Probabilistic Neural Topic Models for Text Understanding

Gabriele Pergola

Department of Computer Science
University of Warwick

16th April, 2021

Outline

- Motivation
- Research Objectives
- Contributions
- Publications
- Discussion
- Future Work and Conclusion

Motivation

Making sense of text in **large corpora**

- User Reviews, Electronic Health Records, Social Networks, etc.

Motivation

Making sense of text in **large corpora**

- User Reviews, Electronic Health Records, Social Networks, etc.

Dealing with **linguistic features** across different domains and contexts

- User reviews might be characterized by colloquial *idioms*, *slang* or *contractions*

Motivation

Making sense of text in **large corpora**

- User Reviews, Electronic Health Records, Social Networks, etc.

Dealing with **linguistic features** across different domains and contexts

- User reviews might be characterized by colloquial ***idioms***, ***slang*** or ***contractions***

★★★★★

Our children didn't manage to clean their plates! Plenty of food!

★★

After one cycle the crockery is still dirty, it doesn't clean the plates even at full power.

Motivation

Making sense of text in **large corpora**

- User Reviews, Electronic Health Records, Social Networks, etc.

Dealing with **linguistic features** across different domains and contexts

- User reviews might be characterized by colloquial ***idioms***, ***slang*** or ***contractions***
- Clinical notes often contain ***technical jargon, multiword phrases***, medical ***abbreviations*** and ***polysemous*** terms

White blood cell
Shortness of breath
. . .

Motivation

Making sense of text in **large corpora**

- User Reviews, Electronic Health Records, Social Networks, etc.

Dealing with **linguistic features** across different domains and contexts

- User reviews might be characterized by colloquial ***idioms***, ***slang*** or ***contractions***
- Clinical notes often contain ***technical jargon, multiword phrases***, medical ***abbreviations*** and ***polysemous*** terms
- Authors mix their ***opinions*** with ***factual descriptions***

Overall, this is a good film and an excellent adaption. [..] the myriad inhabitants of Middle-earth, the legendary Rings of Power, and the fellowship of hobbits, elves, dwarfs, and humans–led by the wizard Gandalf (Ian McKellen) and the brave hobbit Frodo.

Motivation

- **Topic models** have established themselves as effective tools to generate concise and expressive representations of high volumes of documents.
    - *Global Context* - Document-based embedding (e.g. LDA, LSI).

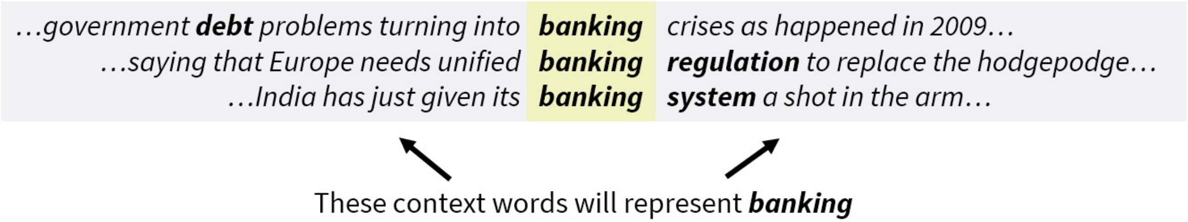
"An introduction to Topic models", M. J. Paul 2013

Motivation

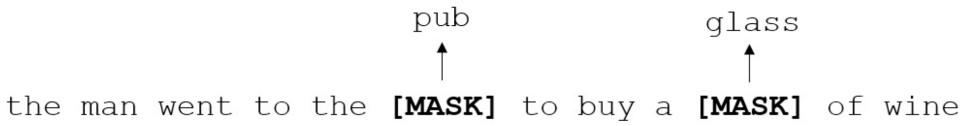
- **Topic models** have established themselves as effective tools to generate concise and expressive representations of high volumes of documents.
  - *Global Context* - Document-based embedding (e.g. LDA, LSI).

- Distributional representations of **word** and **language models**
  - *Local Context* - Window-based embedding (e.g. word2vec, BERT).

Word Embeddings



Masked Language Models



Research Objectives

Our hypothesis is that **topic models** and **neural architectures** are a suitable combination for capturing text semantic, with benefits connected to several NLP open problems.

RO 1 Combining global and local context of words

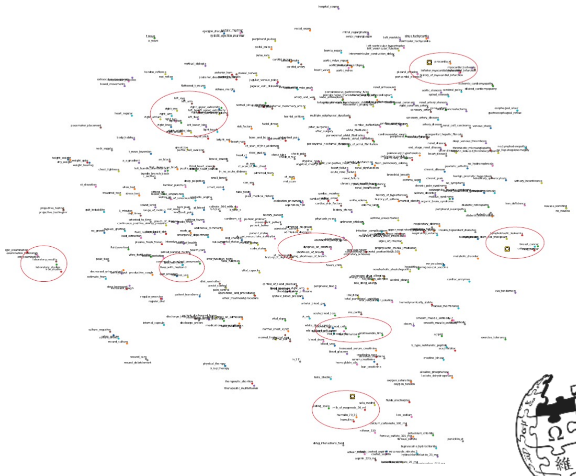
*…government **debt** problems turning into* ***banking*** *crises as happened in 2009…*
…saying that Europe needs unified ***banking*** ***regulation*** *to replace the hodgepodge…*
…India has just given its ***banking*** ***system*** *a shot in the arm…*

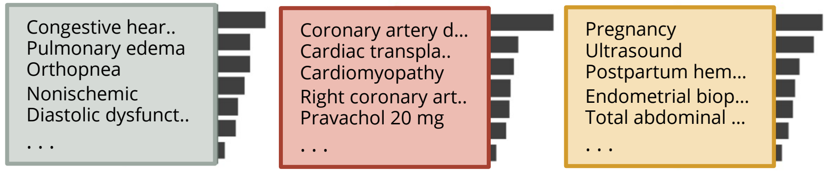
Research Objectives

Our hypothesis is that **topic models** and **neural architectures** are a suitable combination for capturing text semantic, with benefits connected to several NLP open problems.

RO 1 Combining global and local context of words

*…government **debt** problems turning into* ***banking*** *crises as happened in 2009…*
…saying that Europe needs unified ***banking*** ***regulation*** *to replace the hodgepodge…*
…India has just given its ***banking*** ***system*** *a shot in the arm…*





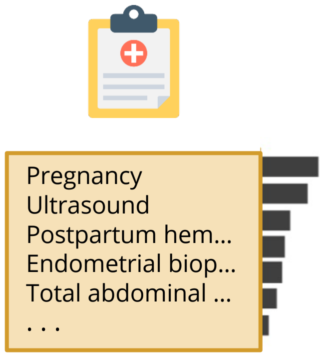
Research Objectives

Our hypothesis is that **topic models** and **neural architectures** are a suitable combination for capturing text semantic, with benefits connected to several NLP open problems.

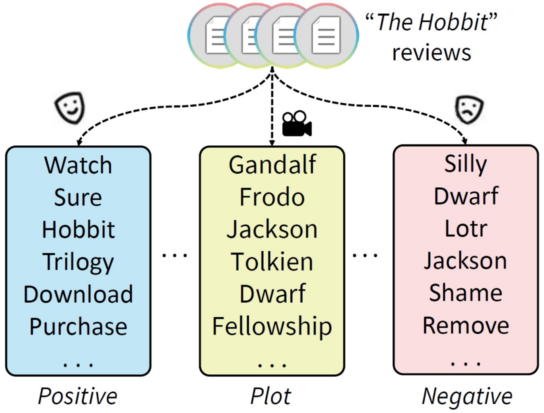
RO 1 Combining global and local context of words

RO 2 Generating fine-grained topics.

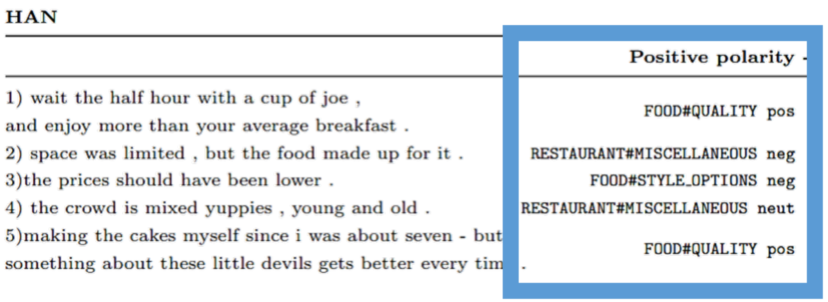
Topical phrases



Polarity-disentangled Topics



Sentiment-oriented topics



Research Objectives

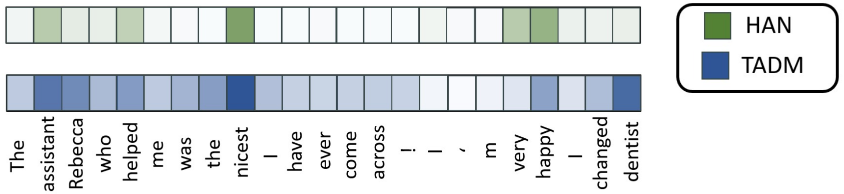
Our hypothesis is that **topic models** and **neural architectures** are a suitable combination for capturing text semantic, with benefits connected to several NLP open problems.

RO 1 Combining global and local context of words

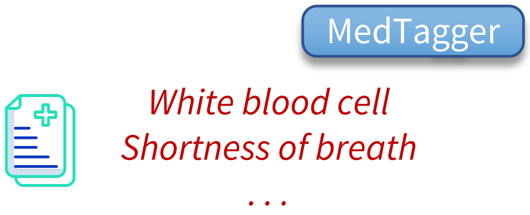
RO 2 Generating fine-grained topics.

RO 3 Incorporating unstructured knowledge.

Word Polarity



Domain-specific lexicons and concepts



Research Objectives

Our hypothesis is that **topic models** and **neural architectures** are a suitable combination for capturing text semantic, with benefits connected to several NLP open problems.

RO 1 Combining global and local context of words

RO 2 Generating fine-grained topics.

RO 3 Incorporating unstructured knowledge.

RO 4 Incorporating structured knowledge.

Domain-specific lexicons and concepts

PubMed

Patients ENTITY with diabetes ENTITY (HR ENTITY 1.59) were more likely to reach to the composite ENTITY endpoints ENTITY than those without.

Research Objectives

Our hypothesis is that **topic models** and **neural architectures** are a suitable combination for capturing text semantic, with benefits connected to several NLP open problems.

RO 1 Combining global and local context of words

RO 2 Generating fine-grained topics.

RO 3 Incorporating unstructured knowledge.

RO 4 Incorporating structured knowledge.

RO 5 Evaluation on downstream tasks.

