Explainable and Discourse Topic-aware Neural Language Understanding

Presenter: **Yatin Chaudhary** @ ICML 2020, July 12-18, 2020

Author(s): Yatin Chaudhary[1,2], Hinrich Schütze[2] and Pankaj Gupta[1]

[1]Corporate Technology, Machine-Intelligence (MIC-DE), Siemens AG Munich, Germany

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

Table of contents

1 Motivations and Contributions

2 Architecture: Neural Language Model

3 Architecture: Neural Topic Model

4 Architecture: Neural Composite Language Model

5 Experiments and Analysis

6 Conclusion

Background: Complementary Learning

Neural Topic Model (NTM):

Neural Language Model (NLM):

Background: Complementary Learning

Neural Topic Model (NTM):

✓ **Statistical modeling** of documents (via word co-occurrences)

Neural Language Model (NLM):

Document corpus

Document 1

sent#1: An integrated **circuit** (IC) is a set of **electronic circuits** used for **computer processor**.

sent#2: Production of **chip** is a multi-step process.

sent#3: **Sales** are expected to grow 2.7 % to 79.3 **billion dollars**, the largest **market share** in future.

⋮

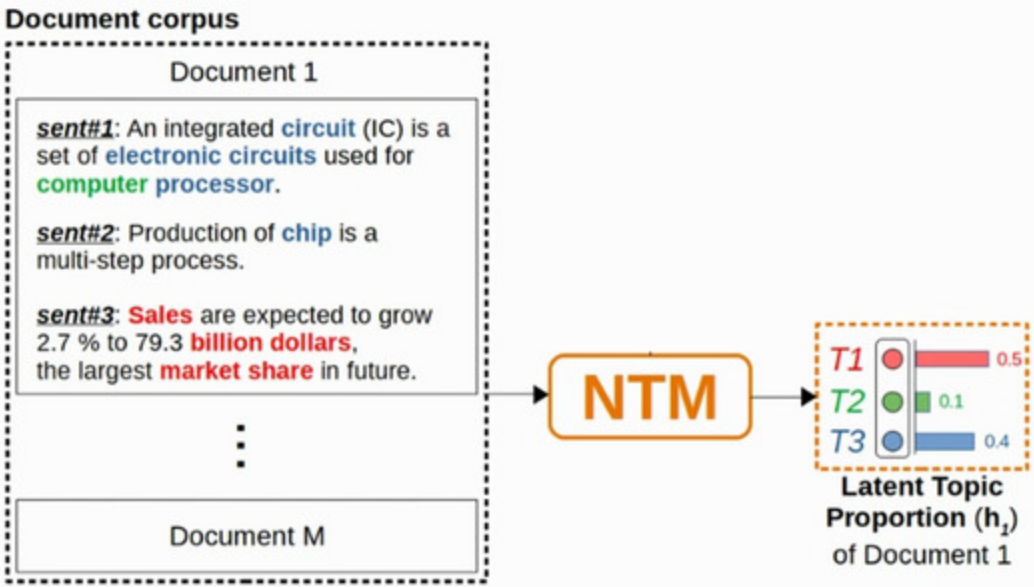
Document M

Background: Complementary Learning

Neural Topic Model (NTM):

- ✓ **Statistical modeling** of documents (via word co-occurrences)
- ✓ **Capture** document semantics via **latent topic proportions**

Neural Language Model (NLM):

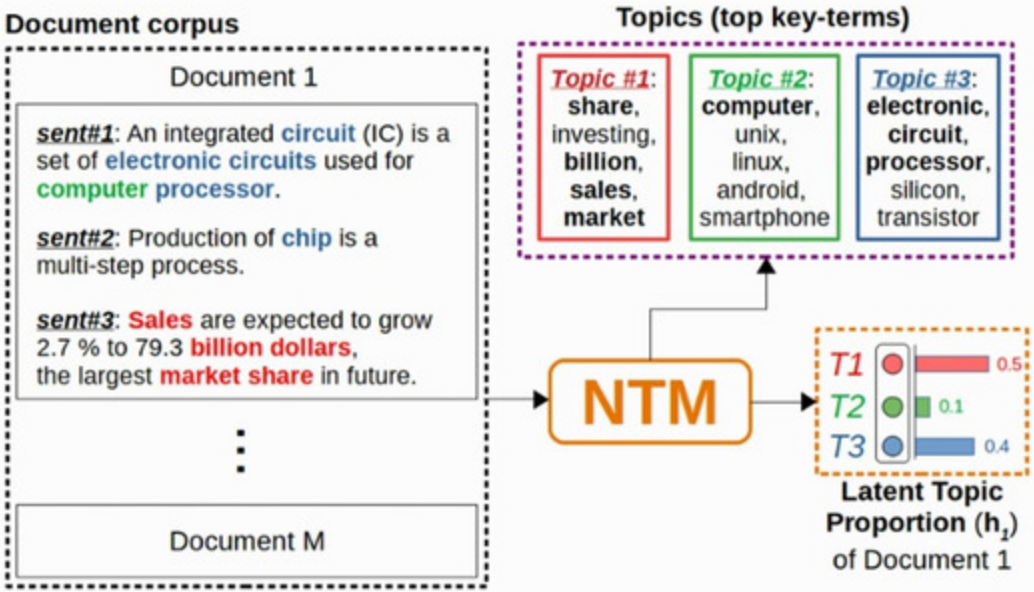


Background: Complementary Learning

Neural Topic Model (NTM):

- ✓ **Statistical modeling** of documents (via word co-occurrences)
- ✓ **Capture** document semantics via **latent topic proportions**
- ✓ **Topics** capture thematic structures defined by key-terms

Neural Language Model (NLM):



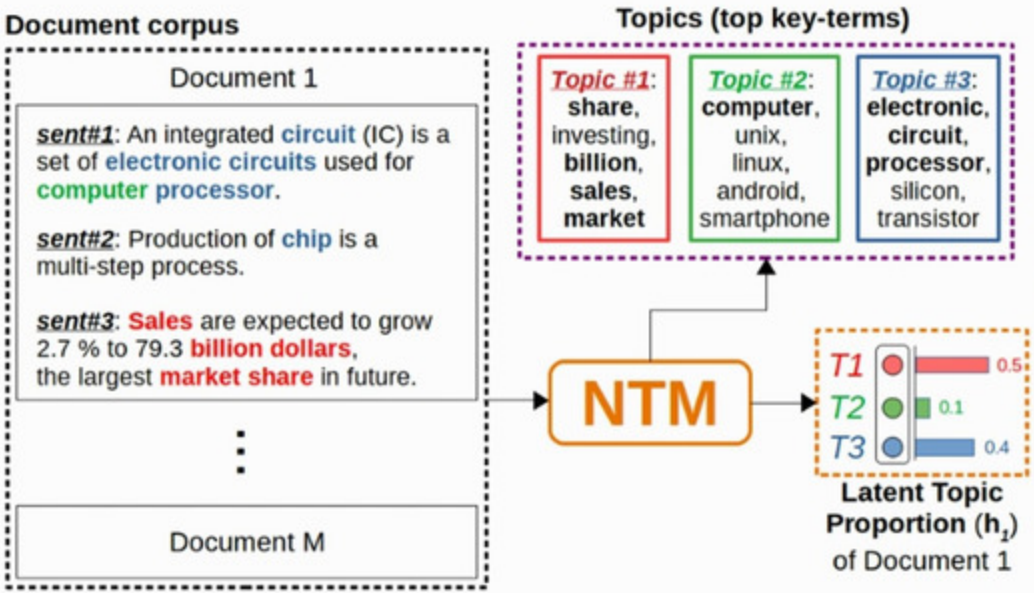
Background: Complementary Learning

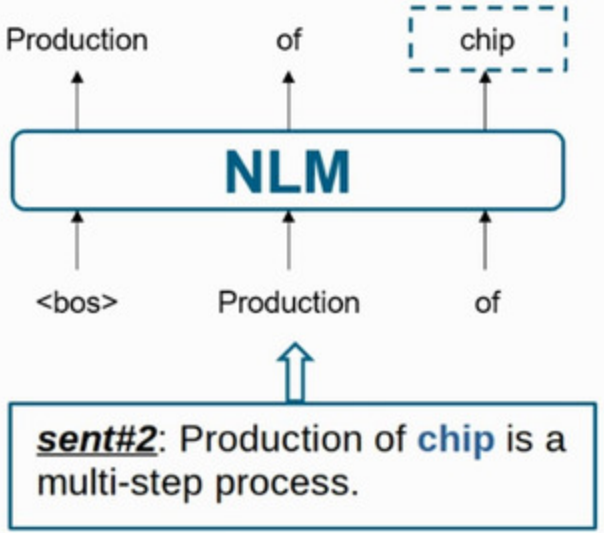
Neural Topic Model (NTM):

- ✓ **Statistical modeling** of documents (via word co-occurrences)
- ✓ **Capture** document semantics via **latent topic proportions**
- ✓ **Topics** capture thematic structures defined by key-terms

Neural Language Model (NLM):

- ✓ **Sequence modeling** via next word prediction
- ✓ **Captures** syntactic structure
- ➢ **Difficult to capture** document-level semantic dependencies





