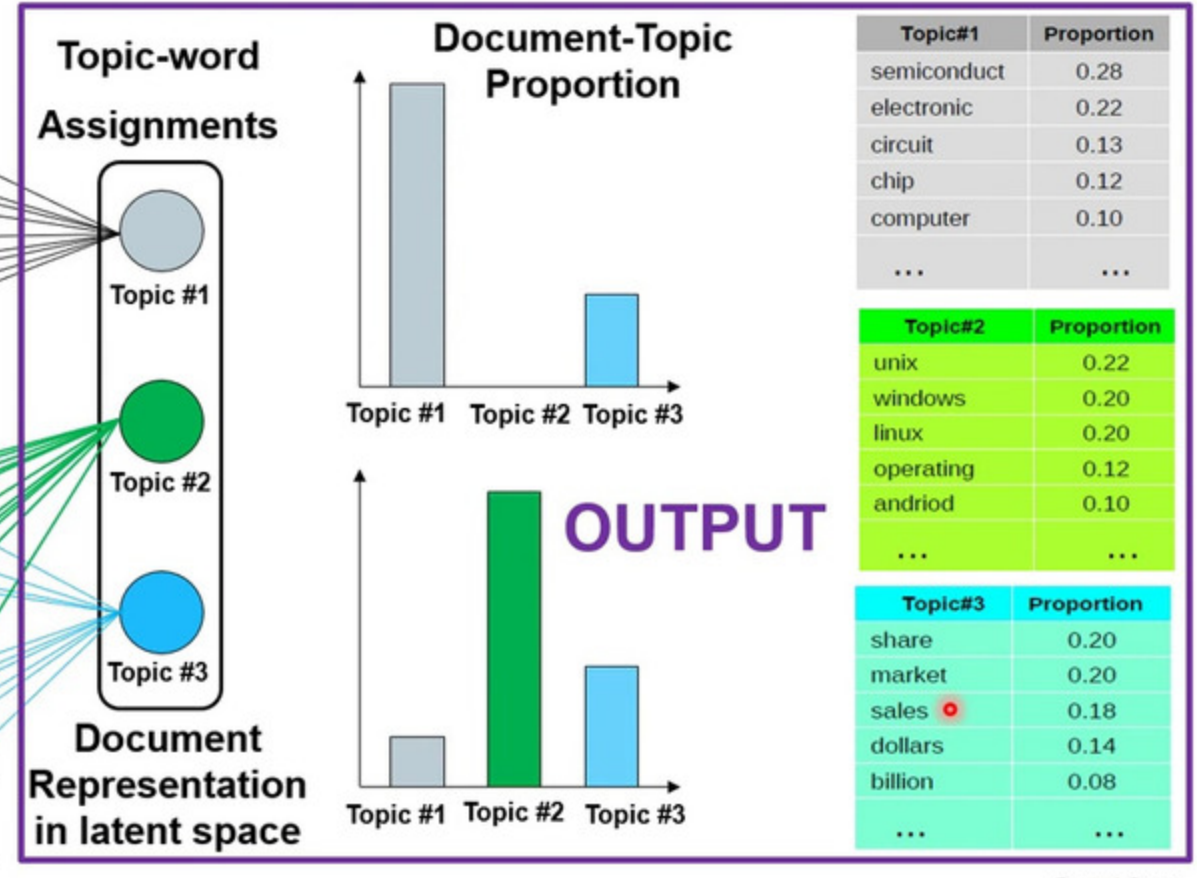
Background: Topic Modeling

SIEMENS
Ingenuity for life

A chip is an integrated circuit, an assembly of electronic components that are used for computer processors, memory, etc. There are transistors and diodes, capacitors and resistors made of semiconductor material.

Semiconductor is a crucial component of electronic devices, e.g., computer. In 2018, the global semiconductor industry sales reached 477.94 billion dollars, where Samsung Electronics holds the largest market share.

Linux, Unix…Windows…computer operating systems. Linux…Windows, and popular a operating system for supercomputers. Android…Linux kernel…operating system…smartphones…billion dollars…financial quarter …smartphones…Android…sales… market share…



Topic#1	Proportion
semiconduct	0.28
electronic	0.22
circuit	0.13
chip	0.12
computer	0.10
…	…

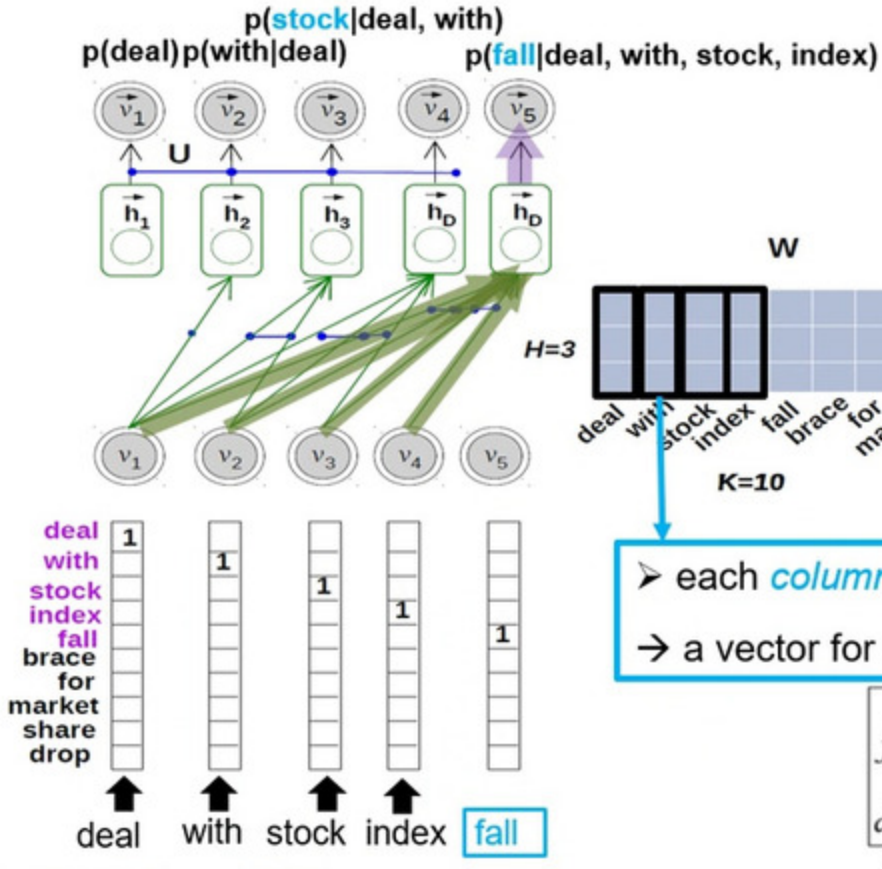
Topic#2	Proportion
unix	0.22
windows	0.20
linux	0.20
operating	0.12
andriod	0.10
…	…