Sentiment Classification

Aspect Extraction

Biomedical QA

Contributions

"TDAM: a Topic-Dependent Attention Model for Sentiment Analysis"

Information Processing and Management, 2019

	Research Objectives	Contributions
RO 1	Combining global and local context of words	**External memory** keeping track of word co-occurrences
RO 2	Generating fine-grained topics.	**Aspect** extraction through sentiment topics
RO 3	Incorporating unstructured knowledge.	**Word embeddings** induced from sentiment polarity
RO 4	Incorporating structured knowledge.	
RO 5	Evaluation on downstream tasks.	Sentiment classification and Aspect Extraction

Contributions

"A Disentangled Adversarial Neural Topic Model for Separating Opinions from Plots in User Reviews"

North American Chapter of the Association for Computational Linguistics (NAACL), 2021

Research Objectives		Contributions
RO 1	Combining global and local context of words	Initializing with **pre-training word embedding**
RO 2	Generating fine-grained topics.	Generation of **disentangled** topics
RO 3	Incorporating unstructured knowledge.	
RO 4	Incorporating structured knowledge.	
RO 5	Evaluation on downstream tasks.	**MOBO dataset** and introduction of **disentangling rate** for topics

Contributions

"Topical Phrase Extraction from Clinical Reports by Incorporating both Local and Global Context"

The 2nd AAAI Workshop on Health Intelligence, AAAI 2018

Research Objectives		Contributions
RO 1	Combining global and local context of words	Modified **inference process** joining word embedding information
RO 2	Generating fine-grained topics.	Identification of **topical phrases** encoded by character-based word embedding
RO 3	Incorporating unstructured knowledge.	
RO 4	Incorporating structured knowledge.	**Domain-specific** word embedding
RO 5	Evaluation on downstream tasks.	Identification of **topical phrases** encoded by character-based word embedding

Contributions

"Boosting Low-Resource Biomedical QA via Entity-Aware Masking Strategies"

European Chapter of the Association for Computational Linguistics (EACL) 2021

Research Objectives	Contributions
RO 1 Combining global and local context of words	
RO 2 Generating fine-grained topics.	
RO 3 Incorporating unstructured knowledge.	
RO 4 Incorporating structured knowledge.	**Biomedical entity detection** via a scientific NER model
RO 5 Evaluation on downstream tasks.	Impact on several language models

Publications

“A Disentangled Adversarial Neural Topic Model for Separating Opinions from Plots in User Reviews”

North American Chapter of the Association for Computational Linguistics (NAACL), 2021

“Boosting Low-Resource Biomedical QA via Entity-Aware Masking Strategies”

European Chapter of the Association for Computational Linguistics (EACL), 2021

“TDAM: a Topic-Dependent Attention Model for Sentiment Analysis”

Information Processing and Management, 2019

“Topical Phrase Extraction from Clinical Reports by Incorporating both Local and Global Context”

The 2nd AAAI Workshop on Health Intelligence (AAAI), 2018

Publications - 2

“Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews”

Runcong Zhao, Lin Gui, Gabriele Pergola, Yulan He.

European Chapter of the Association for Computational Linguistics (EACL), 2021

“CHIME: Crosspassage Hierarchical Memory Network for Generative Review Question Answering”

Junru Lu, Gabriele Pergola, Lin Gui, Binyang Li, Yulan He.

The 28th International Conference on Computational Linguistics (COLING), 2020

“Neural Topic Model with Reinforcement Learning”

Lin Gui, Jia Leng, Gabriele Pergola, Yu Zhou, Ruifeng Xu, Yulan He.

The 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019

TDAM: a Topic-Dependent Attention Model for Sentiment Analysis

Gabriele Pergola, Lin Gui, Yulan He

University of Warwick

Information Processing and Management
2019

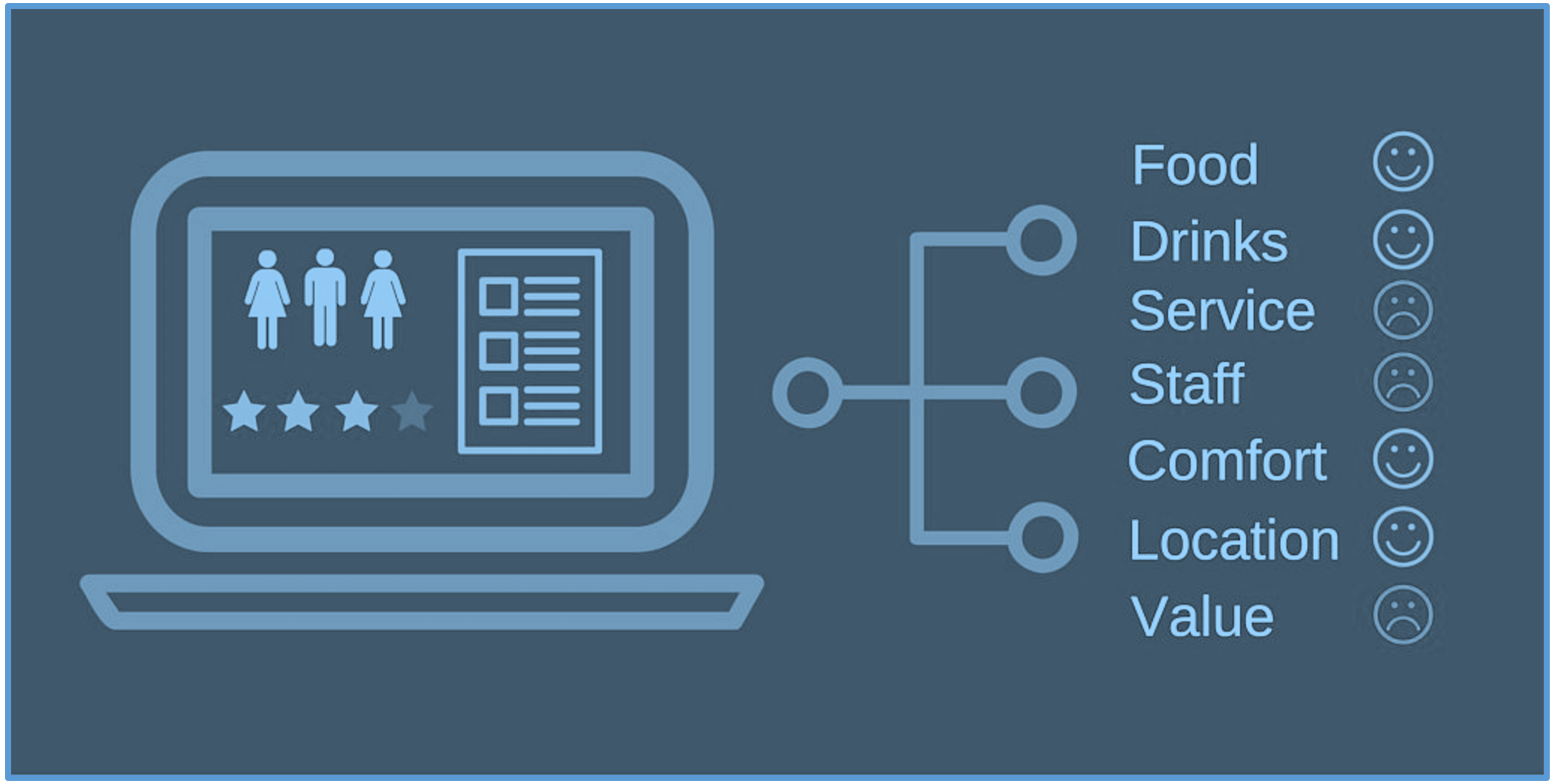
Chosen for poster presentation at

EurNLP

Outline

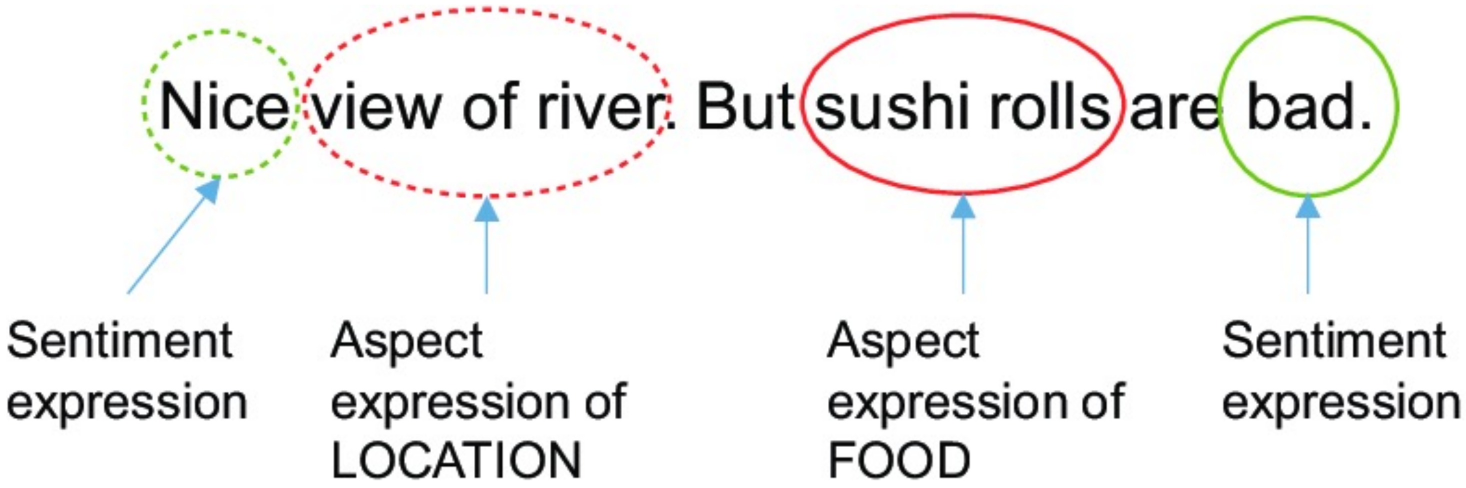
- What is Sentiment Analysis? What are “Topics” ?
- Problem - How to Extract Polarity-Bearing Topics from Text using Neural Models
- The Consciousness Prior
- TDAM: A Topic-Dependent Attention Model
- Summary of Results

Sentiment Analysis



Credits: https://medium.com/@nitesh10126/aspect-based-sentiment-analysis-in-product-reviews-unsupervised-way

Aspect-Based Sentiment Analysis



Topic Modeling - 1

"An introduction to Topic models", M. J. Paul 2013

Topic Modeling - 2

Sentiment analysis meets topic modeling

Word representations should depend on the ***topical context*** in which they appear in.
This context emerges from the co-occurrence of words across documents.

Our children didn’t manage to clean their plates! Plenty of food!

After one cycle the crockery is still dirty, it doesn’t clean the plates even at full power.

Polarity-bearing Topics from Neural Models

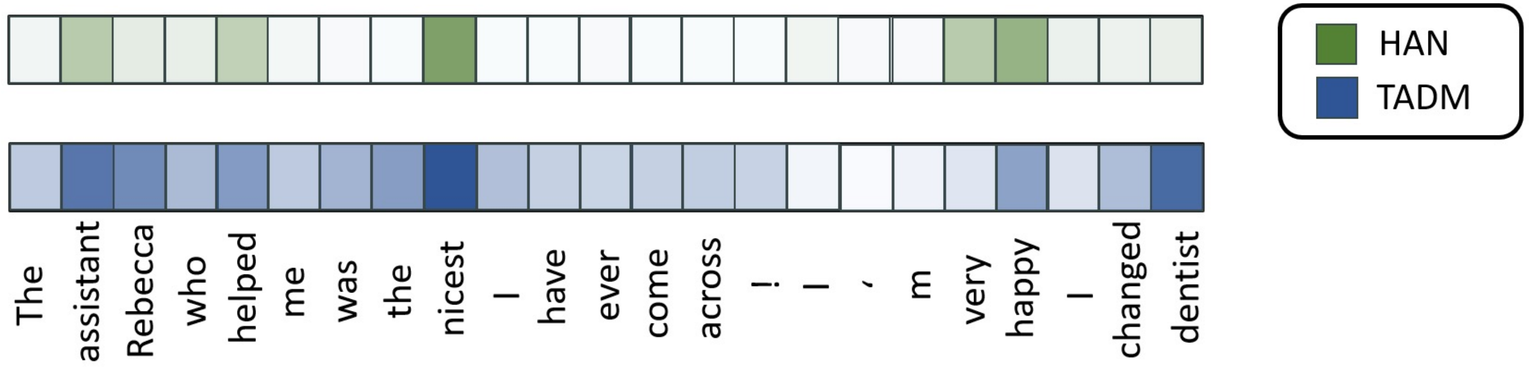
PROBLEM

We want to extract **topics** by looking at the **attention weights** learned by neural models; however, there is no mechanism to separate words into multiple clusters representing polarity-bearing topics.