Motivation #1: Improve NLM via Topical semantics

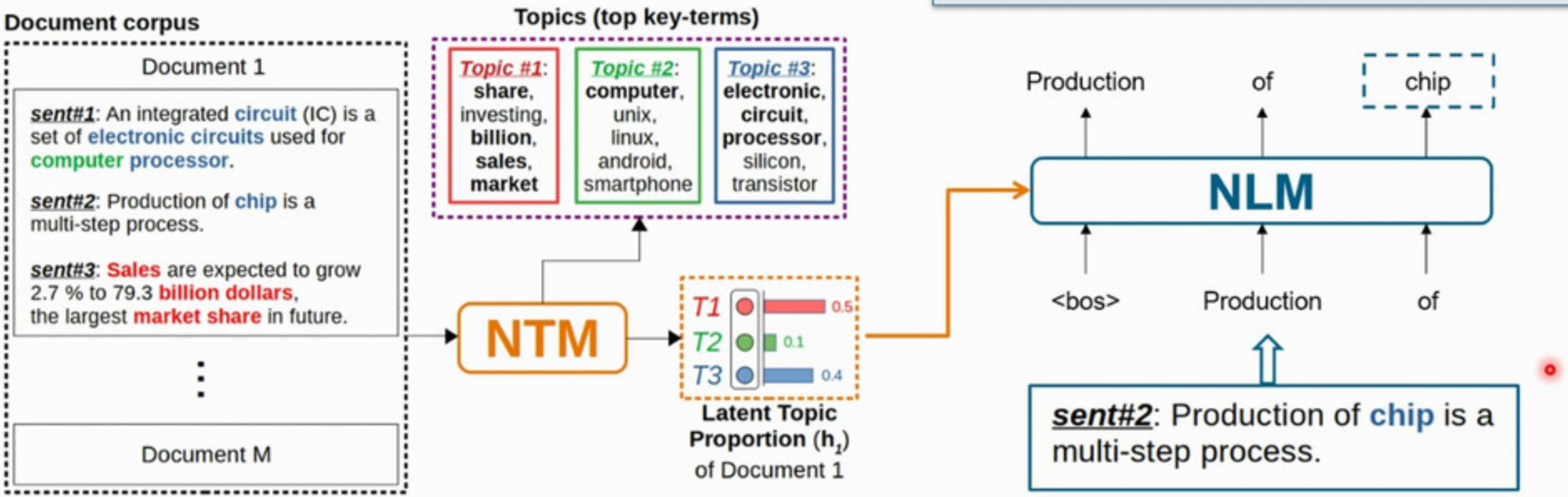
SIEMENS
Ingenuity for life

Neural Topic Model (NTM):

- ✓ **Statistical modeling** of documents (via word co-occurrences)
- ✓ **Capture** document semantics via **latent topic proportions**
- ✓ **Topics** capture thematic structures defined by key-terms

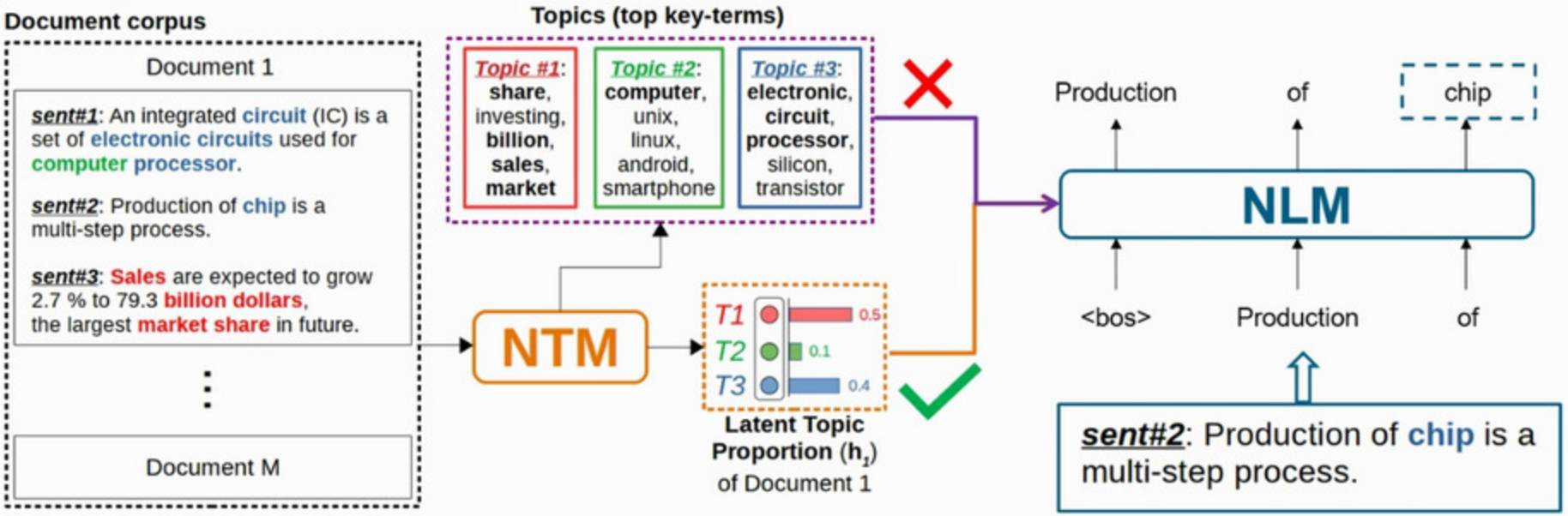
Neural Language Model (NLM):

- ✓ **Sequence modeling** via next word prediction
- ✓ **Captures** syntactic structure
- ➢ **Difficult to capture** document-level semantic dependencies



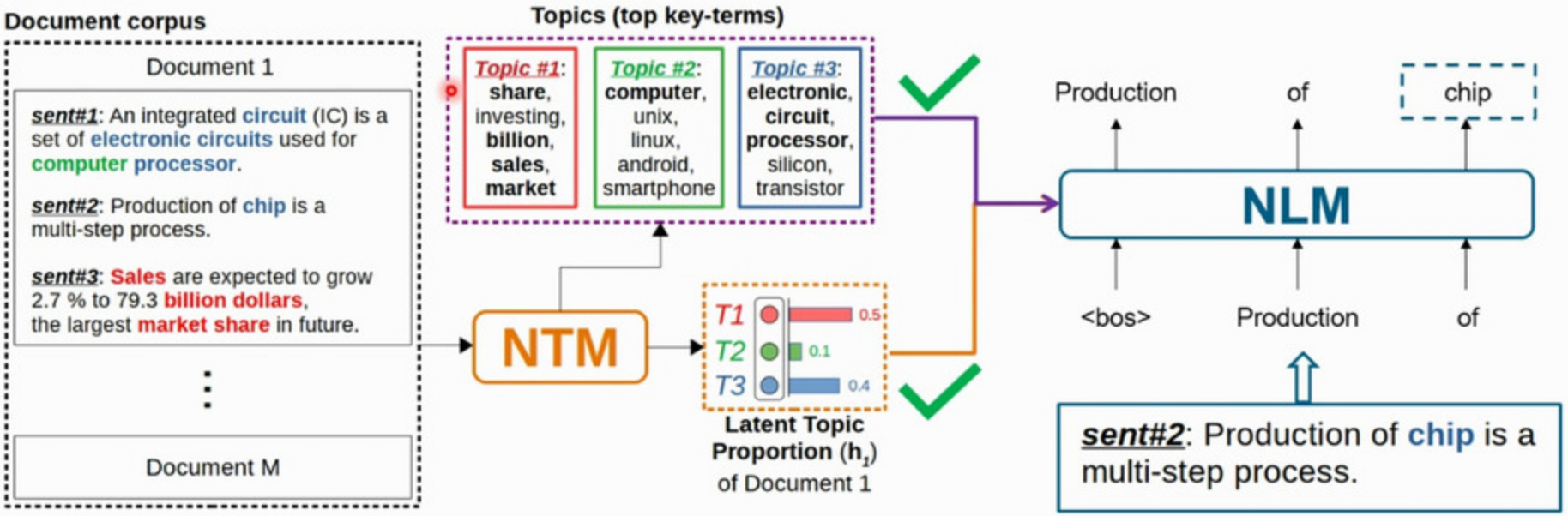
Related work: Improve NLM via Topical semantics

Existing approaches only utilize **latent topic proportions**
(Ignore explainable topics key-terms)



This work: Incorporate Latent and Explainable Topics

Introduce **document-level** **latent** and **explainable** topic representations during language modeling for a **fine-grained explanatory** outlook.