Topic#3	Proportion
share	0.20
market	0.20
sales	0.18
dollars	0.14
billion	0.08
…	…

Unrestricted © Siemens AG 2020

Background: Neural Autoregressive Topic Model (DocNADE)

SIEMENS
Ingenuity for life

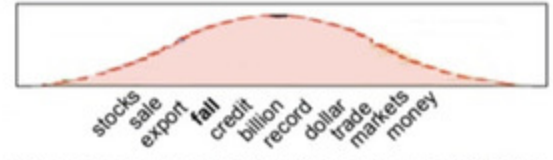


S1: ***Deal with stock index fall***
S2: *Brace for market share drop*

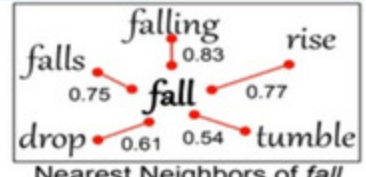
Corpus of two documents

➢ each *row-vector* $\mathbf{W}_{j,:}$
→ encodes the jth topic

➢ each *column-vector* $\mathbf{W}_{:,vi}$
→ a vector for the word at v_i index



Topic due to Word distributions *across* documents



Nearest Neighbors of *fall*

Unrestricted © Siemens AG 2020

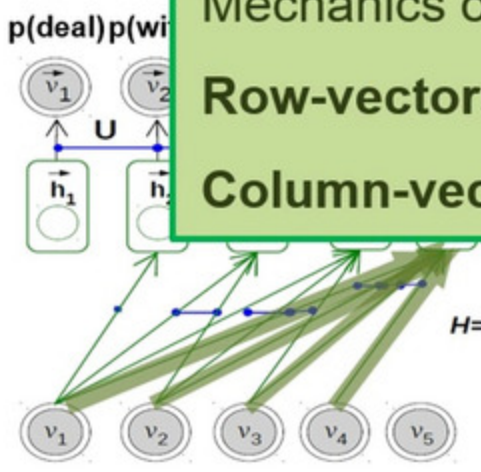
Background: Neural Autoregressive Topic Model (DocNADE)

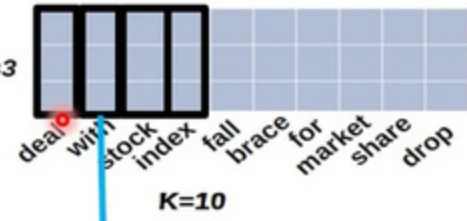
SIEMENS
Ingenuity for life

Mechanics of Knowledge Transfer via:

Row-vectors of W → Topic Embedding Transfer

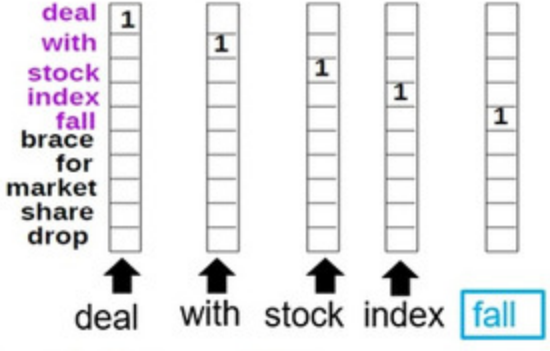
Column-vectors of W → Word Embedding Transfer