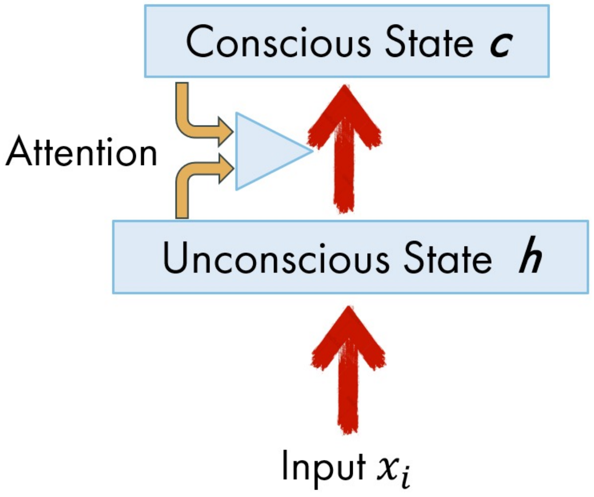
Polarity-bearing Topics from Neural Models

PROBLEM

We want to extract **topics** by looking at the **attention weights** learned by neural models; however, there is no mechanism to separate words into multiple clusters representing polarity-bearing topics.



The Consciousness Prior



<< [...] the consciousness RNN ***C()*** a tool for exploring interpretations or plans or to sample predictions about the future. We can also think of the consciousness RNN as the tool to isolate a particular high-level abstraction and extract the information about it >> [1].

[1]: "The Consciousness Prior", Bengio 2017, arXiv:1709.08568

TopicGRU – Attention to extract topics

Representation State

$$h_t = F(s_t, h_{t-1})$$

Modified GRU

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + V_r \boxed{\mathbf{q_{t-1}}})$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + V_z \boxed{\mathbf{q_{t-1}}})$$

$$\hat{h}_t = \tanh(W_h x_t + r_t \odot (U_h h_{t-1} + V_h \boxed{\mathbf{q_{t-1}}}))$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t$$

Conscious State

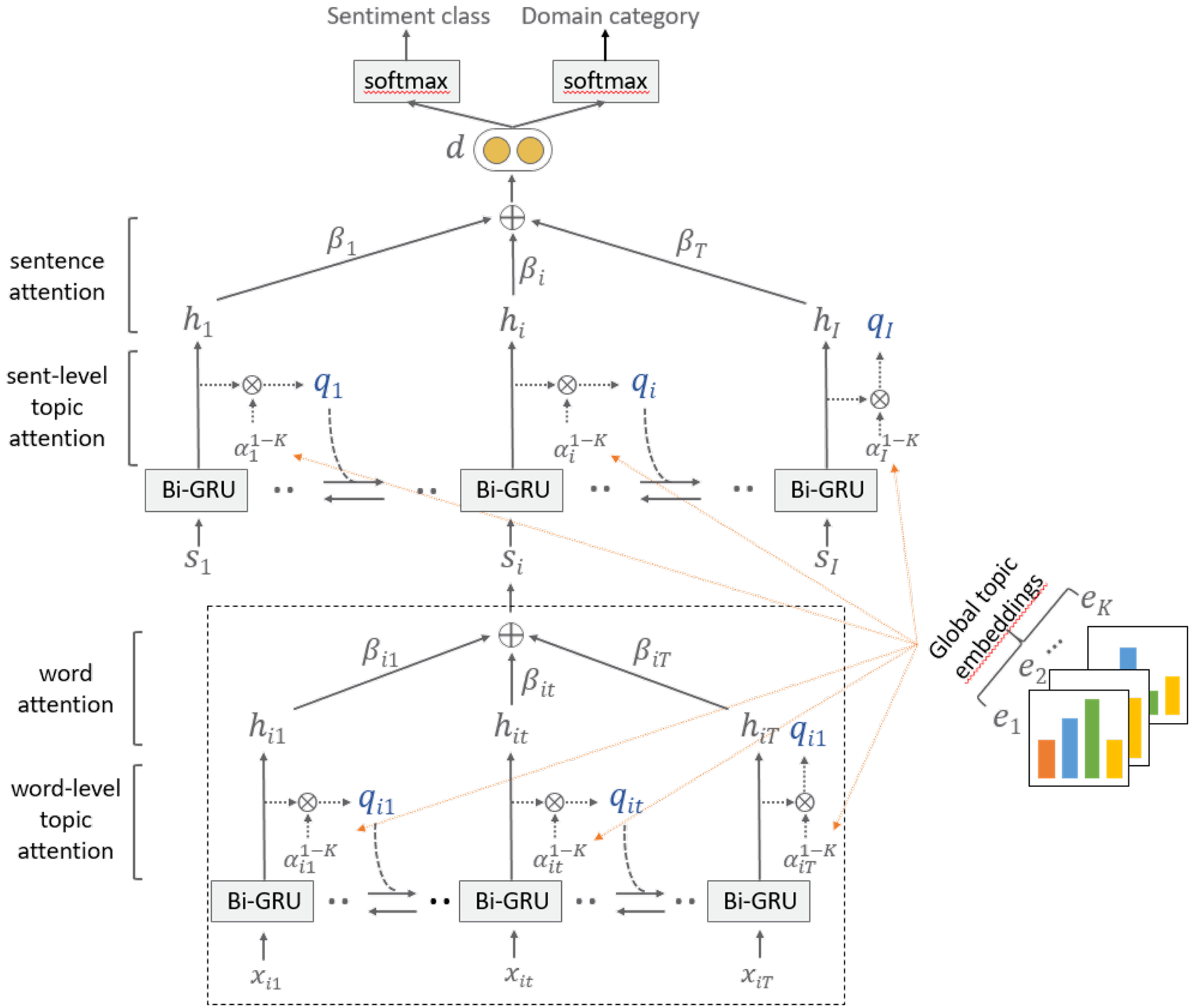
$$\boxed{c_t} = C(h_t, c_{t-1}, z_t)$$

$$u_{it} = \tanh(W_w h_{it})$$

$$\alpha_{it}^k = \mathrm{softmax}(u_{it}^\intercal e_k)$$

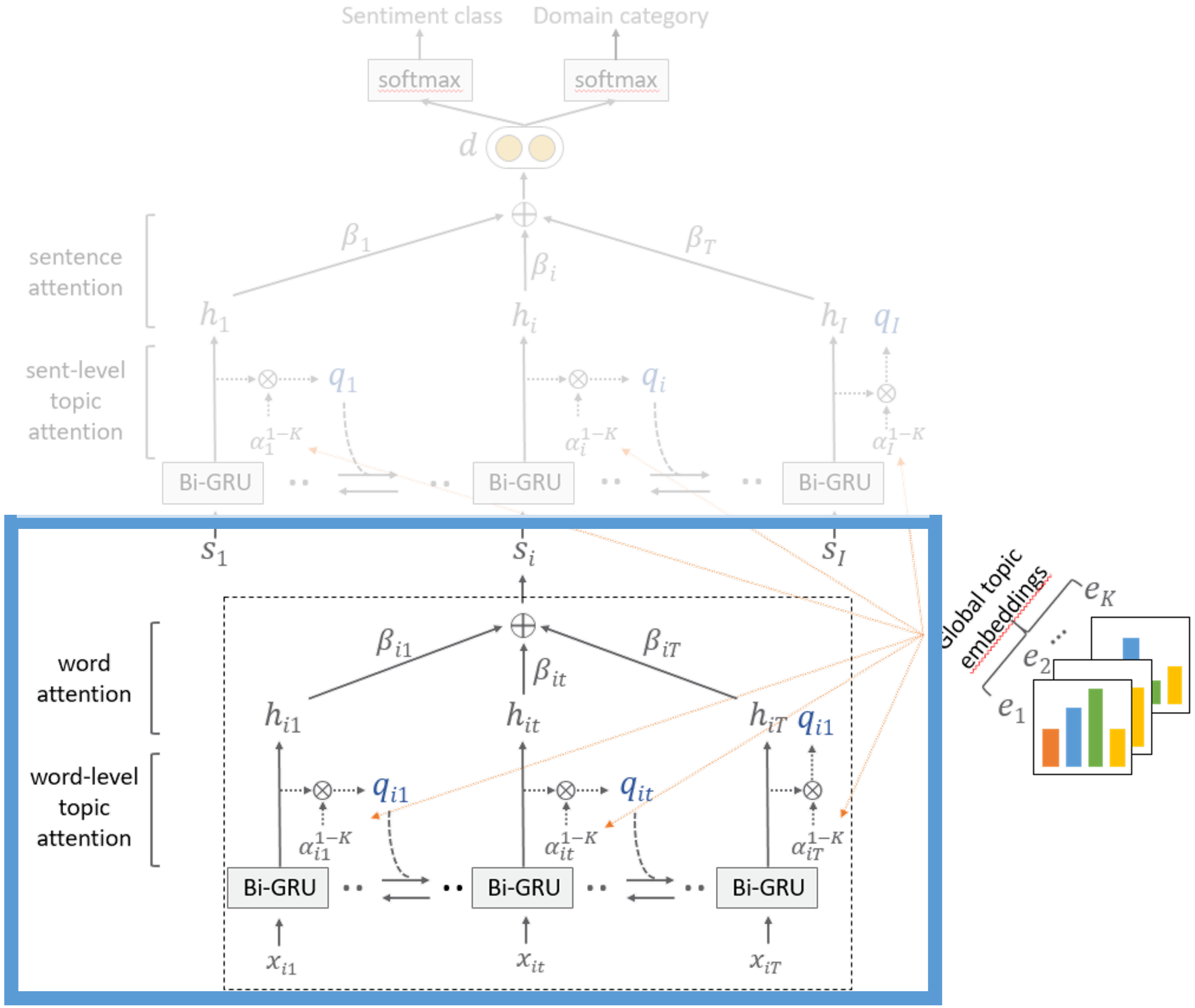
$$\boxed{q_{it}} = \sum_{k=1}^{K} \alpha_{it}^k \otimes e_k$$