Motivation #2: Sentence-level Topical Discourse

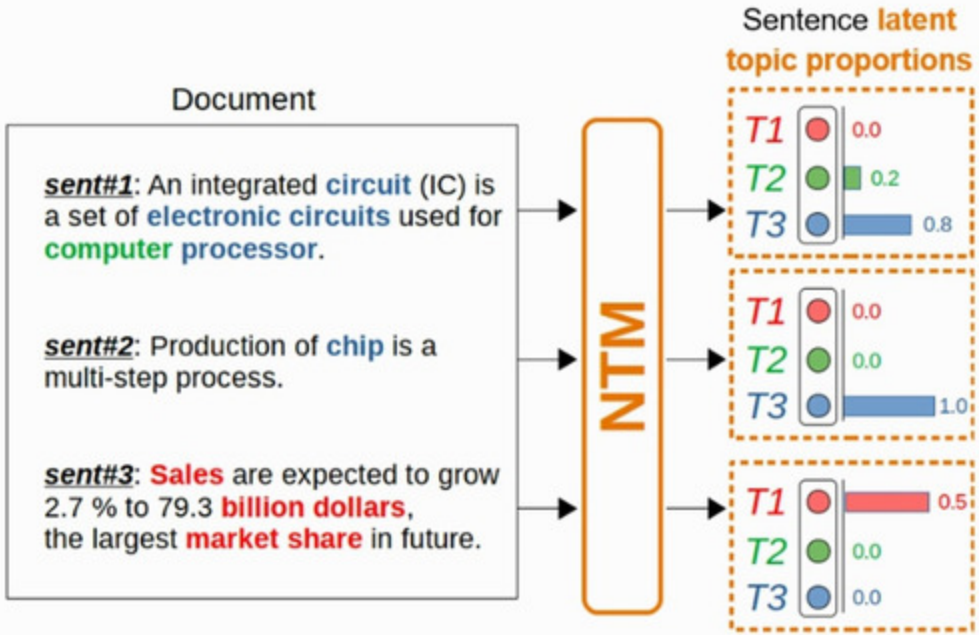
Document

sent#1: An integrated **circuit** (IC) is a set of **electronic circuits** used for **computer processor**.

sent#2: Production of **chip** is a multi-step process.

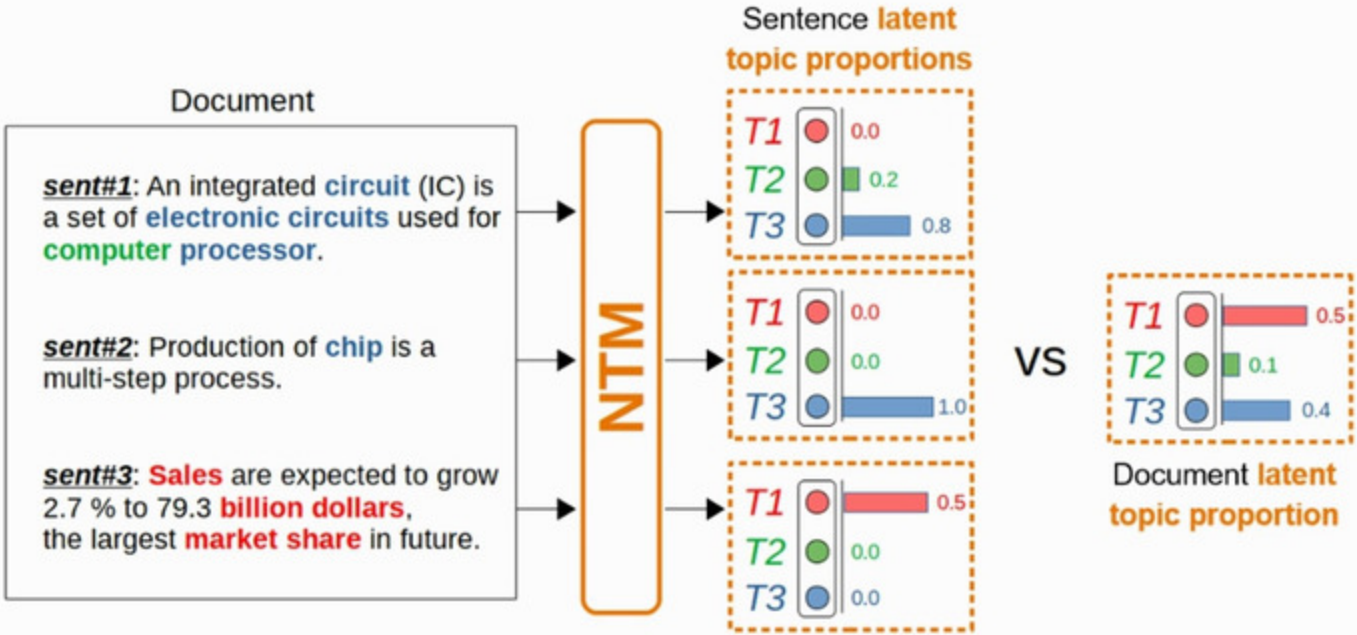
sent#3: **Sales** are expected to grow 2.7 % to 79.3 **billion dollars**, the largest **market share** in future.