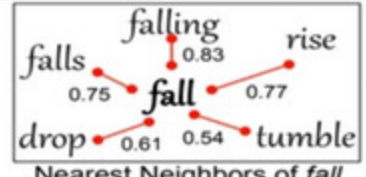
- each *row-vector* $\mathbf{W}_{j,:}$

→ encodes the jth topic

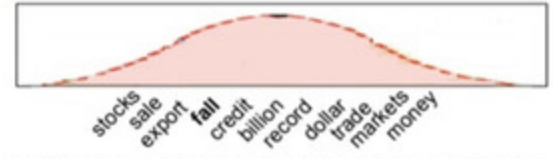


- each *column-vector* $\mathbf{W}_{:,vi}$

→ a vector for the word at v_i index



Nearest Neighbors of *fall*

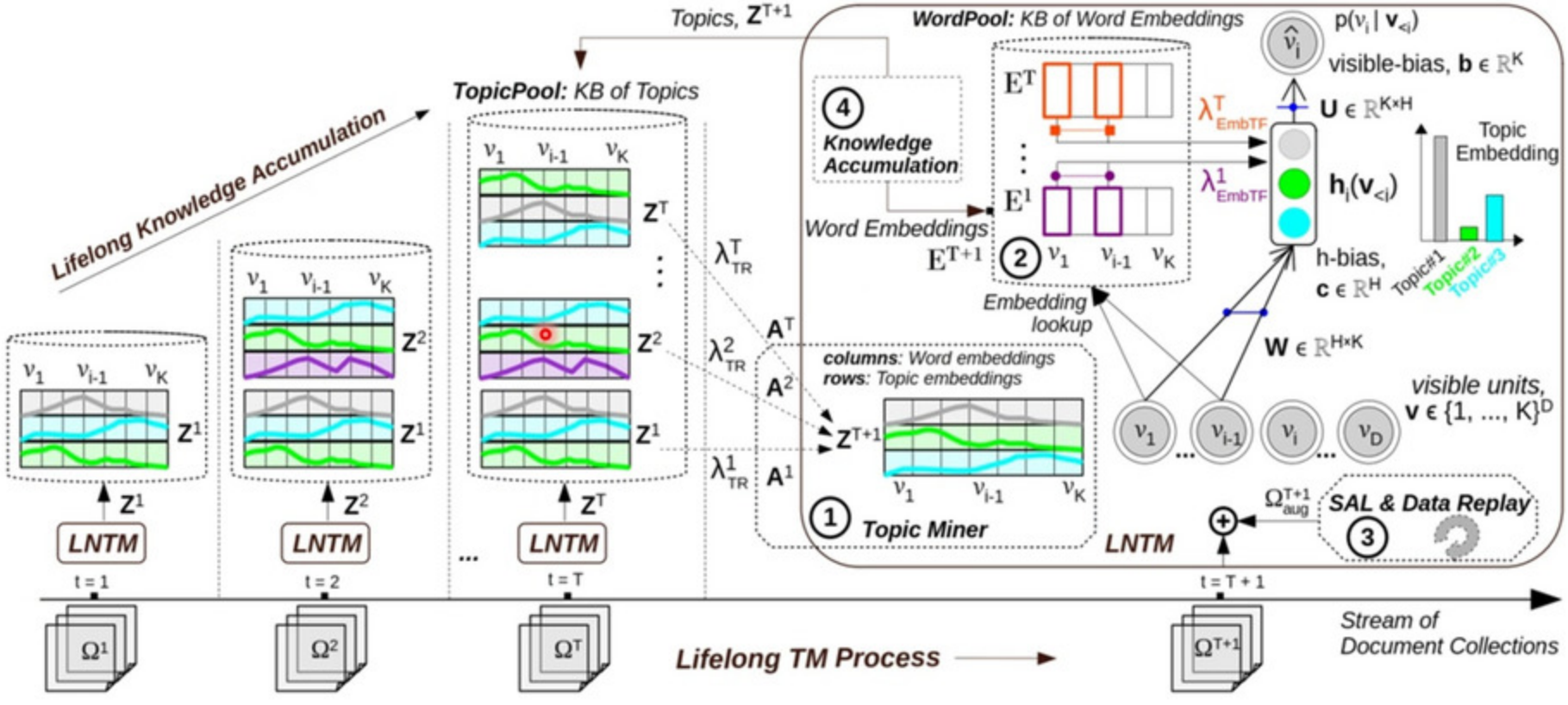


Topic due to Word distributions *across* documents

Unrestricted © Siemens AG 2020

Proposed: Lifelong Neural Topic Modeling (LNTM) framework

SIEMENS
Ingenuity for life



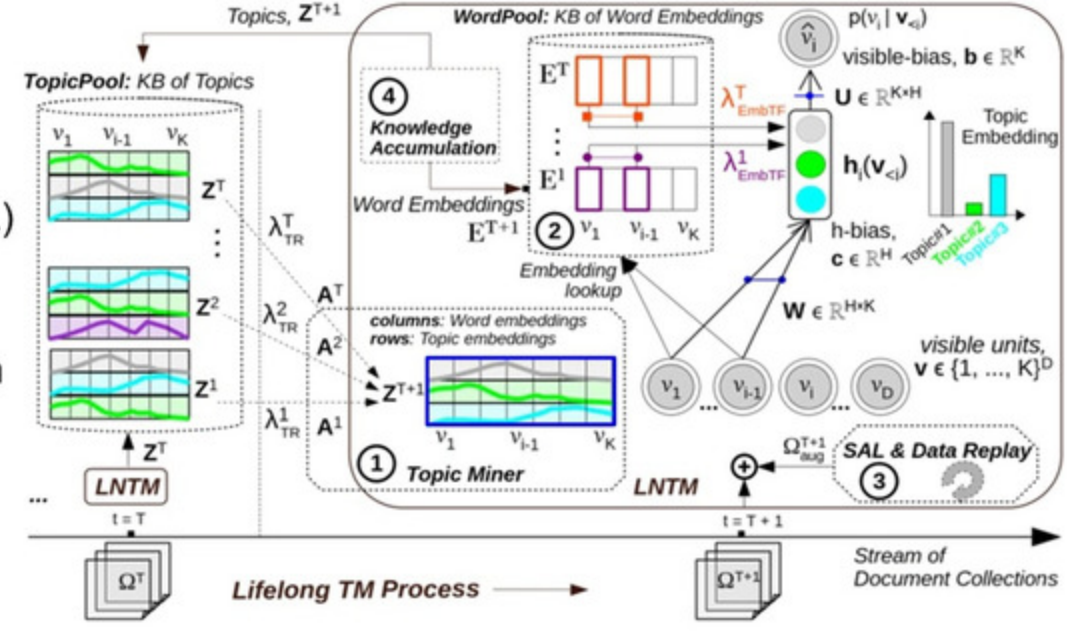
Proposed: Lifelong Neural Topic Modeling (LNTM) framework

SIEMENS
Ingenuity for life

Phase 1: Joint Topic Mining, Transfer and Retention:

1. Topic Regularization with TopicPool (TR)
   1.1 *Topic Extractor*
   1.2 *Topic Regularizer*
2. Transfer Learning with WordPool (EmbTF)
3. Selective-Data Augmentation Learning (SAL)
   3.1 *Document Distillation*
   3.2 *Selective Co-training*

Phase 2: Lifelong Knowledge Accumulation



Proposed: Lifelong Neural Topic Modeling (LNTM) framework

SIEMENS
Ingenuity for life

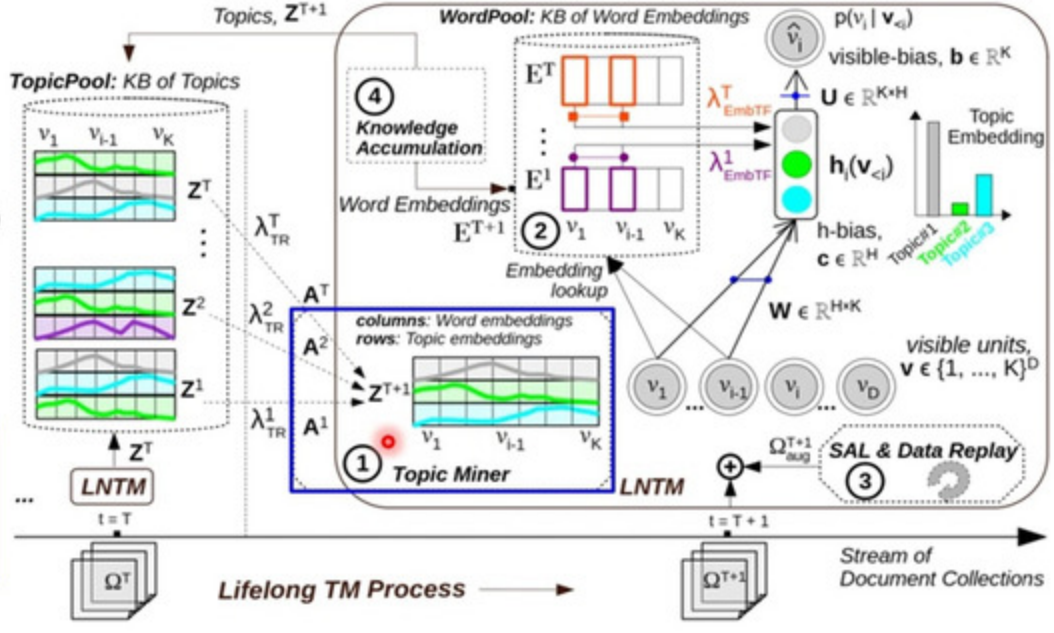
Phase 1: Joint Topic Mining, Transfer and Retention:

1. Topic Regularization with TopicPool (TR)
   1.1 *Topic Extractor*
   1.2 *Topic Regularizer*
2. Transfer Learning with WordPool (EmbTF)
3. Selective-Data Augmentation Learning (SAL)
   3.1 *Document Distillation*
   3.2 *Selective Co-training*

Phase 2: Lifelong Knowledge Accumulation

$$\Delta_{TR} = \sum_{t=1}^{T} \lambda_{TR}^{t} \Big(\underbrace{\|\mathbf{Z}^t - \mathbf{A}^t \mathbf{Z}^{T+1}\|_2^2}_{topic-imitation} + \underbrace{\|\mathbf{U}^t - \mathbf{P}^t \mathbf{U}\|_2^2}_{decoder-proximity} \Big)$$

$$\mathcal{L}(\Omega^{T+1}; \Theta^{T+1}) = \sum_{\mathbf{v} \in \Omega^{T+1}} \mathcal{L}(\mathbf{v}; \Theta^{T+1}) + \Delta_{TR}$$



Proposed: Lifelong Neural Topic Modeling (LNTM) framework

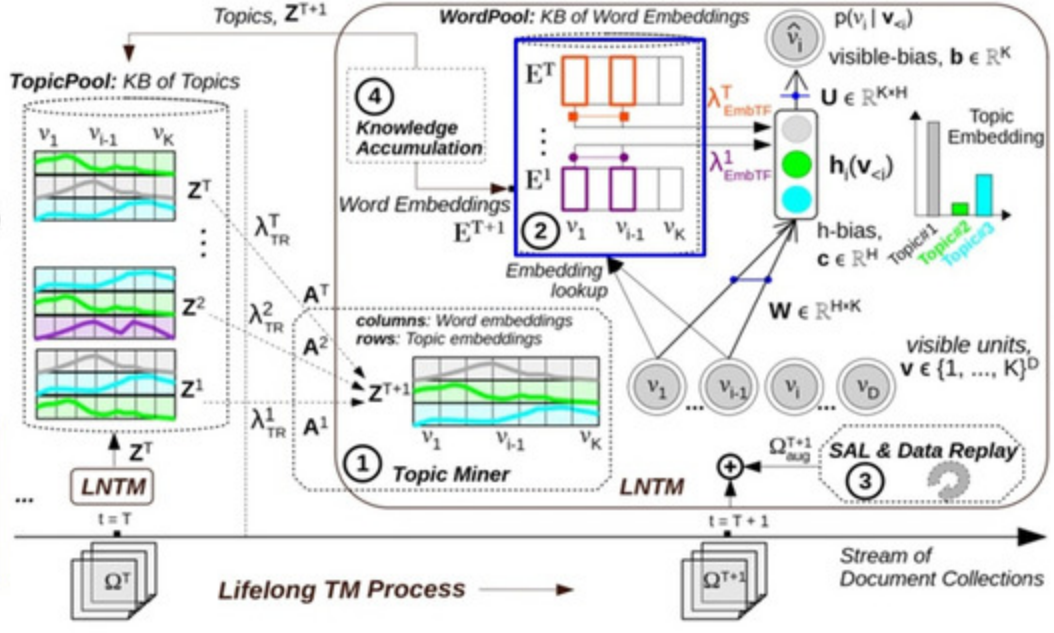
SIEMENS
Ingenuity for life

Phase 1: Joint Topic Mining, Transfer and Retention:

1. Topic Regularization with TopicPool (TR)
   1.1 *Topic Extractor*
   1.2 *Topic Regularizer*
2. Transfer Learning with WordPool (EmbTF)
3. Selective-Data Augmentation Learning (SAL)
   3.1 *Document Distillation*
   3.2 *Selective Co-training*

Phase 2: Lifelong Knowledge Accumulation

$$\mathbf{h}(\mathbf{v}_{<i}) = g(\mathbf{c} + \sum_{q<i} \mathbf{W}_{:,v_q} + \sum_{q<i} \sum_{t=1}^{T} \lambda^t_{EmbTF}\, \mathbf{E}^t_{:,v_q})$$



Proposed: Lifelong Neural Topic Modeling (LNTM) framework

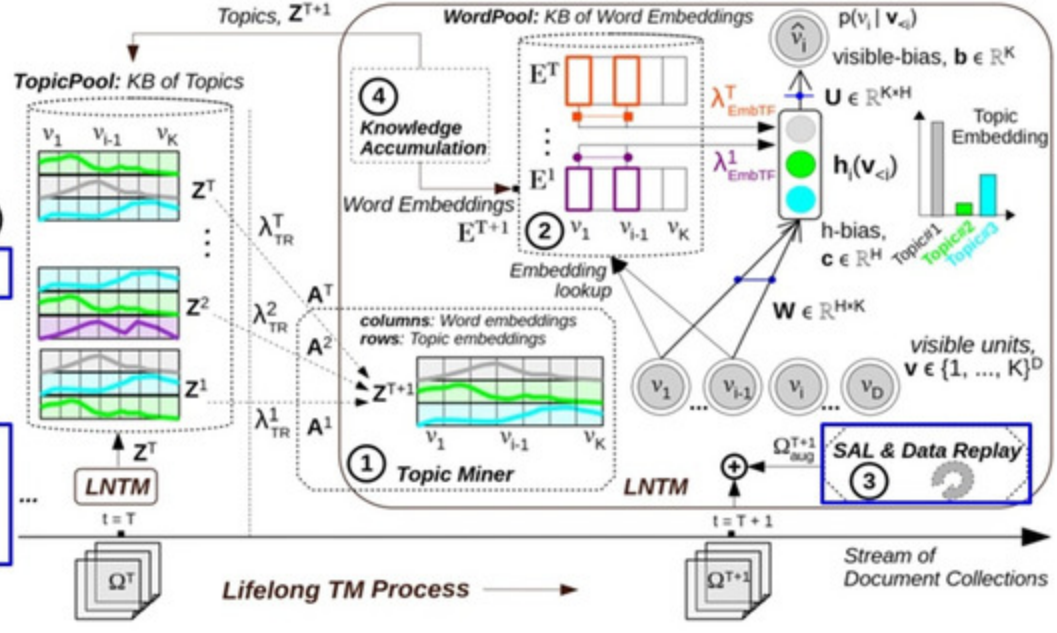
SIEMENS
Ingenuity for life

Phase 1: Joint Topic Mining, Transfer and Retention:

1. Topic Regularization with TopicPool (TR)
   1.1 *Topic Extractor*
   1.2 *Topic Regularizer*
2. Transfer Learning with WordPool (EmbTF)
3. Selective-Data Augmentation Learning (SAL)
   3.1 *Document Distillation*
   3.2 *Selective Co-training*

Phase 2: Lifelong Knowledge Accumulation

$$\textbf{if } \mathrm{PPL}(\mathbf{v}^t; \boldsymbol{\Theta}^{T+1}) \leq \mathrm{PPL}^{T+1} \textbf{ then}$$
$$\text{Document selected: } \Omega_{aug}^{T+1} \leftarrow \Omega_{aug}^{T+1} \cup (\mathbf{v}^t, t)$$



Proposed: Lifelong Neural Topic Modeling (LNTM) framework

SIEMENS
Ingenuity for life

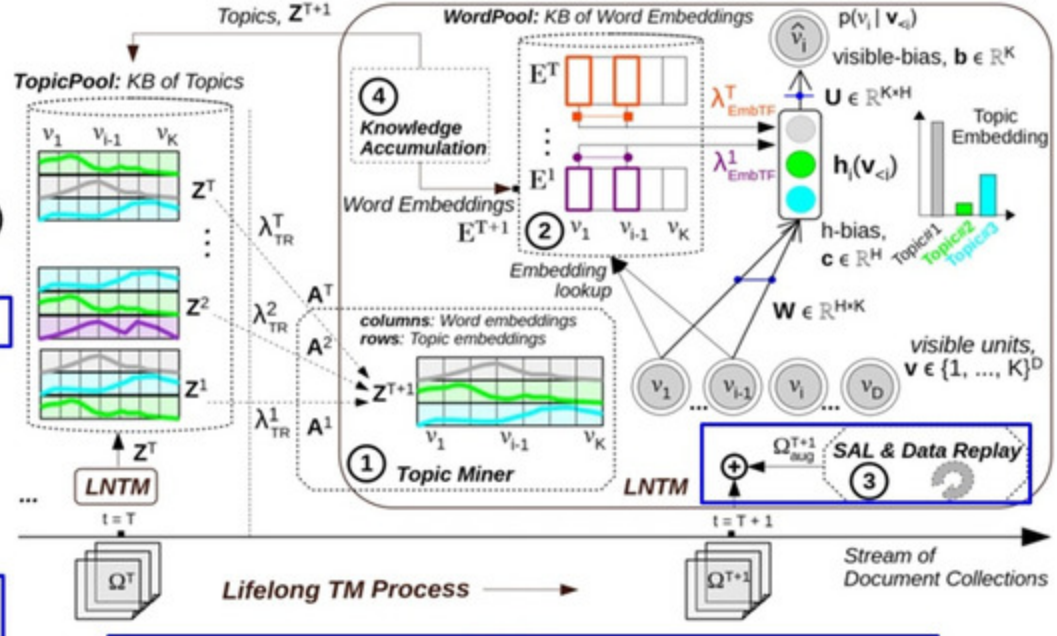
Phase 1: Joint Topic Mining, Transfer and Retention:

1. Topic Regularization with TopicPool (TR)
   1.1 *Topic Extractor*
   1.2 *Topic Regularizer*
2. Transfer Learning with WordPool (EmbTF)
3. Selective-Data Augmentation Learning (SAL)
   3.1 *Document Distillation*
   3.2 *Selective Co-training*

Phase 2: Lifelong Knowledge Accumulation

if $\mathrm{PPL}(\mathbf{v}^t; \boldsymbol{\Theta}^{T+1}) \leq \mathrm{PPL}^{T+1}$ **then**
Document selected: $\Omega_{aug}^{T+1} \leftarrow \Omega_{aug}^{T+1} \cup (\mathbf{v}^t, t)$

$$\Delta_{SAL} = \sum_{(\mathbf{v}^t, t) \in \Omega_{aug}^{T+1}} \lambda_{SAL}^t \, \mathcal{L}(\mathbf{v}^t; \boldsymbol{\Theta}^{T+1})$$



$$\mathcal{L}(\Omega^{T+1}; \boldsymbol{\Theta}^{T+1}) = \sum_{\mathbf{v} \in \Omega^{T+1}} \mathcal{L}(\mathbf{v}; \boldsymbol{\Theta}^{T+1}) + \Delta_{SAL}$$

Unrestricted © Siemens AG 2020

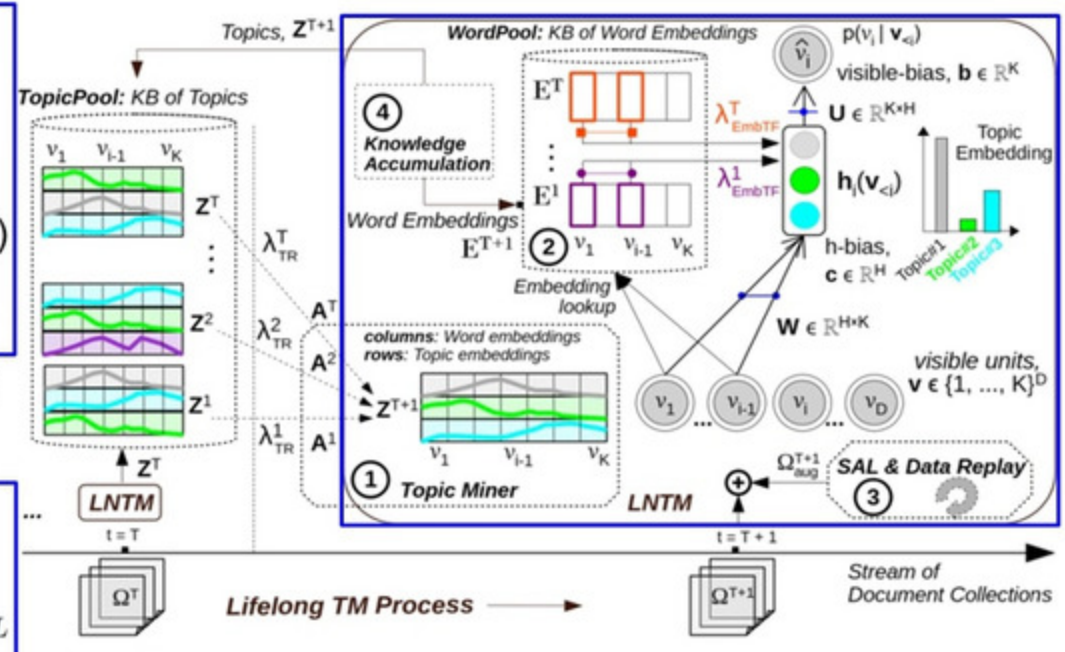
Proposed: Lifelong Neural Topic Modeling (LNTM) framework

SIEMENS
Ingenuity for life

Phase 1: Joint Topic Mining, Transfer and Retention:

1. Topic Regularization with TopicPool (TR)
   1.1 *Topic Extractor*
   1.2 *Topic Regularizer*
2. Transfer Learning with WordPool (EmbTF)
3. Selective-Data Augmentation Learning (SAL)
   3.1 *Document Distillation*
   3.2 *Selective Co-training*

Phase 2: Lifelong Knowledge Accumulation



Overall loss in LNTM framework

$$\mathcal{L}(\Omega^{T+1}; \Theta^{T+1}) = \sum_{\mathbf{v} \in \Omega^{T+1}} \mathcal{L}(\mathbf{v}; \Theta^{T+1}) + \Delta_{TR} + \Delta_{SAL}$$