Topic Dependent Attention Model



Topic Dependent Attention Model

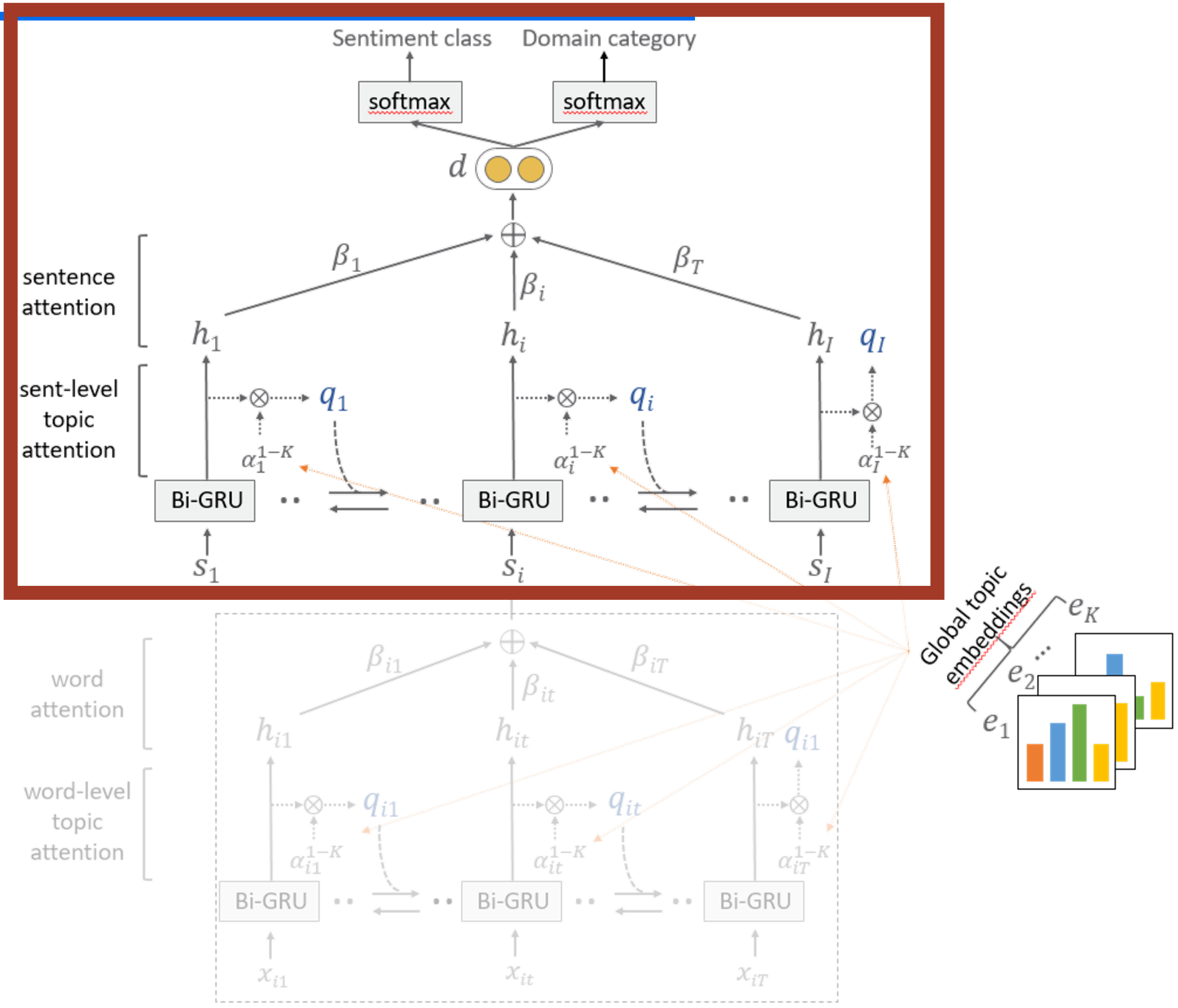
FROM WORDS
TO
SENTENCES



Topic Dependent Attention Model

FROM SENTENCES
TO
DOCUMENTS

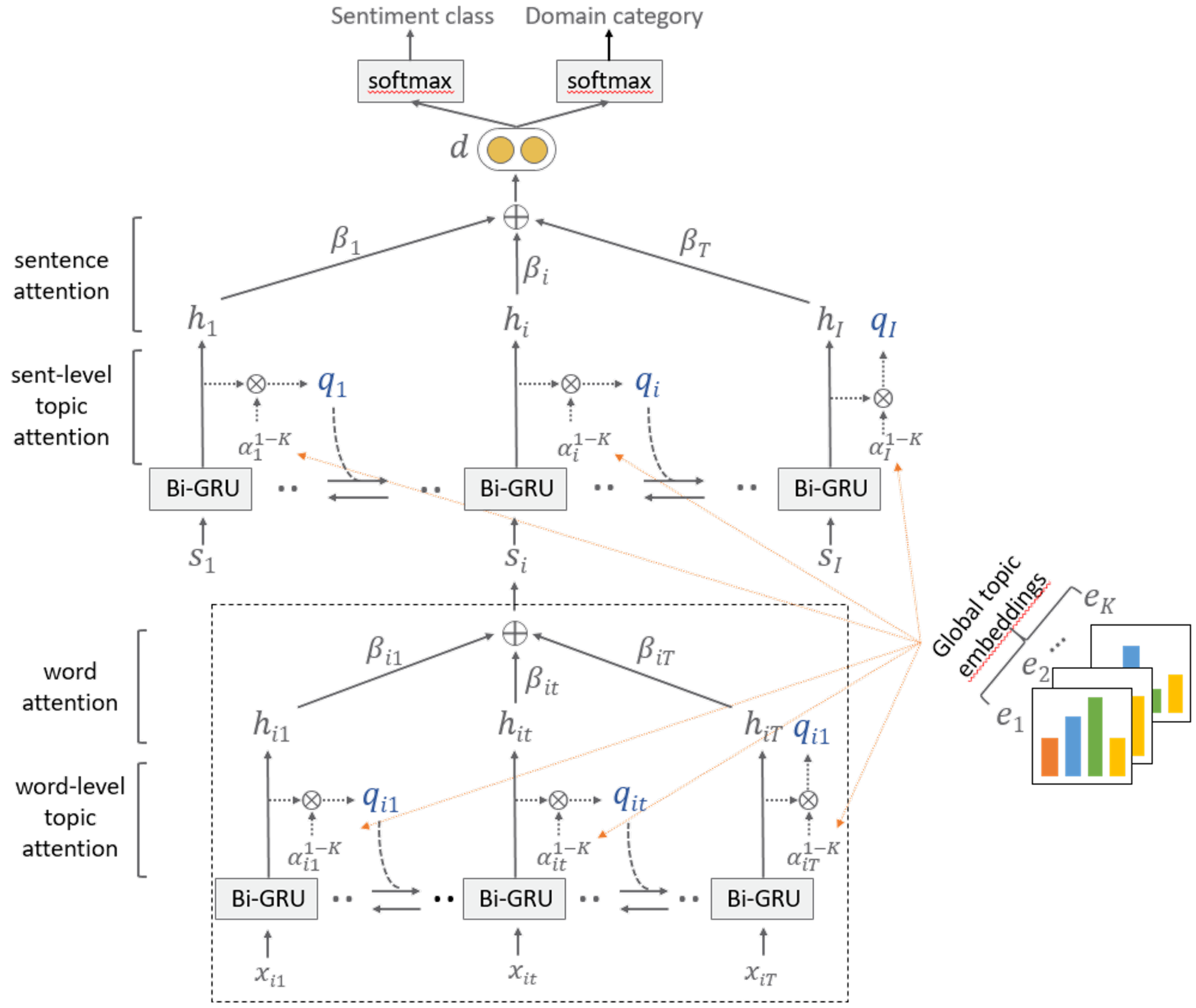
FROM WORDS
TO
SENTENCES



Topic Dependent Attention Model

FROM SENTENCES
TO
DOCUMENTS

FROM WORDS
TO
SENTENCES



Sentiment Classification

Methods	Yelp 18	Amazon
BiLSTM	74.5 ± 0.2	72.1 ± 0.2
BiLSTM - Mtl	74.2 ± 0.2	71.8 ± 0.1
BiGRU	75.5 ± 0.1	72.5 ± 0.3
BiGRU - Mtl	75.4 ± 0.2	72.1 ± 0.3
HAN	83.7 ± 0.2	78.4 ± 0.2
HAN - Mtl	83.6 ± 0.3	78.2 ± 0.3
S-LDA	70.8 ± 0.2	64.6 ± 0.1
SCHOLAR	77.3 ± 0.2	71.4 ± 0.2
TDAM	84.2 ± 0.2	78.9 ± 0.2
TDAM - Mtl	$\mathbf{84.5} \pm 0.3$	$\mathbf{79.1} \pm 0.2$

Sentiment Classification

Methods	Yelp 18	Amazon
BiLSTM	74.5 ± 0.2	72.1 ± 0.2
BiLSTM - Mtl	74.2 ± 0.2	71.8 ± 0.1
BiGRU	75.5 ± 0.1	72.5 ± 0.3
BiGRU - Mtl	75.4 ± 0.2	72.1 ± 0.3
HAN	83.7 ± 0.2	78.4 ± 0.2
HAN - Mtl	83.6 ± 0.3	78.2 ± 0.3
S-LDA	70.8 ± 0.2	64.6 ± 0.1
SCHOLAR	77.3 ± 0.2	71.4 ± 0.2
TDAM	84.2 ± 0.2	78.9 ± 0.2
TDAM - Mtl	$\mathbf{84.5} \pm 0.3$	$\mathbf{79.1} \pm 0.2$

Aspect-based Sentiment Analysis

HAN		TDAM	
Positive polarity - Food#Quality			
1) wait the half hour with a cup of joe , and enjoy more than your average breakfast .	FOOD#QUALITY pos	1) the food was all good but it was way too	FOOD#QUALITY neg
2) space was limited , but the food made up for it .	RESTAURANT#MISCELLANEOUS neg	2) the pizza 's are light and scrumptious .	FOOD#STYLE_OPTIONS pos
3)the prices should have been lower .	FOOD#STYLE_OPTIONS neg	3) the food is great and they make a mean bloody mary .	FOOD#QUALITY pos
4) the crowd is mixed yuppies , young and old .	RESTAURANT#MISCELLANEOUS neut	4) great draft and bottle selection and the pizza rocks .	FOOD#QUALITY pos
5)making the cakes myself since i was about seven - but something about these little devils gets better every time .	FOOD#QUALITY pos	5) the food is simply unforgettable !	FOOD#QUALITY pos

Negative polarity - Food#Quality			
1) the pancakes were certainly inventive but $ 8.50 for 3 - 6 " pancakes (one of them was more like 5 ")	FOOD#STYLE_OPTIONS neg	1) i may not be a sushi guru	FOOD#QUALITY neg
2)a beautiful assortment of enormous white gulf prawns , smoked albacore tuna, [..] and a tiny pile of dungeness	FOOD#STYLE_OPTIONS pos	2) rice is too dry , tuna was n't so fresh either .	FOOD#QUALITY neg
3) space was limited , but the food made up for it .	RESTAURANT#MISCELLANEOUS neg	3)the only way this place survives with such average food is because most customers are one-time customer tourists	FOOD#QUALITY neg
4) the portions are big though , so do not order too much .	FOOD#STYLE_OPTIONS neut	4) the portions are big though , so do not order too much .	FOOD#STYLE_OPTIONS neut
5) not the biggest portions but adequate .	FOOD#QUALITY pos	5) the only drawback is that this place is really expensive and the portions are on the small side .	RESTAURANT#PRICES neg

Detecting **aspect** (e.g. food, service, options) and **polarity** (positive, negative, neutral)

Aspect-based Sentiment Analysis

HAN | TDAM

Positive polarity - Food#Quality			
1) wait the half hour with a cup of joe , and enjoy more than your average breakfast .	FOOD#QUALITY pos	1) the food was all good but it was way too	FOOD#QUALITY neg
2) space was limited , but the food made up for it .	RESTAURANT#MISCELLANEOUS neg	2) the pizza 's are light and scrumptious .	FOOD#STYLE_OPTIONS pos
3)the prices should have been lower .	FOOD#STYLE_OPTIONS neg	3) the food is great and they make a mean bloody mary .	FOOD#QUALITY pos
4) the crowd is mixed yuppies , young and old .	RESTAURANT#MISCELLANEOUS neut	4) great draft and bottle selection and the pizza rocks .	FOOD#QUALITY pos
5)making the cakes myself since i was about seven - but something about these little devils gets better every tim .	FOOD#QUALITY pos	5) the food is simply unforgettable !	FOOD#QUALITY pos