Motivation #2: Sentence-level Topical Discourse

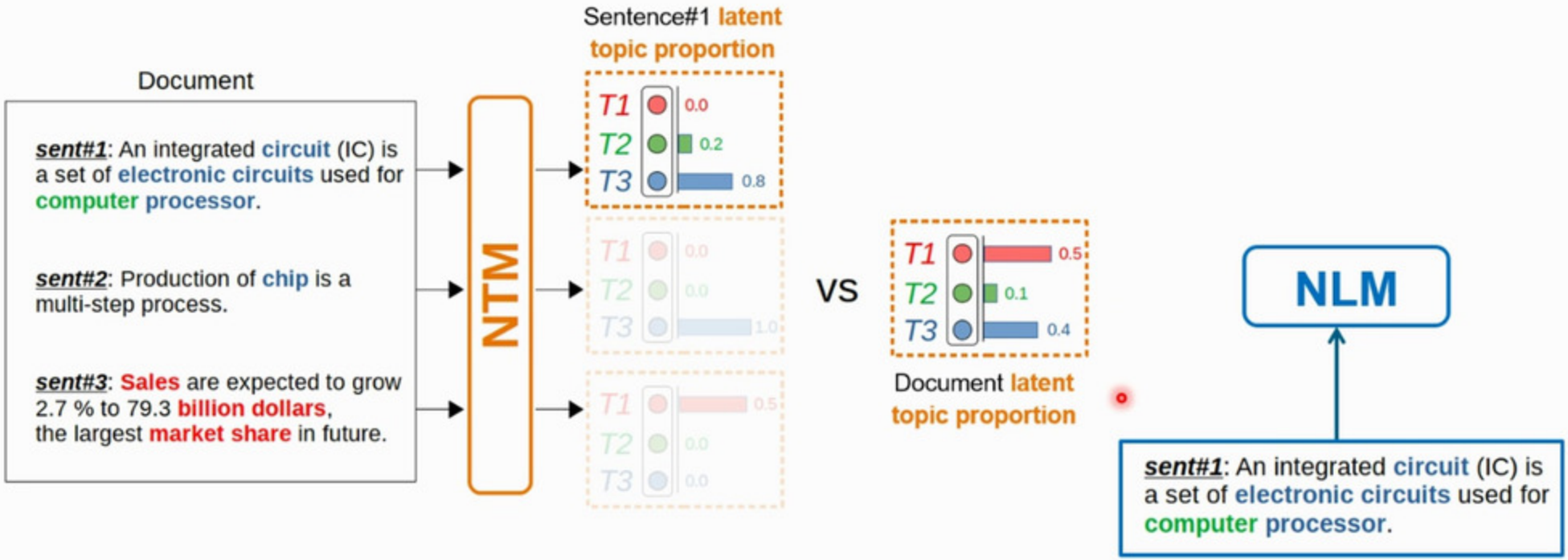


Motivation #2: Sentence-level Topical Discourse

A sentence can have **different topic** than the **document** itself

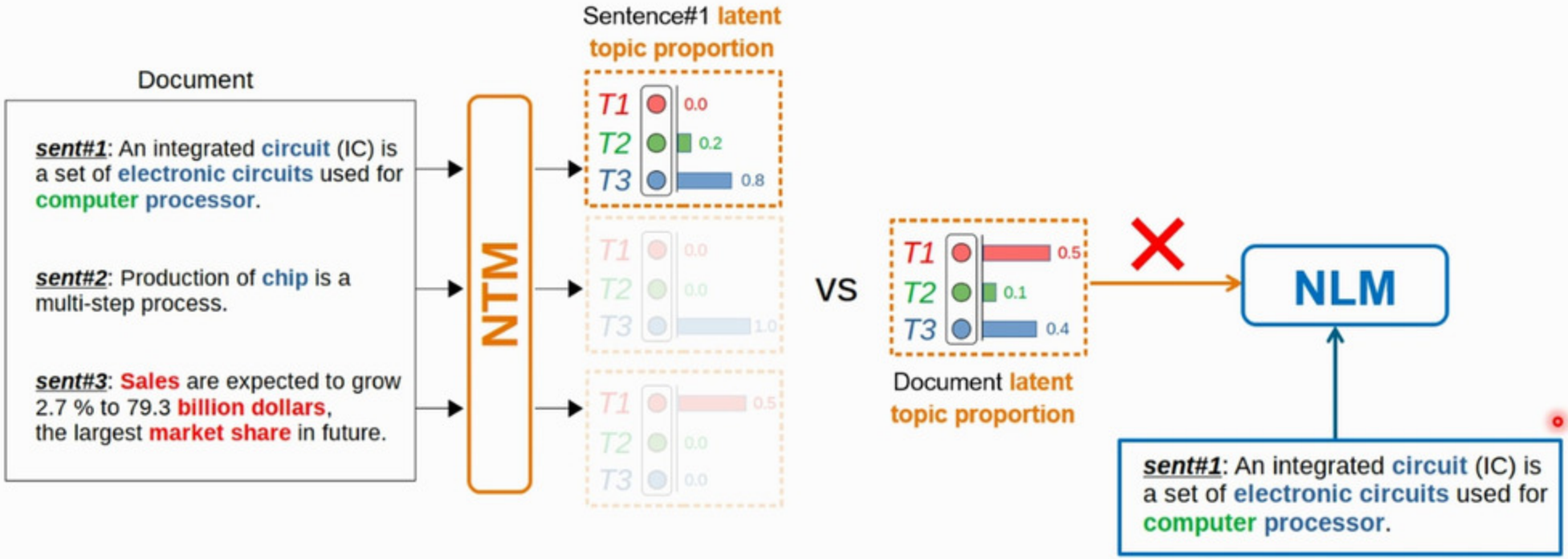


Motivation #2: Sentence-level Topical Discourse



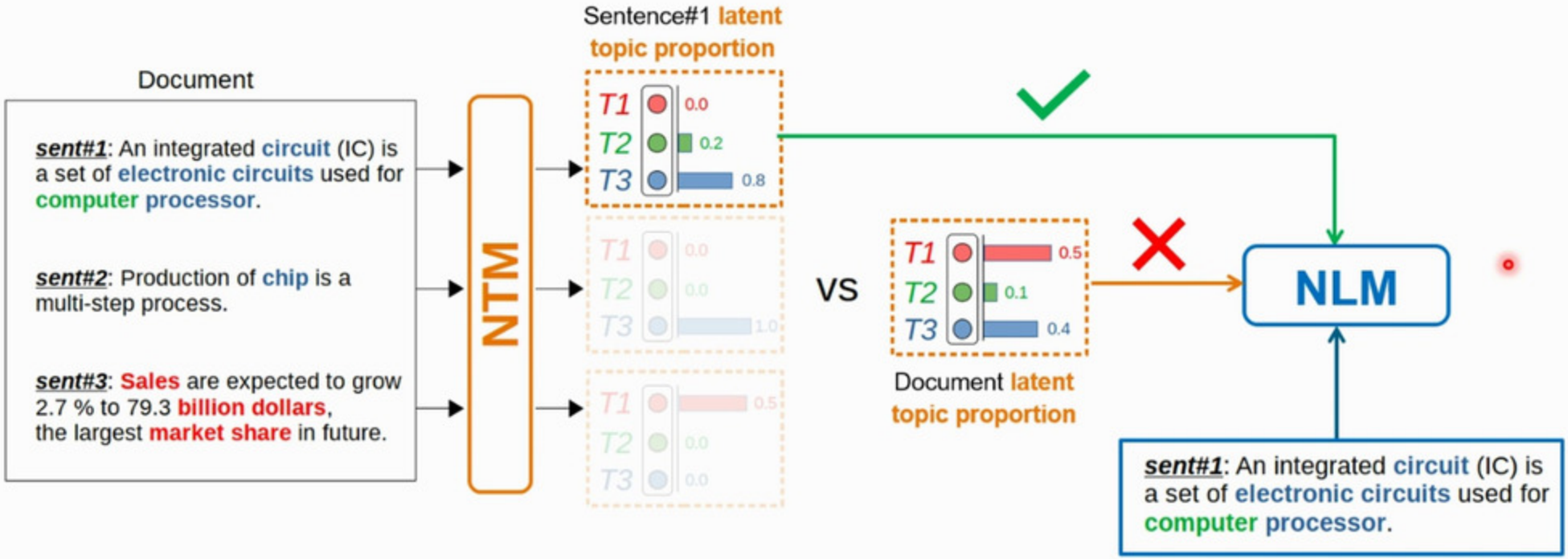
Motivation #2: Sentence-level Topical Discourse

Mismatch in **Sentence-level** and **Document-level** topical semantics

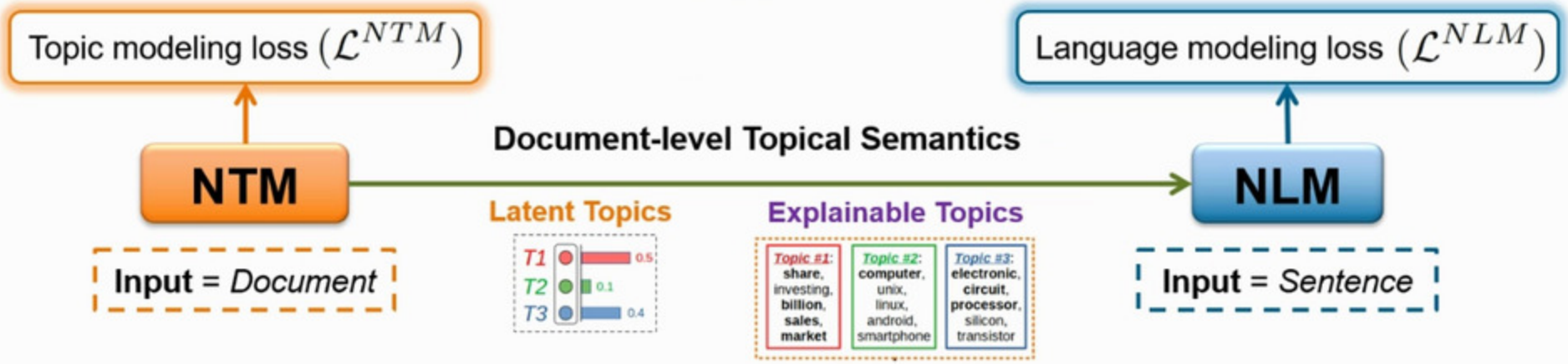


This work: Incorporate Sentence-level Topics into NLM

Introduce Sentence-level **latent** and **explainable topic representation** into NLM to account for **sentence-level topical discourse (SDT)**



Proposed: Neural Composite Language Model (NCLM)



Proposed: Neural Composite Language Model (NCLM)

Hybrid Objective Loss

$$\mathcal{L} = \alpha \cdot \mathcal{L}^{NTM} + (1-\alpha) \cdot \mathcal{L}^{NLM}$$