Unrestricted © Siemens AG 2020

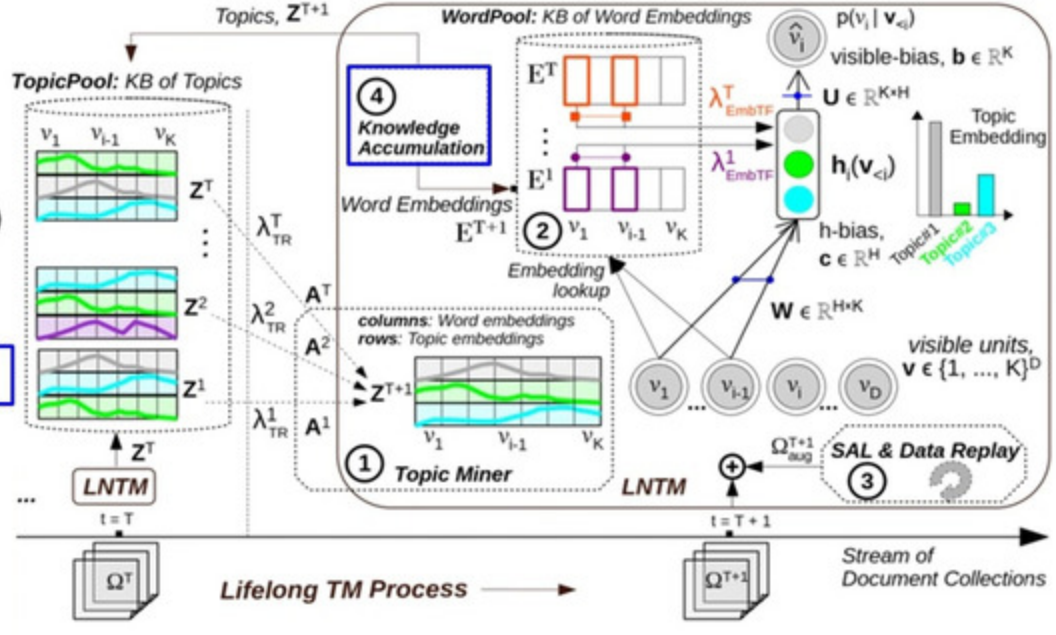
Proposed: Lifelong Neural Topic Modeling (LNTM) framework

SIEMENS
Ingenuity for life

Phase 1: Joint Topic Mining, Transfer and Retention:

1. Topic Regularization with TopicPool (TR)
   1.1 *Topic Extractor*
   1.2 *Topic Regularizer*
2. Transfer Learning with WordPool (EmbTF)
3. Selective-Data Augmentation Learning (SAL)
   3.1 *Document Distillation*
   3.2 *Selective Co-training*

Phase 2: Lifelong Knowledge Accumulation



Experiments and Analysis

SIEMENS
Ingenuity for life

Three Streams of Document Collections

Future learning tasks

S1: AGnews → TMN → R21578 → 20NS → 20NSshort
S2: AGnews → TMN → R21578 → 20NS → TMNtitle
S3: AGnews → TMN → R21578 → 20NS → R21578title

Evaluation

→ Topic Quality via Coherence (**COH**)
→ Information Retrieval via Precision@Recall (**P@R**)
→ Generalization via Perplexity (**PPL**)

Results: PPL, P@R and COH (Max. Knowledge Transfer + Min. Forgetting)

SIEMENS
Ingenuity for life

S1: AGnews → TMN → R21578 → 20NS → 20NSshort

	Scores on historical data incurring Catastrophic Forgetting								Scores with Lifelong Knowledge Transfer					
	PPL	P@0.02	PPL	P@0.02	PPL	P@0.02	PPL	P@0.02	PPL	P@5	P@10	P@0.02	COH	r-time (second)
LNTM + EmbTF + TR + SAL	471	0.786	696	0.647	380	0.724	541	0.248	**641**	**0.418**	**0.375**	**0.324**	**0.735**	50.1
LNTM + EmbTF + TR	469	0.786	698	0.649	382	0.724	541	0.247	672	0.376	0.341	0.297	0.728	2.80
LNTM + TR	466	0.786	690	0.649	382	0.724	538	0.249	655	0.380	0.345	0.306	0.719	2.76
LNTM + EmbTF	468	0.786	696	0.649	383	0.724	540	0.247	647	0.386	0.348	0.310	0.670	2.37
NTM without Lifelong Learning	454	0.785	584	0.651	311	0.726	470	0.268	646	0.375	0.335	0.290	0.667	1.98

AGnews ***TMN*** ***R21578*** ***20NS*** ***20NSshort*** (Sparse Data)

Lifelong Neural Topic Modeling over a stream of document collections →

Unrestricted © Siemens AG 2020

Results: PPL, P@R and COH (Max. Knowledge Transfer + Min. Forgetting)

SIEMENS
Ingenuity for life

S2: AGnews → TMN → R21578 → 20NS → TMNtitle

	Scores on historical data incurring Catastrophic Forgetting								Scores with Lifelong Knowledge Transfer					
	PPL	P@0.02	PPL	P@0.02	PPL	P@0.02	PPL	P@0.02	PPL	P@5	P@10	P@0.02	COH	r-time (second)
LNTM + EmbTF + TR + SAL	493	0.786	698	0.650	390	0.720	535	0.249	707	**0.689**	**0.662**	**0.562**	0.745	852
LNTM + EmbTF + TR	484	0.786	701	0.650	390	0.720	534	0.249	723	0.679	0.656	0.556	**0.750**	29.38
LNTM + TR	483	0.784	700	0.650	389	0.725	534	0.249	736	0.676	0.652	0.554	0.743	29.75
LNTM + EmbTF	485	0.784	700	0.650	390	0.722	534	0.249	**666**	0.675	0.649	0.548	0.726	23.21
NTM without Lifelong Learning	454	0.785	584	0.651	311	0.726	470	0.268	706	0.665	0.634	0.521	0.709	22.23

AGnews *TMN* *R21578* *20NS* *TMNtitle* (Sparse Data)