Negative polarity - Food#Quality			
1) the pancakes were certainly inventive but $ 8.50 for 3 - 6 " pancakes (one of them was more like 5 ")	FOOD#STYLE_OPTIONS neg	1) i may not be a sushi guru	FOOD#QUALITY neg
2)a beautiful assortment of enormous white gulf prawns , smoked albacore tuna, [..] and a tiny pile of dungeness	FOOD#STYLE_OPTIONS pos	2) rice is too dry , tuna was n't so fresh either .	FOOD#QUALITY neg
3) space was limited , but the food made up for it .	RESTAURANT#MISCELLANEOUS neg	3) the only way this place survives with such average food is because most customers are one-time customer tourists	FOOD#QUALITY neg
4) the portions are big though , so do not order too much .	FOOD#STYLE_OPTIONS neut	4) the portions are big though , so do not order too much .	FOOD#STYLE_OPTIONS neut
5) not the biggest portions but adequate .	FOOD#QUALITY pos	5) the only drawback is that this place is really expensive and the portions are on the small side .	RESTAURANT#PRICES neg

Detecting **aspect** (e.g. food, service, options) and **polarity** (positive, negative, neutral)

Aspect-based Sentiment Analysis

HAN		TDAM	
Positive polarity - Food#Quality			
1) wait the half hour with a cup of joe , and enjoy more than your average breakfast .	FOOD#QUALITY pos	1) the food was all good but it was way too	FOOD#QUALITY neg
2) space was limited , but the food made up for it .	RESTAURANT#MISCELLANEOUS neg	2) the pizza 's are light and scrumptious .	FOOD#STYLE_OPTIONS pos
3)the prices should have been lower .	FOOD#STYLE_OPTIONS neg	3) the food is great and they make a mean bloody mary .	FOOD#QUALITY pos
4) the crowd is mixed yuppies , young and old .	RESTAURANT#MISCELLANEOUS neut	4) great draft and bottle selection and the pizza rocks .	FOOD#QUALITY pos
5)making the cakes myself since i was about seven - but something about these little devils gets better every time .	FOOD#QUALITY pos	5) the food is simply unforgettable !	FOOD#QUALITY pos

Negative polarity - Food#Quality			
1) the pancakes were certainly inventive but $ 8.50 for 3 - 6 " pancakes (one of them was more like 5 ")	FOOD#STYLE_OPTIONS neg	1) i may not be a sushi guru	FOOD#QUALITY neg
2)a beautiful assortment of enormous white gulf prawns , smoked albacore tuna, [..] and a tiny pile of dungeness	FOOD#STYLE_OPTIONS pos	2) rice is too dry , tuna was n't so fresh either .	FOOD#QUALITY neg
3) space was limited , but the food made up for it .	RESTAURANT#MISCELLANEOUS neg	3) the only way this place survives with such average food i because most customers are one-time customer tourists	FOOD#QUALITY neg
4) the portions are big though , so do not order too much .	FOOD#STYLE_OPTIONS neut	4) the portions are big though , so do not order too much	FOOD#STYLE_OPTIONS neut
5) not the biggest portions but adequate .	FOOD#QUALITY pos	5) the only drawback is that this place is really expensive and the portions are on the small side .	RESTAURANT#PRICES neg

Detecting **aspect** (e.g. food, service, options) and **polarity** (positive, negative, neutral)

Conclusion

- A possible extension of the hierarchical model is to introduce an intermediate level between words and sentences, defining a **discourse-level layer**.
  This discourse-level layer would determine the **elementary discourse units (EDUs).**

- Use a **Neural Topic Model** to jointly update the external memory inferring topical representations.

- Employment of **Contextualized Language Models** in place of GloVe embedding.

Disentangled Adversarial Topic Model

Gabriele Pergola, Lin Gui, Yulan He
University of Warwick

"*A Disentangled Adversarial Neural Topic Model for Separating Opinions from Plots in User Reviews*"

North American Chapter of the Association for Computational Linguistics (NAACL)

2021

Problem: **Disentangling Topics**

- The application of topic models to text with users' opinions (e.g. book reviews) often leads to the generation of **topics mixing different aspects.** For example, information about book's plot or characters gets mixed with user's feelings and opinions.

- Therefore, we want to separate **polarity-bearing topics** from **neutral topics.**
  Providing features disentangling different aspects in text.

- As a case study, we applied our model to movie and book reviews to disentangle topics about their **plots** and topics about the users' **opinions**.
  Fig. 1 shows an example of polarity-disentangled topics.

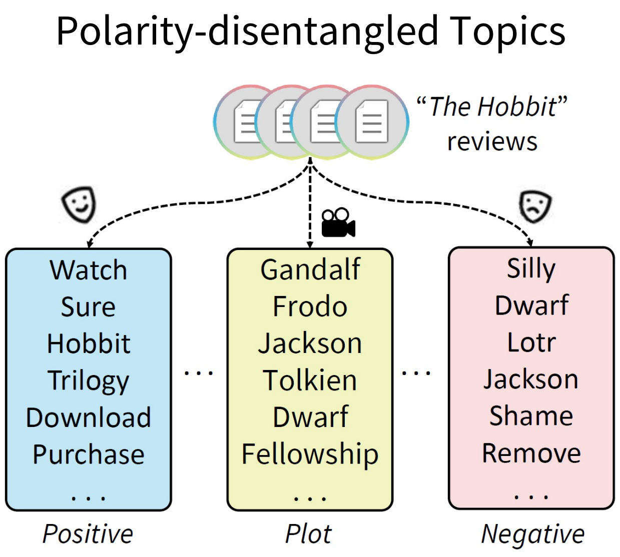
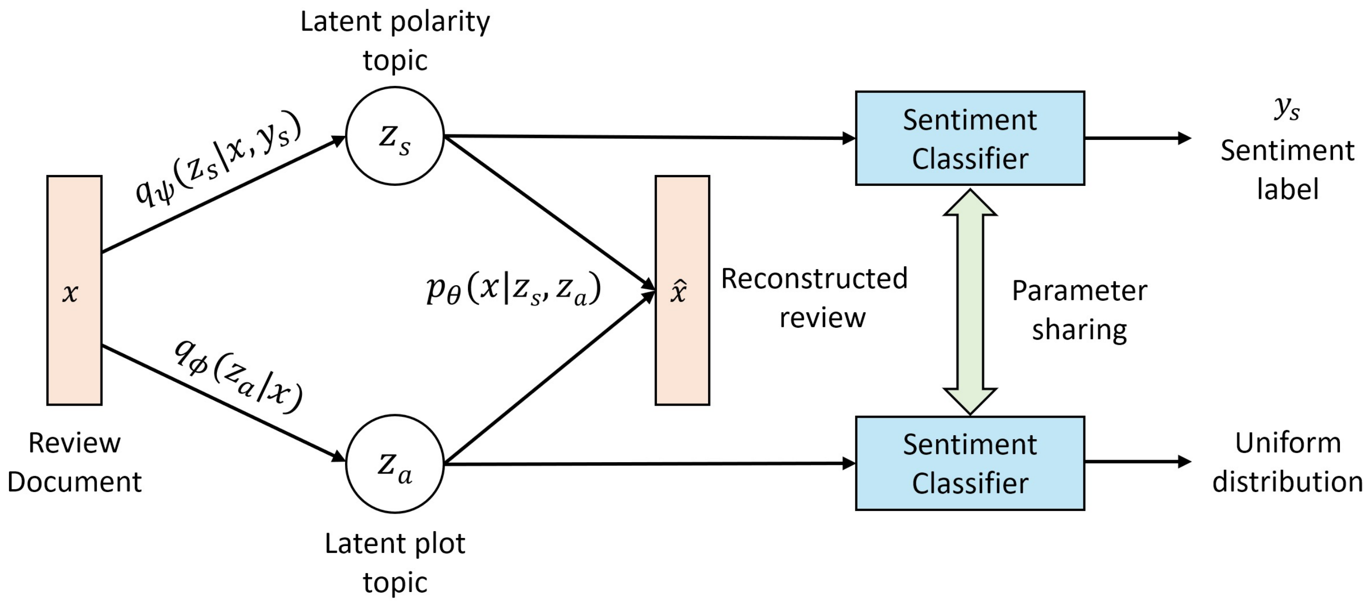


Fig. 1: Disentangled topics extracted by DIATOM from the Amazon reviews for "The Hobbit".

Model: **Neural Topic Model with Adversarial Training**

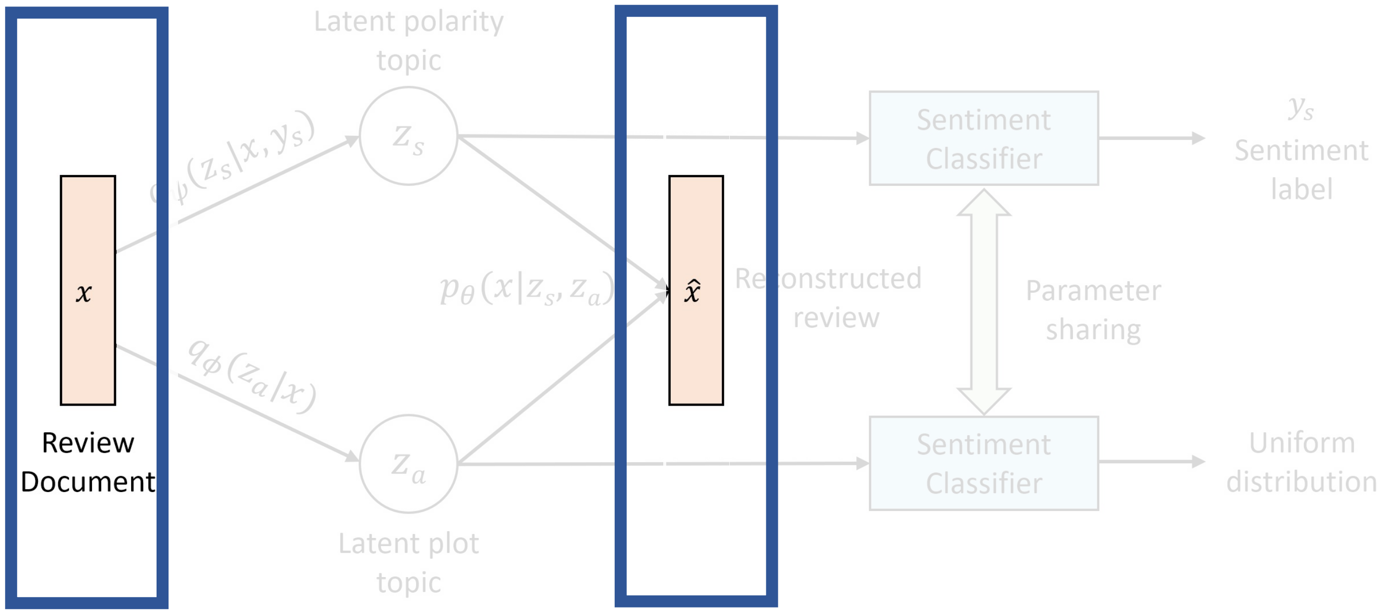
We propose to combine a **Neural Topic Model** architecture with **adversarial training** to disentangle polarity-bearing topics from neutral ones.



Model: **Neural Topic Model with Adversarial Training**

We propose to combine a **Neural Topic Model** architecture with **adversarial training** to disentangle polarity-bearing topics from neutral ones.