$(0 \leq \alpha \leq 1)$

$(0 \leq 1-\alpha \leq 1)$

Composite modeling loss $(\mathcal{L})$

α

$1-\alpha$

Topic modeling loss $(\mathcal{L}^{NTM})$

Language modeling loss $(\mathcal{L}^{NLM})$

Document-level Topical Semantics

NTM

NLM

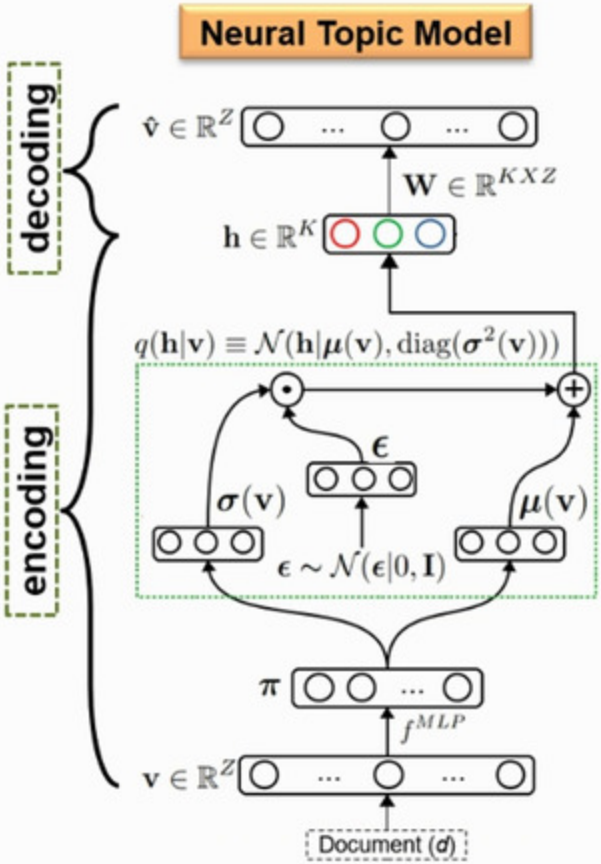
Latent Topics

Explainable Topics

Input = *Document*

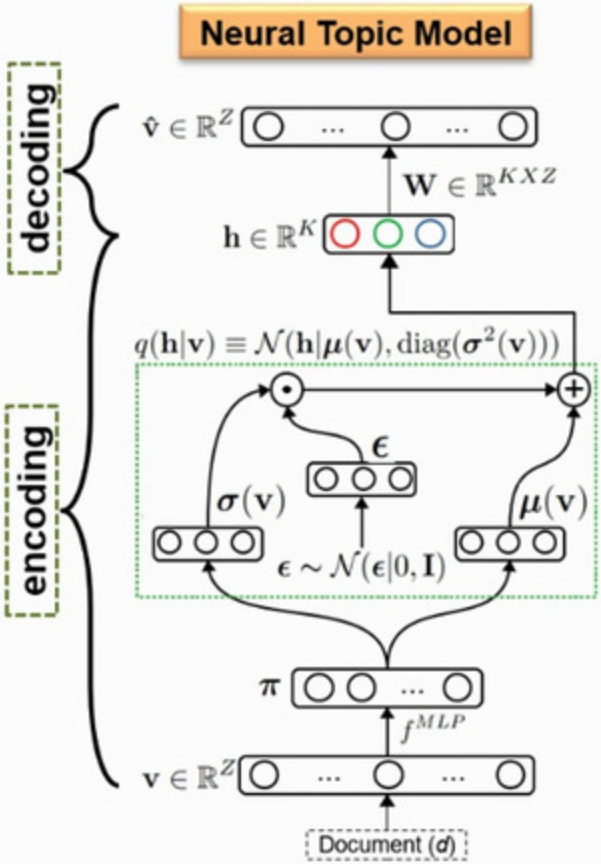
Input = *Sentence*

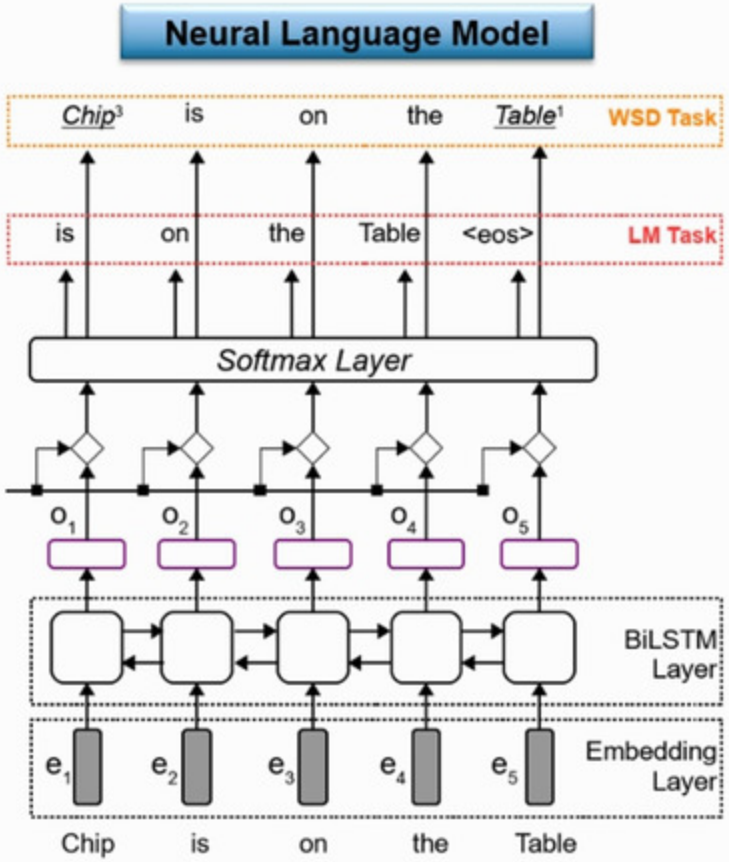
Architecture: Neural Topic Model (NTM)



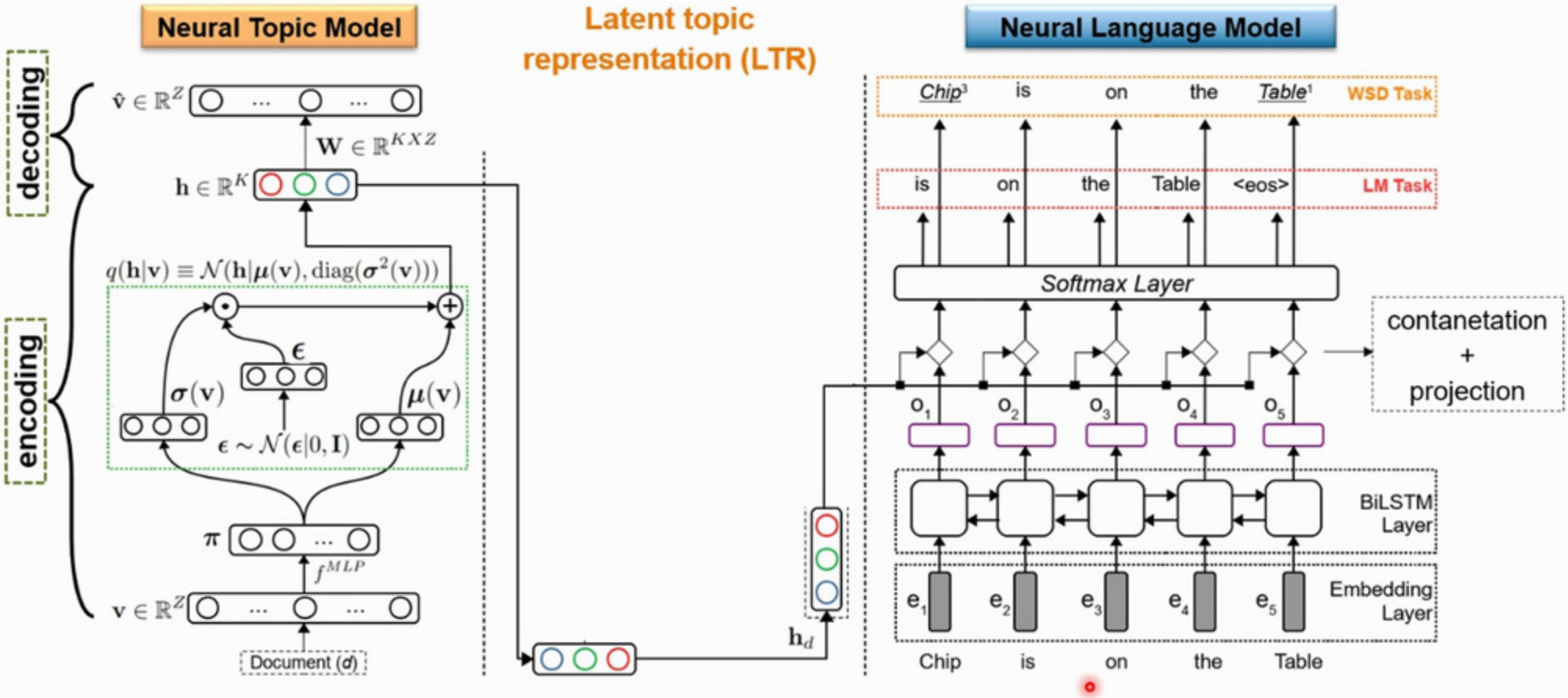
Reference: Miao, Y., Yu, L. and Blunsom, P., Neural variational inference for text processing. In *ICML2016* (pp. 1727-1736).

Architecture: Neural Language Model (NLM)