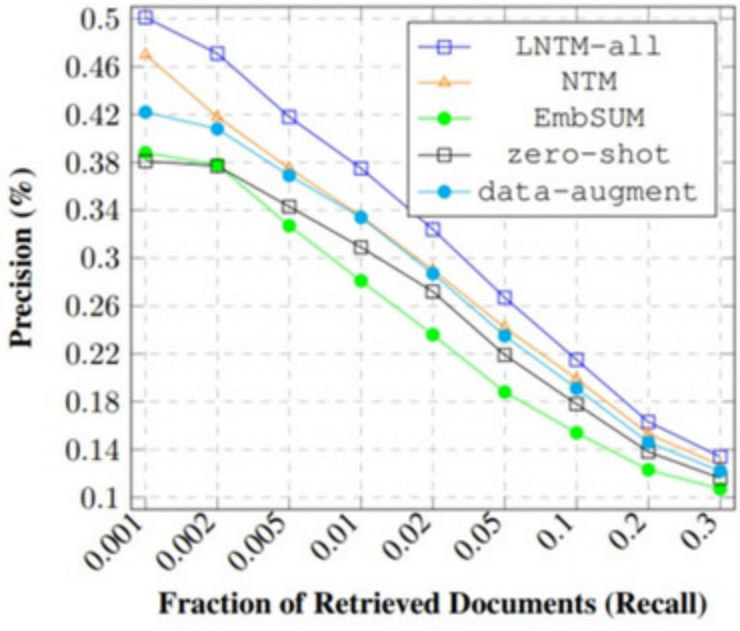
Lifelong Neural Topic Modeling over a stream of document collections

See scores for R21578title data in the paper

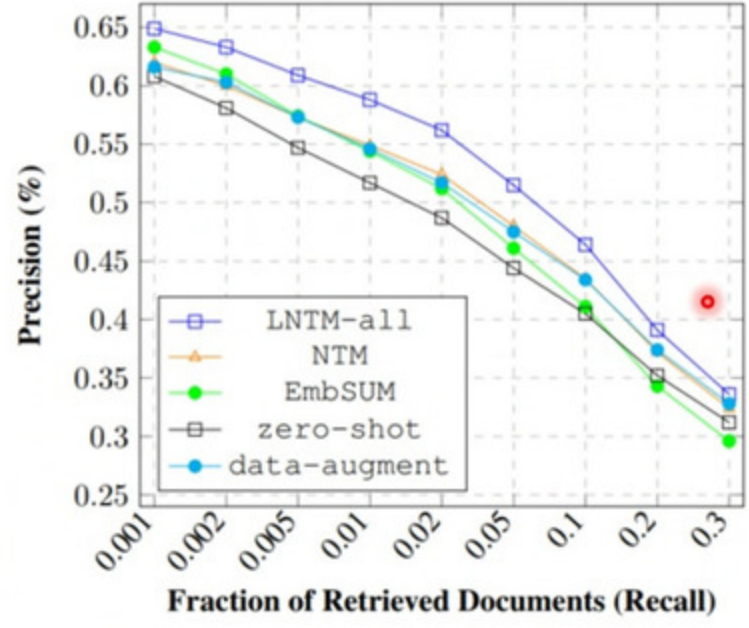
Unrestricted © Siemens AG 2020

IR Analysis: P@R Curve (Max. Knowledge Transfer + Min. Forgetting)

SIEMENS
Ingenuity for life



(a) **IR:** 20NSshort



(b) **IR:** TMNtitle

See IR plot for R21578title data in the paper

Unrestricted © Siemens AG 2020

Topic Analysis: Qualitative (Max. Knowledge Transfer + Min. Forgetting)

SIEMENS
Ingenuity for life

Model	Topic-words (Top 5)
NTM	T1: nuclear, break, jobs, afghanistan, ipad T2: gulf, bruins, japanese, michigan, radiation
LNTM + TR	T1: arts, android, iphone, tablet, ipad T2: rail, medicare, wildfire, radioactive, recession
LNTM-all	T1: linkedin, android, tablet, ipad, iphone T2: tornadoes, fukushima, radioactive, radiation, medicare

Analysis: Qualitative topics of TMNtitle

See ablation studies in the paper

Unrestricted © Siemens AG 2020

Conclusion

SIEMENS
Ingenuity for life

Novel **Lifelong Neural Topic Modeling** framework over Streaming Document Collections

- *Transfer Learning and Domain Adaptation:*
  - **Topic Regularization** with TopicPool (TR)
  - **Transfer Learning** with WordPool (EmbTF)
- *Knowledge Retention/Minimizing Catastrophic Forgetting:*
  - **Topic Regularization** with TopicPool (TR)
  - **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training
- **Joint** Topic Mining, Knowledge Transfer, Domain Adaptation and Knowledge Retention
- State-of the-art results (**PPL/IR/Topic Coherence**) in modeling **sparse** data sets

Code: https://github.com/pgcool/Lifelong-Neural-Topic-Modeling

Webpage: https://sites.google.com/view/gupta-pankaj
LinkedIn: https://www.linkedin.com/in/pankaj-gupta-phd-6b95bb17

Unrestricted © Siemens AG 2020

Conclusion

Thanks!!!

SIEMENS
Ingenuity for life

Novel **Lifelong Neural Topic Modeling** framework over Streaming Document Collections

- *Transfer Learning and Domain Adaptation:*
  - **Topic Regularization** with TopicPool (TR)
  - **Transfer Learning** with WordPool (EmbTF)
- *Knowledge Retention/Minimizing Catastrophic Forgetting:*
  - **Topic Regularization** with TopicPool (TR)
  - **Selective-Data Augmentation Learning** (SAL) with Distillation and Co-training
- **Joint** Topic Mining, Knowledge Transfer, Domain Adaptation and Knowledge Retention
- State-of the-art results (**PPL/IR/Topic Coherence**) in modeling **sparse** data sets

Code: https://github.com/pgcool/Lifelong-Neural-Topic-Modeling

Webpage: https://sites.google.com/view/gupta-pankaj
LinkedIn: https://www.linkedin.com/in/pankaj-gupta-phd-6b95bb17

Wish You

A Happy

LIFELONG LEARNING!!!

Unrestricted © Siemens AG 2020