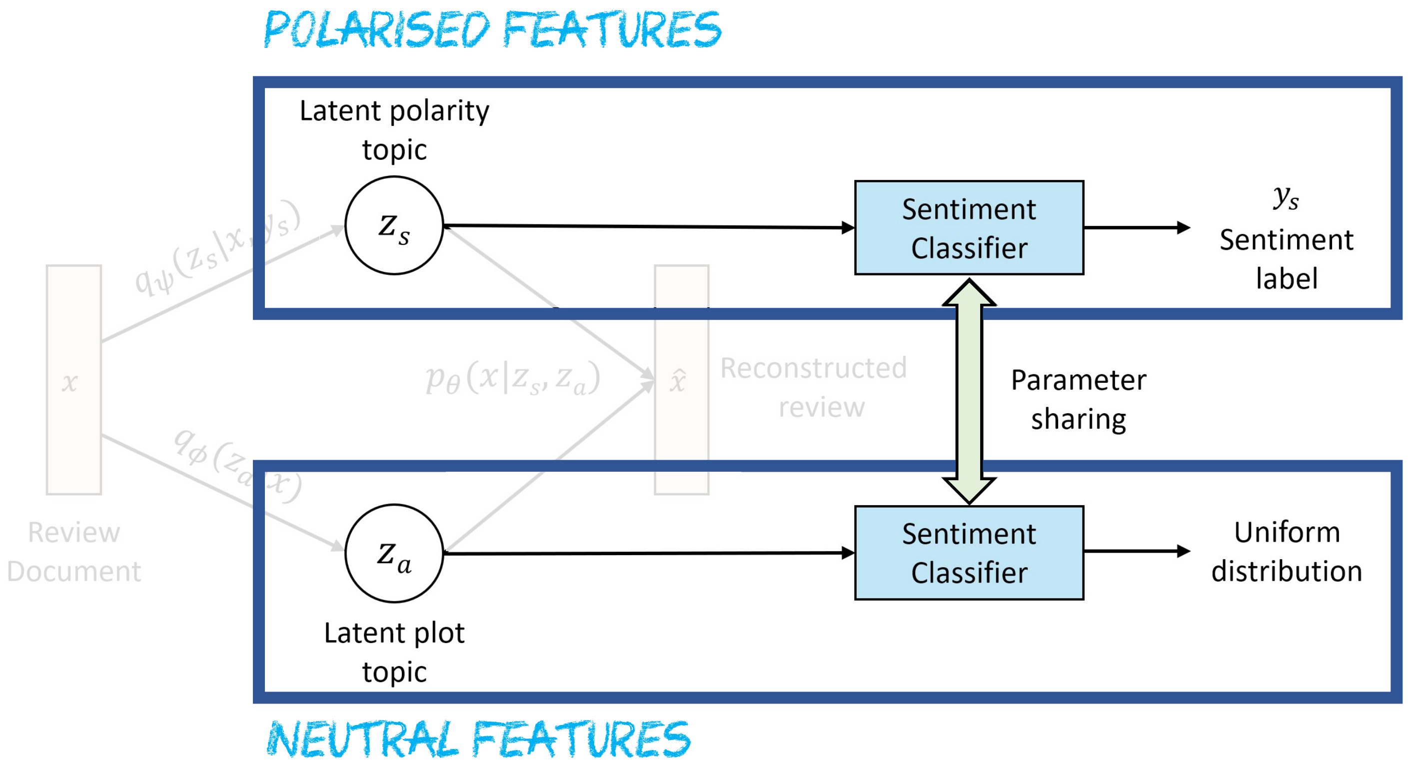
- The **VAE architecture** extracts topics through a bottleneck mechanism, reading and reconstructing the documents



Model: **Neural Topic Model with Adversarial Training**

We propose to combine a **Neural Topic Model** architecture with **adversarial training** to disentangle polarity-bearing topics from neutral ones.

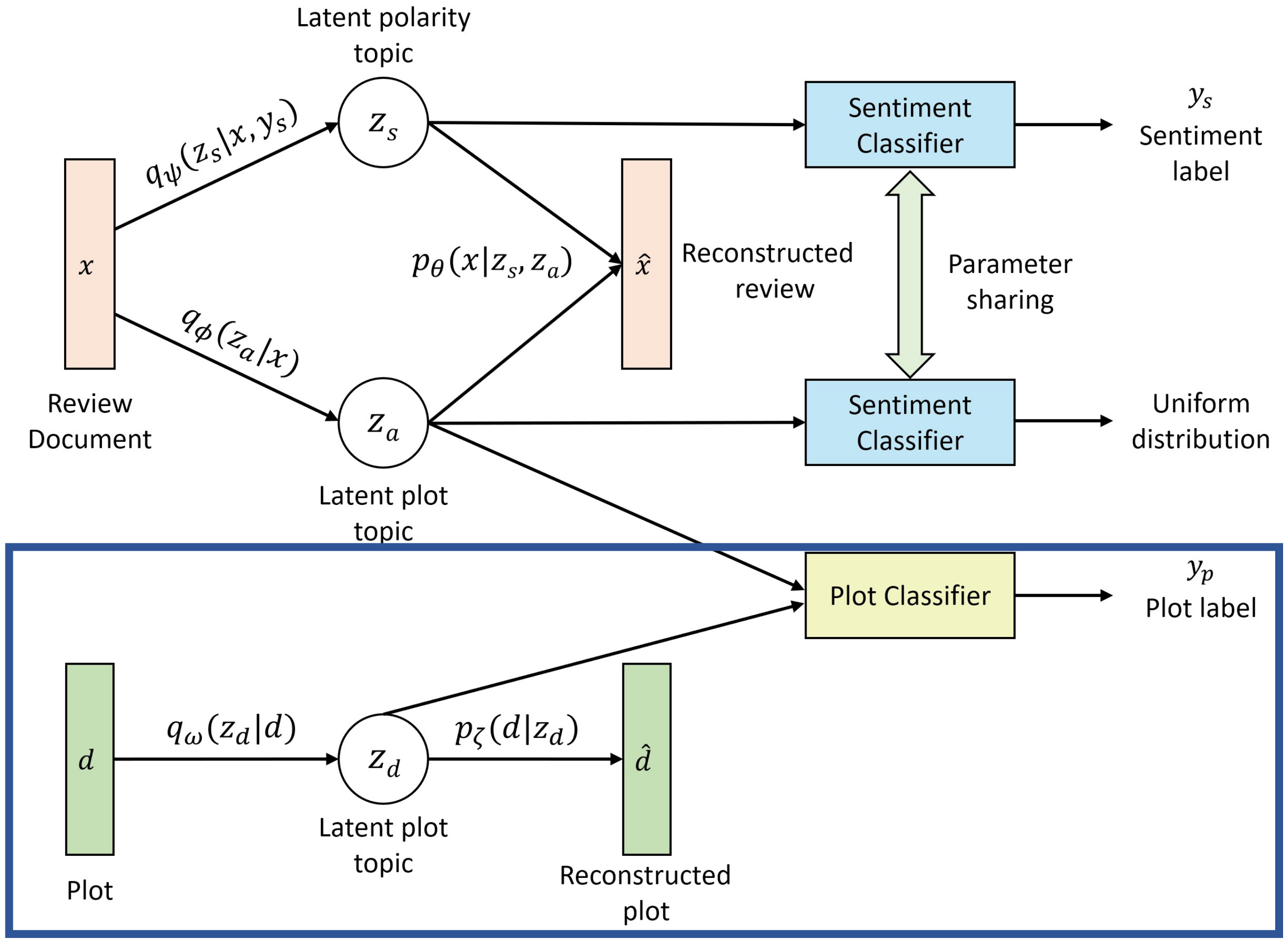
- The **VAE architecture** extracts topics through a bottleneck mechanism, reading and reconstructing the documents
- An **adversarial mechanism** separate polarised and neutral features, corresponding to polarised and neutral topics



Model: **Neural Topic Model with Adversarial Training**

We propose to combine a **Neural Topic Model** architecture with **adversarial training** to disentangle polarity-bearing topics from neutral ones.

- The **VAE architecture** extracts topics through a bottleneck mechanism, reading and reconstructing the documents
- An **adversarial mechanism** separate polarised and neutral features, corresponding to polarised and neutral topics
- An analogous mechanism induces a prevalence of **plot topics** among all the possible neutral aspects.



Results: **the MOBO Dataset**

- We also introduce a new dataset, namely the **MOBO dataset**, made up of **MO**vie and **BO**ok reviews, paired with their related plots. The reviews come from different publicly available datasets: IMDB, GoodReads and Amazon reviews.

- Additionally, around **15.000 sentences** of these reviews were manually annotated and used (only) during the evaluation phase to compute the disentanglement rate between polarity-bearing topics and plot (or neutral) topics.

Statistics	**IMDB**	**GoodReads**	**Amazon**
No. of plots	1,131	150	100
No. of reviews	25,836	83,852	32,375
No. of reviews / plot (avg / max / min)	24 / 30 / 10	954 / 3,000 / 549	464 / 1525 / 272
Pos / Neg / Neutral distribution	0.46 / 0.54 / 0	0.33 / 0.50 / 0.17	0.32 / 0.46 / 0.22
Training set	20,317	65,816	25,883
Development set	2,965	9,007	3,275
Test set	2,554	9,029	3,217
No. of annotated sentences	6,000	6,000	6,000

Results: **Topic Evaluation - 1**

- To automatically evaluate whether a topic is polarised or neutral, we perform **topic labelling** using the annotated sentences in the MOBO dataset.

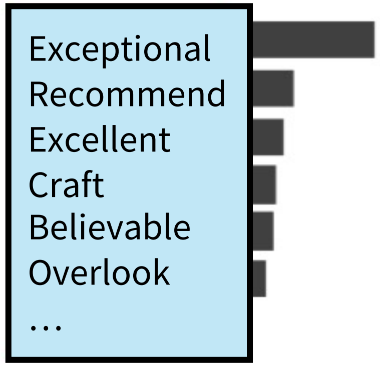
- For each topic we compute a **topic embedding** taking the normalized weighted average of the word embeddings, and we then retrieve the top 10 most similar human-annotated sentences via cosine similarity between the topic embedding and each **sentence embedding**, where the sentence embedding is computed using the *Sentence-BERT* encoder.

- The most frequent label among the retrieved sentences is adopted as the topic's label.

Results: **Topic Evaluation - 2**

0.619 PLOT the myriad inhabitants of Middle-earth, the legendary Rings of Power, and the fellowship of **hobbits**, elves, **dwarfs**, and humans–led by the wizard **Gandalf** (Ian McKellen) and the brave **hobbit Frodo**.