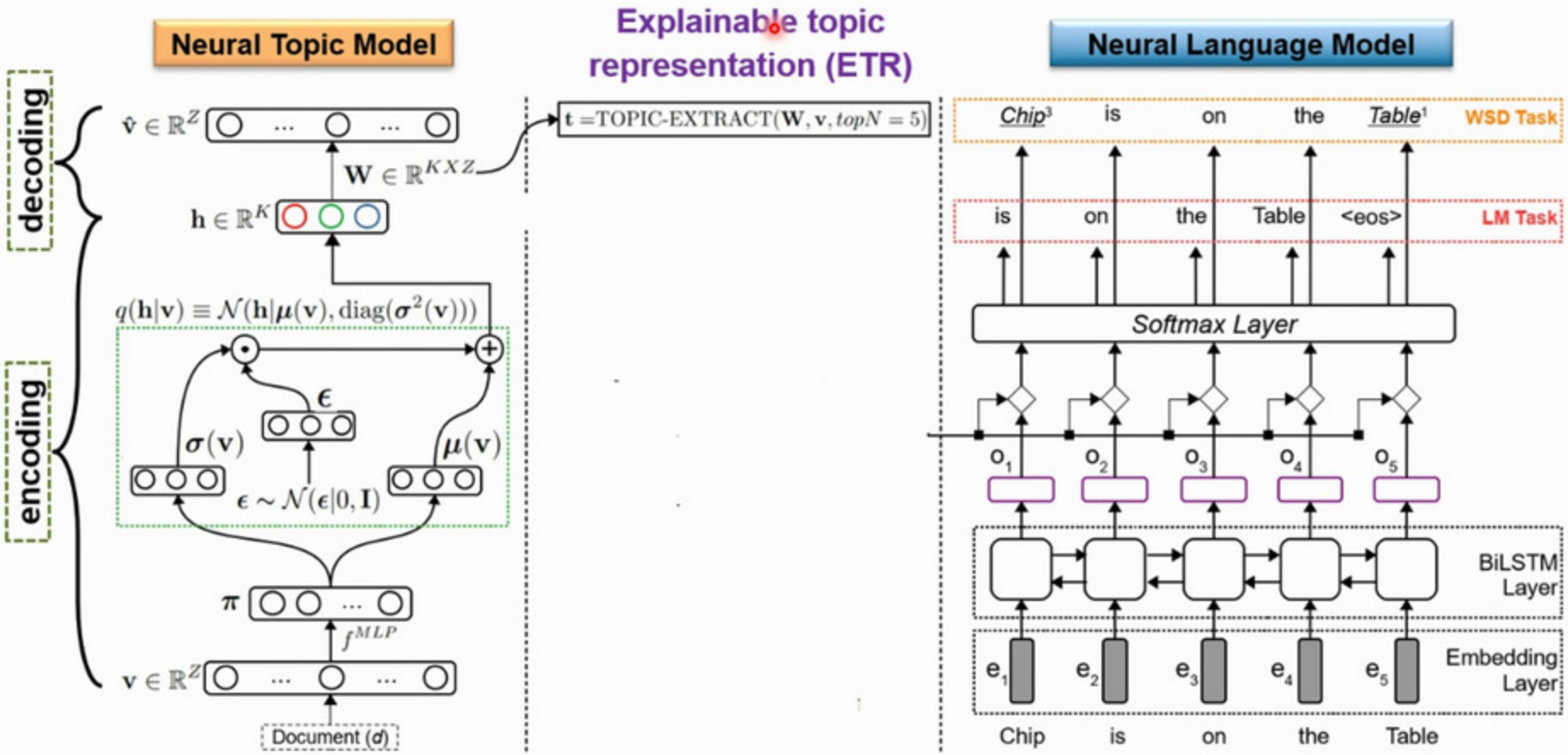




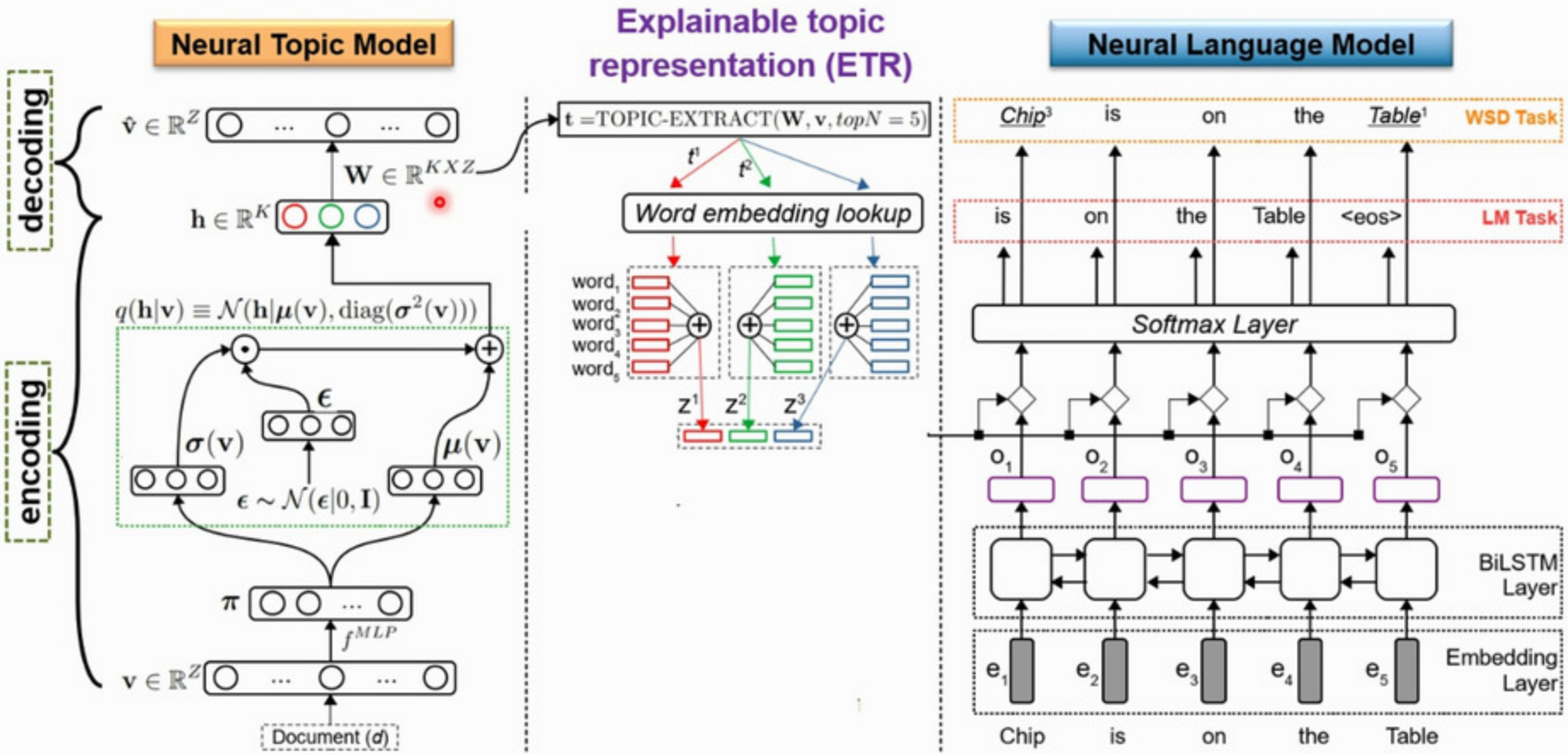
LTA-NLM: Latent Topic-aware NLM



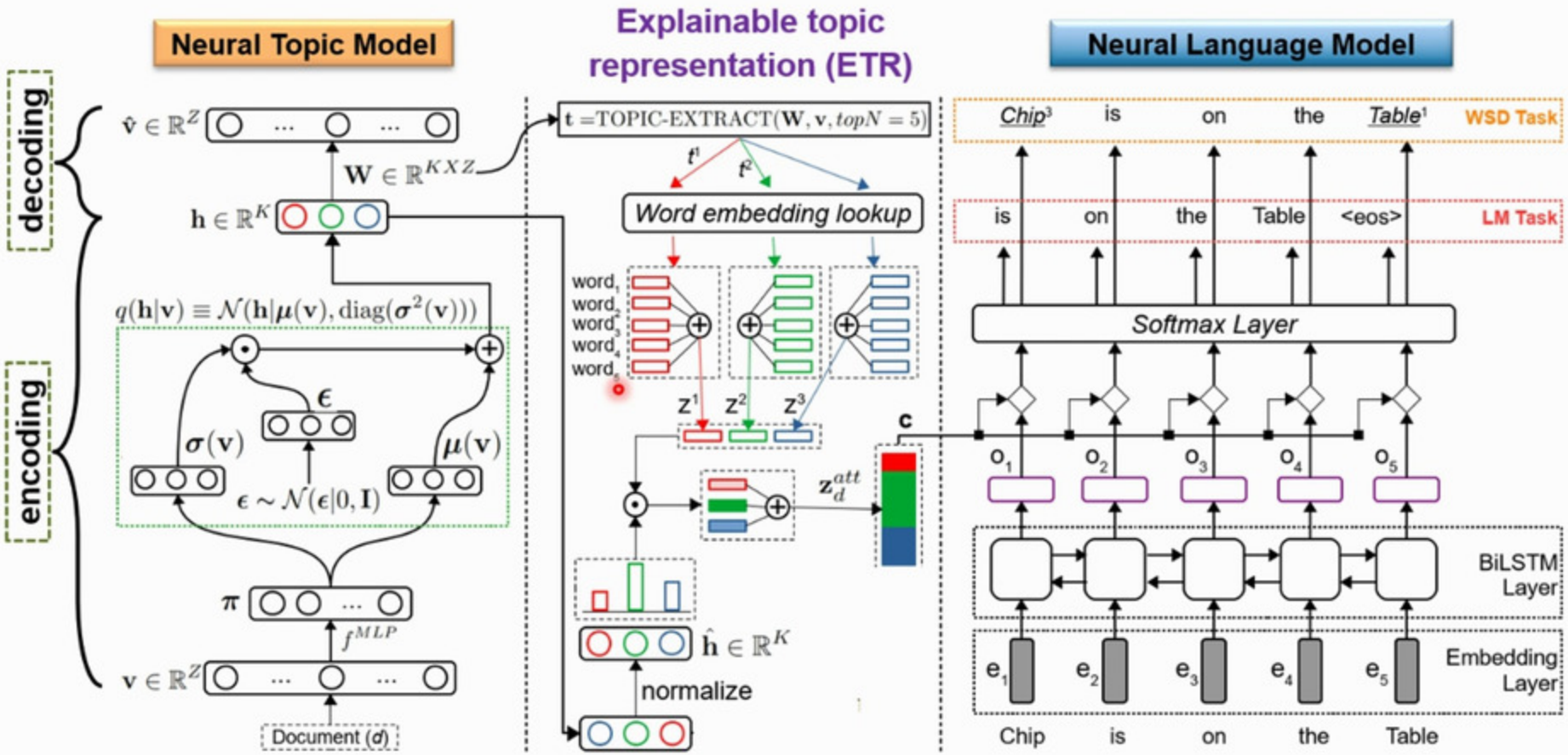
ETA-NLM: Explainable Topic-aware NLM



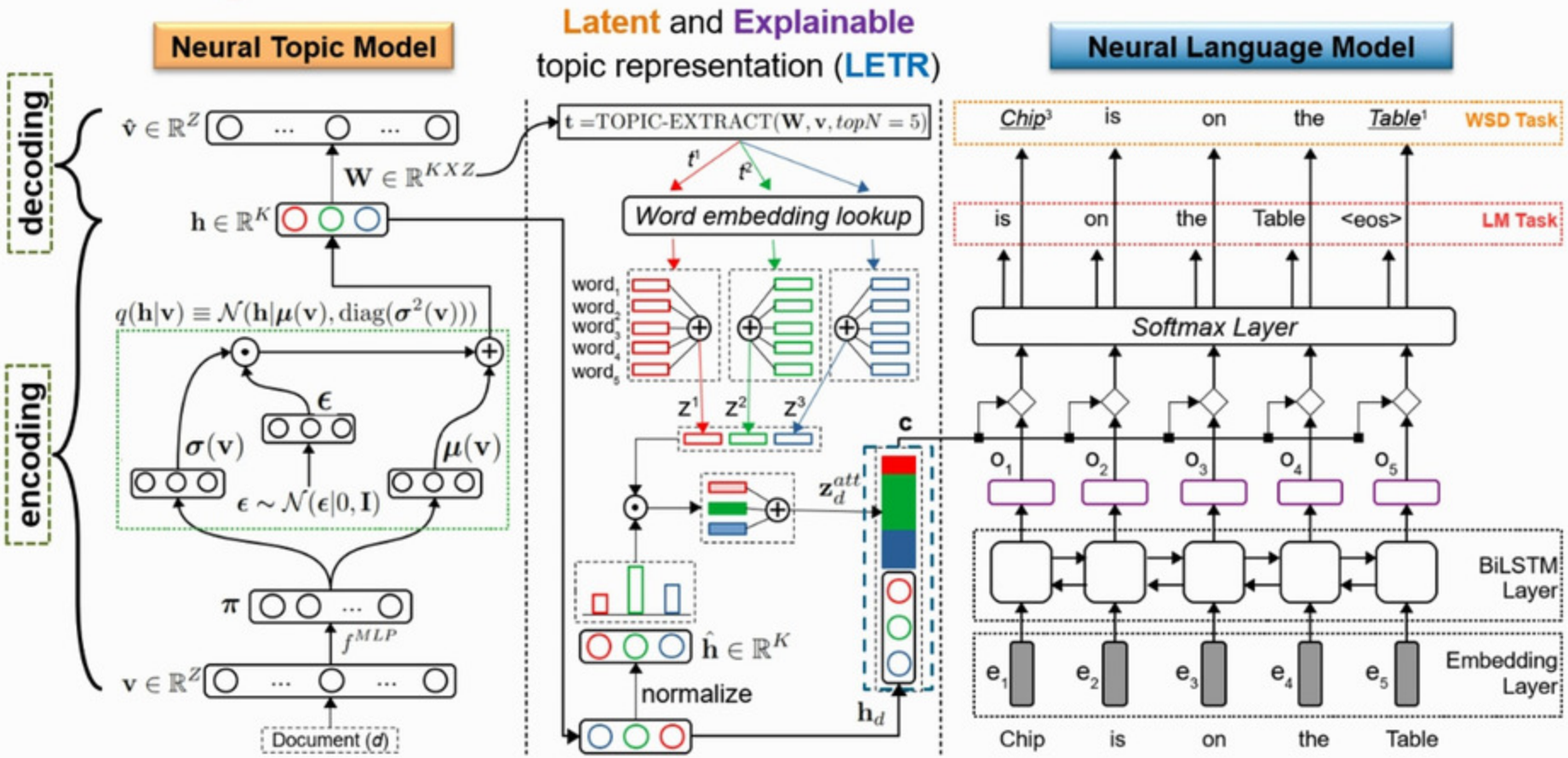
ETA-NLM: Explainable Topic-aware NLM



ETA-NLM: Explainable Topic-aware NLM



LETA-NLM: Latent and Explainable Topic-aware NLM



Evaluation: Natural Language Tasks

1) Language Modeling
2) Topic Modeling
3) Text Classification
4) Information Retrieval
5) Word Sense Disambiguation

Datasets: APNEWS, BNC, 20NewsGroups, R21578, AGnews, IMDB

Evaluation: Language Modeling (Perplexity)

	MODEL	APNEWS		IMDB		BNC	
		S	L	S	L	S	L
BASELINES	*LDA+LSTM**	54.83	50.17	69.62	62.78	96.38	87.28
	*LCLM**	54.18	50.63	67.78	67.86	87.47	80.68
	*TopicRNN**	54.12	50.01	66.45	60.14	93.55	84.12
	*TDLM**	52.65	48.21	63.82	58.59	86.43	80.58
	*TCNLM**	52.59	47.74	62.59	56.12	86.21	80.12
	LSTM-LM	64.95	59.28	72.31	65.54	106.82	98.78
	LTA-NLM	55.48	49.61	68.21	61.49	98.31	89.36
PROPOSED	+ *SDT*	48.23	42.85	63.81	58.90	90.36	**80.30**
	ETA-NLM	49.34	43.19	59.20	51.40	95.62	87.22
	+ *SDT*	48.75	43.50	57.83	50.51	95.64	88.37
	LETA-NLM	48.33	43.17	58.10	52.35	94.78	86.73
	+ *SDT*	**42.98**	**39.41**	**56.65**	**51.05**	**88.30**	81.12
	GAIN(%)	33.8	33.5	21.6	22.1	17.3	18.7

Bold values → best performing proposed model compared to *LSTM-LM*

GAIN (%) → gain in performance **(bold** vs *LSTM-LM***)**

* → values taken from ``Wang, W., Gan, Z., Wang, W., Shen, D., Huang, J., Ping, W., Satheesh, S. and Carin, L., 2018, March. Topic Compositional Neural Language Model. In International Conference on Artificial Intelligence and Statistics (pp. 356-365)''

NLM → *LSTM-LM*

S (small) → 1 layer NLM with 600 hidden units

L (large) → 2 layer NLM with 900 hidden units

(+) → augment this feature to the model in previous row

Evaluation: Language Modeling (Perplexity)

	MODEL	APNEWS		IMDB		BNC	
		S	L	S	L	S	L
BASELINES	*LDA+LSTM**	54.83	50.17	69.62	62.78	96.38	87.28
	*LCLM**	54.18	50.63	67.78	67.86	87.47	80.68
	*TopicRNN**	54.12	50.01	66.45	60.14	93.55	84.12
	*TDLM**	52.65	48.21	63.82	58.59	86.43	80.58
	*TCNLM**	52.59	47.74	62.59	56.12	86.21	80.12
	LSTM-LM	64.95	59.28	72.31	65.54	106.82	98.78
	LTA-NLM	55.48	49.61	68.21	61.49	98.31	89.36
PROPOSED	+ *SDT*	48.23	42.85	63.81	58.90	90.36	**80.30**
	ETA-NLM	49.34	43.19	59.20	51.40	95.62	87.22