0.523 PLOT This is the beginning of a trilogy; soon to be finalized.



0.597 POSITIVE Overall, this is a good film and an excellent adaption.

0.495 POSITIVE It's great acting, superb cinematography and excellent writing.

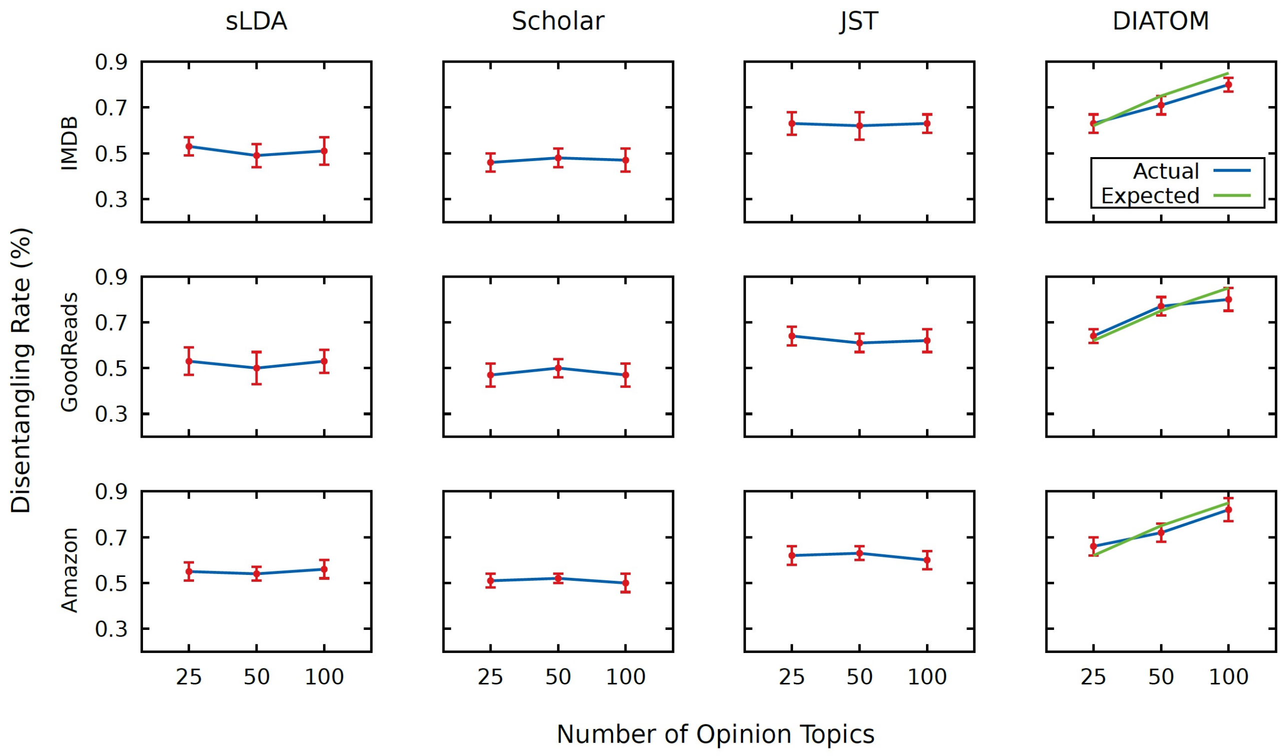
Fig. 3: Examples of topics with the assigned sentences.
Each sentence has a manually assigned label (i.e. Positive, Negative, Plot or None)

Results: **Disentangling rate**

- We analyse how the **proportion of polarity-bearing topics** varies across standard and sentiment topic models.

- We notice that despite the signal from the document labels, sLDA and SCHOLAR tend to produce topics rather **balanced** in terms of neutral and polarity-bearing topics. Joint-Sentiment Topic model has a more skewed distribution towards opinion topics.

- DIATOM instead generates a proportion of opinion topics approaching the expected proportion set up by the model, demonstrating the capability to control the generation of neutral and polarity-bearing topics.



Conclusion

- In our current model, the latent plot topics extracted from reviews are encouraged to have a similar discriminative power as the latent topic learned from plots directly for predicting the plots. It is also possible to impose a **Gaussian prior centred on “z_d”** for the latent plot topics in reviews instead of using the Gaussian prior of zero mean and unit variance.

- Another approach would consist of replacing the plot classifier with a **discriminator** as typically used in GAN training that the learned plot topics from different sources (reviews and plots) are competing

- An additional adversarial mechanism could be employed to differentiate opinion topics based on their overall polarity, to **avoid topics with mixed sentiments**.

Making sense of clinical notes by exploiting local and global context of words

Gabriele Pergola, Yulan He, David Lowe

The 2nd AAAI Workshop on Health Intelligence (AAAI)

2018

Outline

- Making Sense of Text (focusing on clinical documents)
- Topic Modeling and Distributed Representations of Language
- Context-GPU: Combining Local and Global Context
- Conclusions and Future Works

Natural language processing for medical documents

Find me relevant **documents** on this topic!

Find me relevant **facts** about this issue!

Find me **scientific papers** which help treat this patient!

I need more **information** on my health problem

Research Challenges

Is my inflammation your inflammation?

[1]: "Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies", Pisanelli et al. 2004

Research Challenges

- **Compositional Semantic** ‘*spike protein sequence*’ and ‘*white blood cell*’.

Research Challenges

- **COMPOSITIONAL SEMANTIC** ‘*spike protein sequence’* and ‘*white blood cell’.*
- **TECHNICAL JARGON** ‘*basophilia’*, ‘*synechococcus elongatus’.*

Research Challenges

- **Compositional Semantic** ‘*spike protein sequence*’ and ‘*white blood cell*’.
- **Technical jargon** ‘*basophilia*’, ‘*synechococcus elongatus*’.
- **Abbreviations** *S.B.P.* expanded both as:
  - ‘*spontaneous bacterial peritonitis*’
  - ‘*systolic blood pressure*’

Research Challenges

- **COMPOSITIONAL SEMANTIC** — ‘*spike protein sequence*’ and ‘*white blood cell*’.
- **TECHNICAL JARGON** — ‘*basophilia*’, ‘*synechococcus elongatus*’.
- **ABBREVIATIONS** — *S.B.P.* expanded both as:
  - ‘*spontaneous bacterial peritonitis*’
  - ‘*systolic blood pressure*’
- **POLYSEMY** — ‘*Inflammation*’ up to 5 different meanings[1].

[1]: “Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies”, Pisanelli et al. 2004

Research Challenges

- **Compositional Semantic** — '*spike protein sequence*' and '*white blood cell*'.
- **Technical jargon** — '*basophilia*', '*synechococcus elongatus*'.
- **Abbreviations** — *S.B.P.* expanded both as: - '*spontaneous bacterial peritonitis*' - '*systolic blood pressure*'
- **Polysemy** — '*Inflammation*' up to 5 different meanings[1].
- **Lack of Structure**

[1]: "Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies", Pisanelli et al. 2004

Research Challenges

- **Compositional Semantic** — '*spike protein sequence'* and '*white blood cell'.*
- **Technical jargon** — '*basophilia'*, '*synechococcus elongatus'.*
- **Abbreviations** — *S.B.P.* expanded both as: - '*spontaneous bacterial peritonitis'*
  - '*systolic blood pressure'*
- **Polysemy** — '*Inflammation*' up to 5 different meanings[1].
- **Lack of Structure**
- **Data Availability**

[1]: "Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies", Pisanelli et al. 2004

Research Challenges

- COMPOSITIONAL SEMANTIC
- TECHNICAL JARGON
- ABBREVIATIONS
- POLYSEMY
- LACK OF STRUCTURE
- DATA AVAILABILITY

Topical phrases

Topics made of relevant phrases

Contexts

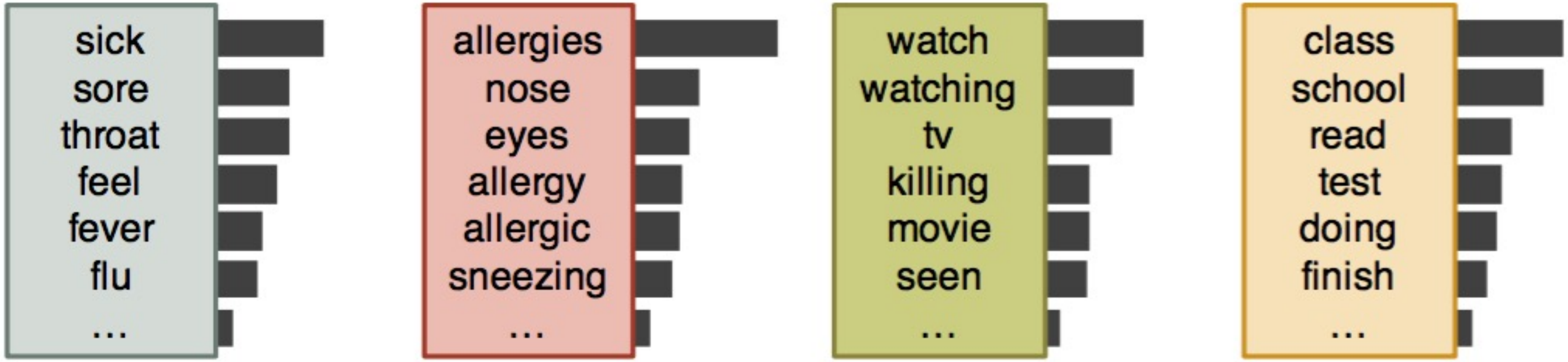
Word local and global context

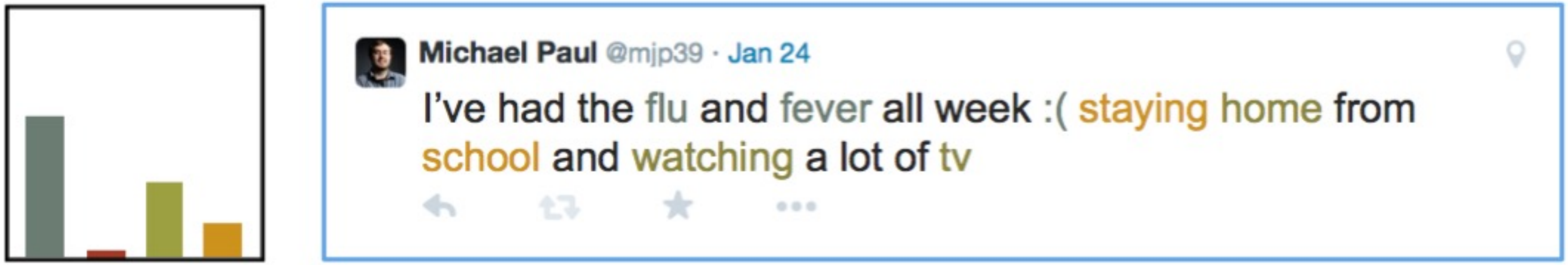
External source of knowledge

Embedding exploited within the inference process

[1]: "Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies", Pisanelli et al. 2004

Topic Modeling - 1





Credits for this slide to "An introduction to Topic models", M. J. Paul 2013

Topic Modeling - 2

Word Embedding

Local context of words

Distributional hypothesis: *“You shall know a word by the company it keeps”*

J. R. Firth, British Linguist, 1957

I love NLP and I like dogs

I = [0 1 0 1 1 0]

Love = [1 0 1 0 0 0]

NLP = [0 1 0 1 0 0]

And = [1 0 1 0 0 0]

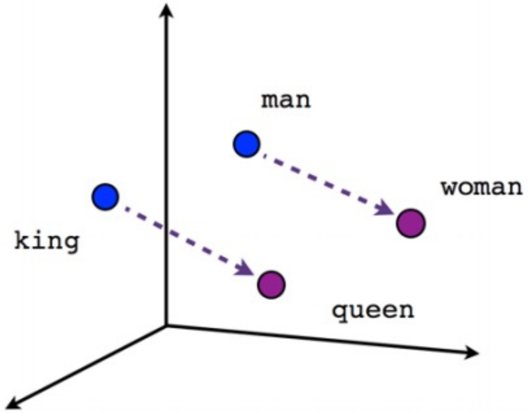
Like = [1 0 0 0 0 1]

Dogs = [0 0 0 0 1 0]