	+ *SDT*	48.75	43.50	57.83	50.51	95.64	88.37
	LETA-NLM	48.33	43.17	58.10	52.35	94.78	86.73
	+ *SDT*	**42.98**	**39.41**	**56.65**	**51.05**	**88.30**	81.12
	GAIN(%)	33.8	33.5	21.6	22.1	17.3	18.7

- *ETA-NLM* **outperforms** *LTA-NLM* due to **explanatory document semantics** provided by **ETR.**

11% gain

Bold values → best performing proposed model compared to *LSTM-LM*

GAIN (%) → gain in performance **(bold** vs *LSTM-LM***)**

* → values taken from ``Wang, W., Gan, Z., Wang, W., Shen, D., Huang, J., Ping, W., Satheesh, S. and Carin, L., 2018, March. Topic Compositional Neural Language Model. In International Conference on Artificial Intelligence and Statistics (pp. 356-365)''

NLM → *LSTM-LM*

S (small) → 1 layer NLM with 600 hidden units

L (large) → 2 layer NLM with 900 hidden units

(+) → augment this feature to the model in previous row

Evaluation: Language Modeling (Perplexity)

	MODEL	APNEWS		IMDB		BNC	
		S	L	S	L	S	L
BASELINES	*LDA+LSTM**	54.83	50.17	69.62	62.78	96.38	87.28
	*LCLM**	54.18	50.63	67.78	67.86	87.47	80.68
	*TopicRNN**	54.12	50.01	66.45	60.14	93.55	84.12
	*TDLM**	52.65	48.21	63.82	58.59	86.43	80.58
	*TCNLM**	52.59	47.74	62.59	56.12	86.21	80.12
	LSTM-LM	64.95	59.28	72.31	65.54	106.82	98.78
	LTA-NLM	55.48	49.61	68.21	61.49	98.31	89.36
PROPOSED	+ *SDT*	48.23	42.85	63.81	58.90	90.36	**80.30**
	ETA-NLM	49.34	43.19	59.20	51.40	95.62	87.22
	+ *SDT*	48.75	43.50	57.83	50.51	95.64	88.37
	LETA-NLM	48.33	43.17	58.10	52.35	94.78	86.73
	+ *SDT*	**42.98**	**39.41**	**56.65**	**51.05**	**88.30**	81.12
	GAIN(%)	33.8	33.5	21.6	22.1	17.3	18.7

- *ETA-NLM* **outperforms** *LTA-NLM* due to **explanatory document semantics** provided by **ETR.**

- *LETA-NLM* **outperforms** both *LTA-NLM* and *ETA-NLM* by leveraging both **LTR** and **ETR.**

13% gain

Bold values → best performing proposed model compared to *LSTM-LM*

GAIN (%) → gain in performance **(bold** vs *LSTM-LM***)**

* → values taken from ``Wang, W., Gan, Z., Wang, W., Shen, D., Huang, J., Ping, W., Satheesh, S. and Carin, L., 2018, March. Topic Compositional Neural Language Model. In International Conference on Artificial Intelligence and Statistics (pp. 356-365)''

NLM → *LSTM-LM*

S (small) → 1 layer NLM with 600 hidden units

L (large) → 2 layer NLM with 900 hidden units

(+) → augment this feature to the model in previous row

Evaluation: Language Modeling (Perplexity)

	MODEL	APNEWS		IMDB		BNC	
		S	L	S	L	S	L
BASELINES	*LDA+LSTM**	54.83	50.17	69.62	62.78	96.38	87.28
	*LCLM**	54.18	50.63	67.78	67.86	87.47	80.68
	*TopicRNN**	54.12	50.01	66.45	60.14	93.55	84.12
	*TDLM**	52.65	48.21	63.82	58.59	86.43	80.58
	*TCNLM**	52.59	47.74	62.59	56.12	86.21	80.12
	LSTM-LM	64.95	59.28	72.31	65.54	106.82	98.78
	LTA-NLM	55.48	49.61	68.21	61.49	98.31	89.36
PROPOSED	+ *SDT*	48.23	42.85	63.81	58.90	90.36	**80.30**
	ETA-NLM	49.34	43.19	59.20	51.40	95.62	87.22
	+ *SDT*	48.75	43.50	57.83	50.51	95.64	88.37
	LETA-NLM	48.33	43.17	58.10	52.35	94.78	86.73
	+ *SDT*	**42.98**	**39.41**	**56.65**	**51.05**	**88.30**	81.12
	GAIN(%)	33.8	33.5	21.6	22.1	17.3	18.7

- *ETA-NLM* **outperforms** *LTA-NLM* due to **explanatory document semantics** provided by **ETR.**
- *LETA-NLM* **outperforms** both *LTA-NLM* and *ETA-NLM* by leveraging both **LTR** and **ETR.**
- *+SDT* further **improves** the performance by leveraging **sentence-level topic representation**.

Bold values → best performing proposed model compared to *LSTM-LM*

GAIN (%) → gain in performance (**bold** vs *LSTM-LM*)

* → values taken from ``Wang, W., Gan, Z., Wang, W., Shen, D., Huang, J., Ping, W., Satheesh, S. and Carin, L., 2018, March. Topic Compositional Neural Language Model. In International Conference on Artificial Intelligence and Statistics (pp. 356-365)''

NLM → *LSTM-LM*

S → 1 layer NLM with 600 hidden units

L → 2 layer NLM with 900 hidden units

(+) → augment this feature to the model in previous row

Evaluation: Word Sense Disambiguation (F1 scores)

Model	Dev	Test				
	SE07	SE13	SE15	SE2	SE3	ALL
MFS	54.5	63.8	67.1	65.6	66.0	65.5
BiLSTM-LM	55.0	53.9	60.8	63.6	60.8	59.8
LTA-NLM	**56.0**	**54.8**	60.9	**64.8**	62.2	**60.7**
ETA-NLM	55.4	54.7	60.6	**64.8**	**62.3**	**60.7**
LETA-NLM	55.6	54.7	**61.1**	64.7	62.1	**60.7**

NLM → *BiLSTM-LM*

MFS → baseline model based on **Most Frequent Sense** extracted from **Wordnet***

BiLSTM-LM → **bidirectional *LSTM-LM*** (baseline model)

Bold values → **best performing** proposed model compared to *BiLSTM-LM*

* → George A. Miller (1995). WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.

- Proposed models **outperforms** the *BiLSTM-LM* model
- Leveraging **latent**, **explainable** and **joint** topic representations **enhance the sense disambiguation** task

Evaluation: Text Classification (F1 scores)

	MODEL	20NS	R21578	IMDB
BASELINES	*CNN-Rand*	.721	.690	.888
	+Topic	.724	.699	.891
	CNN-LSTM	.745	.750	.899
PROPOSED	*CNN-LTA*	.753	.759	.907
	CNN-ETA	**.775**	**.763**	.903
	CNN-LETA	.770	.750	**.908**

- Leveraging **latent**, **explainable** or **joint** topic representations **outperforms** the *CNN-LSTM* model.

CNN-Rand → **CNN*** model with **randomly initialized** input embeddings
+Topic → augment **latent topic representation** in the input of *CNN-Rand*
CNN-LSTM → **CNN*** model with output of **pretrained** *LSTM-LM* as input

* → Kim, Y., 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

CNN-LTA → **CNN*** model with output of **pretrained** *LTA-NLM* as input
CNN-ETA → **CNN*** model with output of **pretrained** *ETA-NLM* as input
CNN-LETA → **CNN*** model with output of **pretrained** *LETA-NLM* as input
Bold values → best performing proposed model compared to *CNN-LSTM*

Evaluation: Text Classification (F1 scores)

	MODEL	20NS	R21578	IMDB
BASELINES	*CNN-Rand*	.721	.690	.888
	+Topic	.724	.699	.891
	CNN-LSTM	.745	.750	.899
PROPOSED	*CNN-LTA*	.753	.759	.907
	CNN-ETA	**.775**	**.763**	.903
	CNN-LETA	.770	.750	**.908**

- Leveraging **latent**, **explainable** or **joint** topic representations **outperforms** the *CNN-LSTM* model.
- *CNN-ETA* **outperforms** *CNN-LTA* which shows that the **ETR** provide **more descriptive document-level semantics** than **LTR**.

CNN-Rand → **CNN*** model with **randomly initialized** input embeddings
+Topic → augment **latent topic representation** in the input of *CNN-Rand*
CNN-LSTM → **CNN*** model with output of **pretrained** *LSTM-LM* as input

* → Kim, Y., 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

CNN-LTA → **CNN*** model with output of **pretrained** *LTA-NLM* as input
CNN-ETA → **CNN*** model with output of **pretrained** *ETA-NLM* as input
CNN-LETA → **CNN*** model with output of **pretrained** *LETA-NLM* as input
Bold values → best performing proposed model compared to *CNN-LSTM*

Evaluation: Information retrieval (Precision@recall scores)

Model	20NS		R21578		AGNews	
	P@5	P@10	P@5	P@10	P@5	P@10
NTM	.198	.190	.581	.567	.607	.600
LTA-NLM	.264	.217	.585	.558	.682	.666
ETA-NLM	**.287**	**.242**	.590	.562	.683	.665
LETA-NLM	.281	.236	**.615**	**.589**	**.694**	**.675**

P@N → precision score computed among the top N retrieved documents

LTA-NLM → IR using document representations from **pretrained** *LTA-NLM*

ETA-NLM → IR using document representations from **pretrained** *ETA-NLM*

LETA-NLM → IR using document representations from **pretrained** *LETA-NLM*

Bold values → best performing proposed model compared to *CNN-LSTM*

- Leveraging **latent**, **explainable** or **joint** topic representations **outperforms** the *NTM* model.
- *ETA-NLM* **outperforms** *LTA-NLM* which shows that the **ETR** provide **more descriptive document-level semantics** than **LTR**.

Evaluation: Information retrieval (Precision@recall scores)

Model	20NS		R21578		AGNews	
	P@5	P@10	P@5	P@10	P@5	P@10
NTM	.198	.190	.581	.567	.607	.600
LTA-NLM	.264	.217	.585	.558	.682	.666
ETA-NLM	**.287**	**.242**	.590	.562	.683	.665
LETA-NLM	.281	.236	**.615**	**.589**	**.694**	**.675**

P@N → precision score computed among the top N retrieved documents

LTA-NLM → IR using document representations from **pretrained** *LTA-NLM*

ETA-NLM → IR using document representations from **pretrained** *ETA-NLM*

LETA-NLM → IR using document representations from **pretrained** *LETA-NLM*

Bold values → best performing proposed model compared to *CNN-LSTM*

- Leveraging **latent**, **explainable** or **joint** topic representations **outperforms** the *NTM* model.
- *ETA-NLM* **outperforms** *LTA-NLM* which shows that the **ETR** provide **more descriptive document-level semantics** than **LTR**.
- *LETA-NLM* **performs best** by combining **LTR** and **ETR**

Analysis: Topic-aware Sentence generation (APNEWS)

TOPIC	GENERATED SENTENCE
fraud	prosecutors pleaded guilty in federal court in bank fraud conspiracy case .
fire	the fire was reported sunday night in the town of <unk> , about 20 miles northeast of los angeles .

Topic top key-terms

fraud	fire
fraud	flames
scheme	drowned
laundering	engulfed
fraudulent	firefighters
restitution	rescuers

Conclusion

- A novel **Neural Composite Language Modeling (NCLM)** framework for:
  - Composing **document-level LTR**, **ETR** and **LETR** representations into NLM
  - Composing **sentence-level topic representations** to account for **SDT**

- **State-of-the-art results** on 5 natural language tasks using 6 datasets
- **NCLM** framework is **model agnostic** and is applicable to any combination of **NTM** and **NLM** architectures.

CODE: https://github.com/YatinChaudhary/NCLM

Website: https://sites.google.com/view/yatin-chaudhary/home

Conclusion

- A novel **Neural Composite Language Modeling (NCLM)** framework for:
  - Composing **document-level LTR**, **ETR** and **LETR** representations into NLM
  - Composing **sentence-level topic representations** to account for **SDT**

- **State-of-the-art results** on 5 natural language tasks using 6 datasets
- **NCLM** framework is **model agnostic** and is applicable to any combination of **NTM** and **NLM** architectures.

CODE: https://github.com/YatinChaudhary/NCLM

Website: https://sites.google.com/view/yatin-chaudhary/home

Thank You