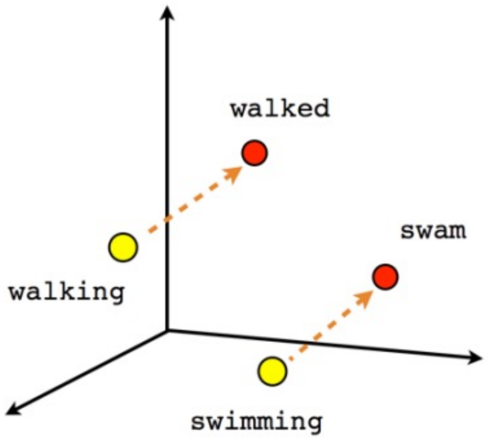
Word Embedding

Distributional hypothesis: *"You shall know a word by the company it keeps"*

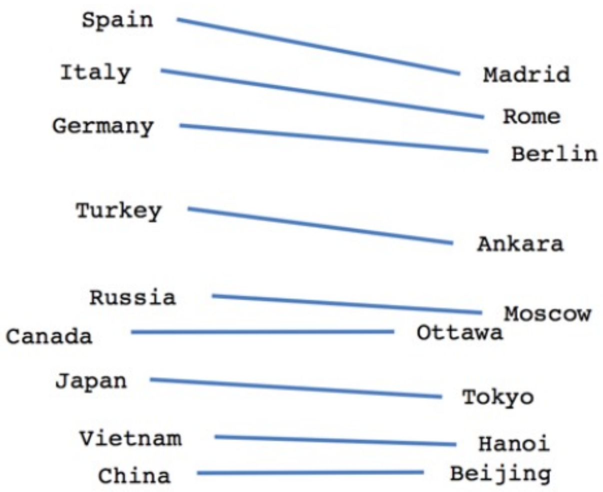
J. R. Firth, British Linguist, 1957



Male-Female



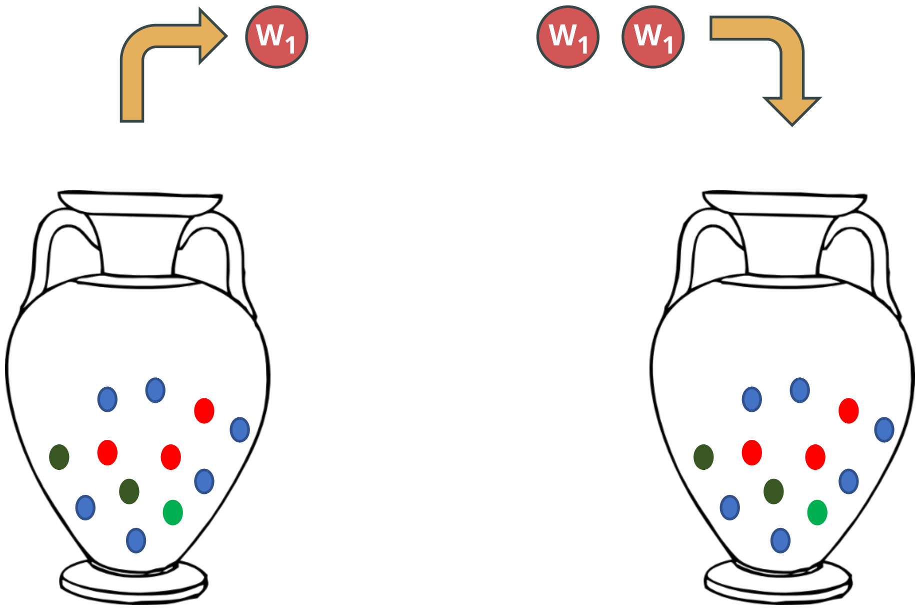
Verb tense



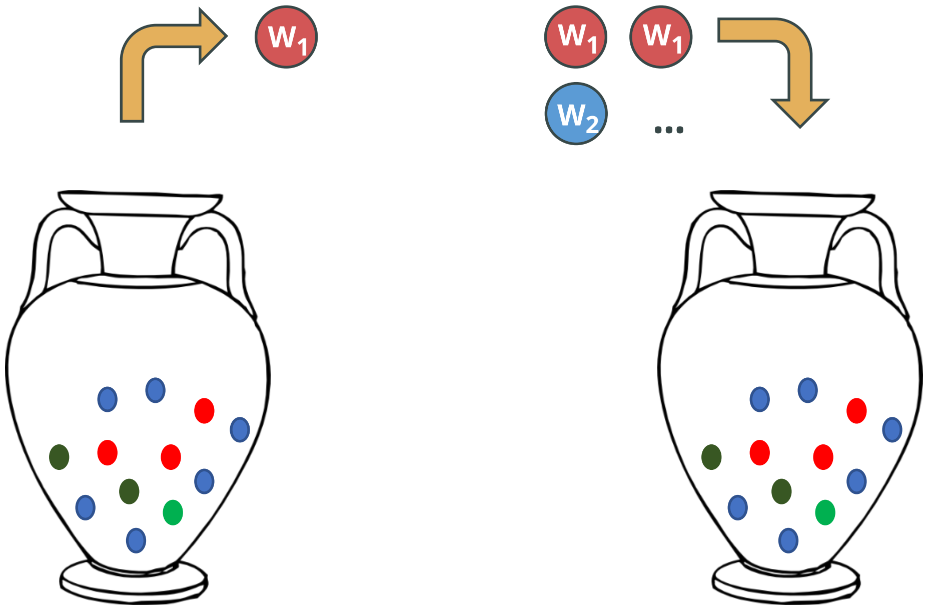
Country-Capital

Context-GPU: Combining local and global context

Simple Polya Urn model



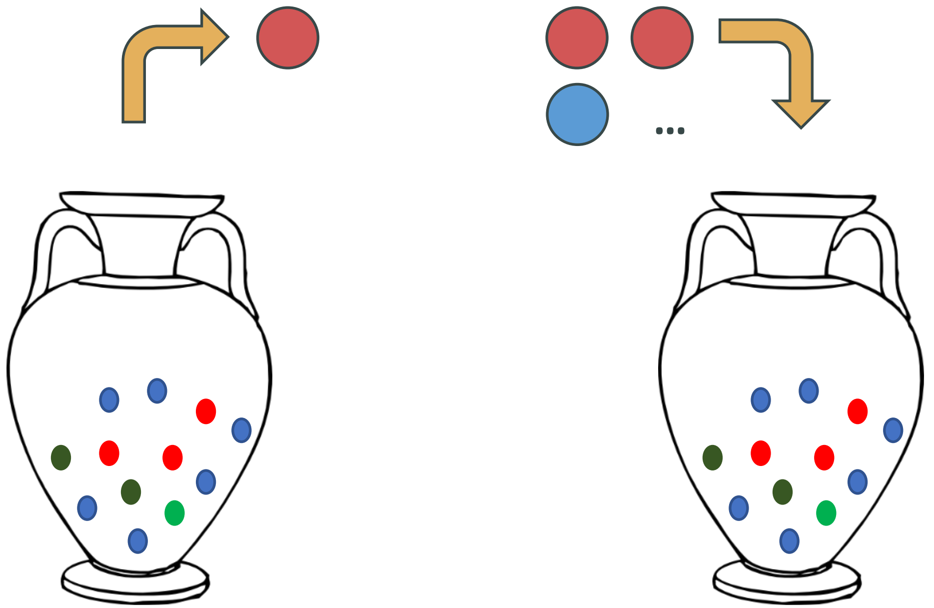
Generalized Polya Urn model

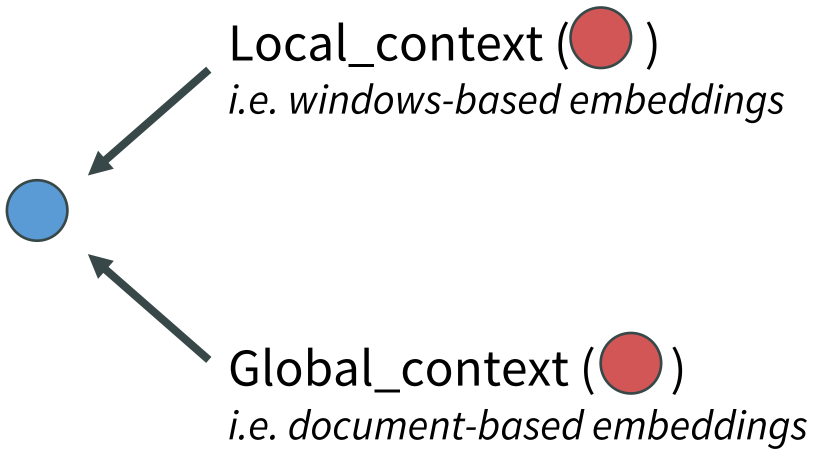


[1]: "Optimizing Semantic Coherence in Topic Models", Mimno et al. 2011

Context-GPU: Combining local and global context

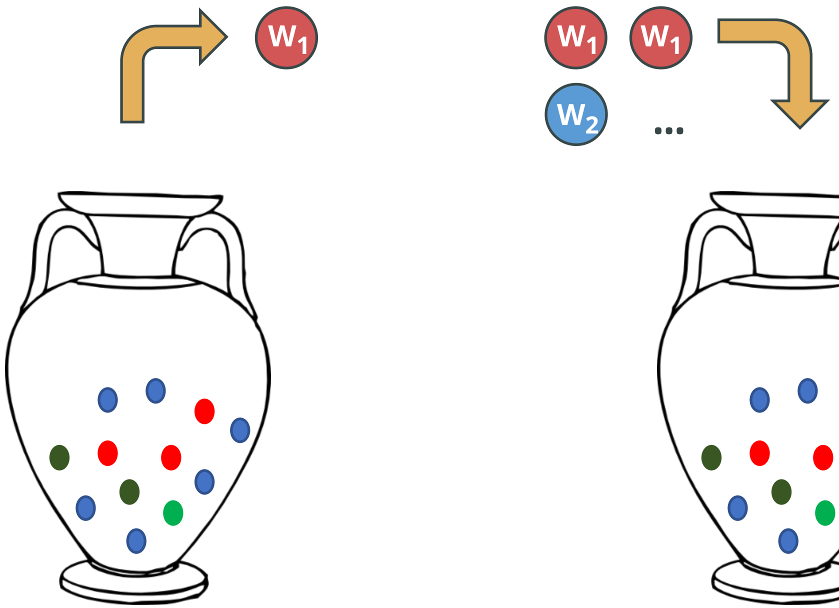
Context - Generalized Polya Urn (GPU) model

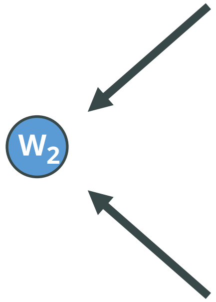




Context-GPU: Combining local and global context

Context - Generalized Polya Urn (GPU) model





Local_context (W_1)
windows-based embedding
(e.g. FastText)

Global_context (W_1)
document-based embedding
(e.g. LSI – Latent Semantic Indexing)

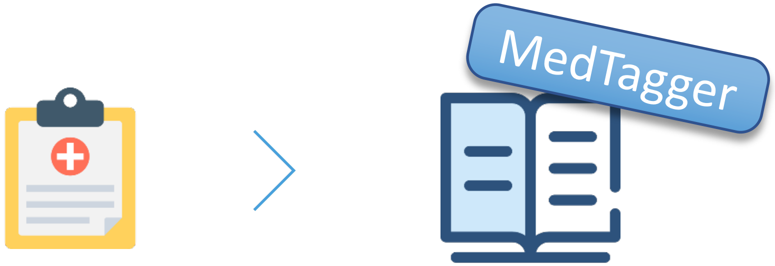
Enhancements

- Improve word semantic
- Add external knowledge
- Simple and efficient process

Context-GPU

[...] white_blood_cell [...]
[...] shortness_of_breath [...]
. . .

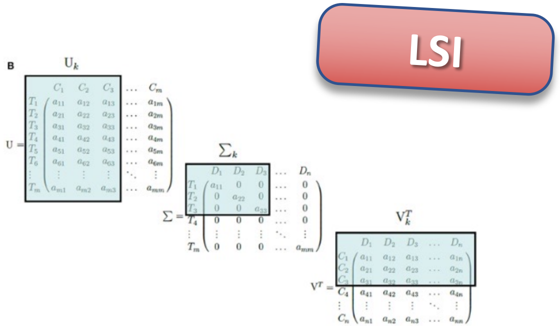
Context-GPU



[...] white_blood_cell [...]
[...] shortness_of_breath [...]
. . .

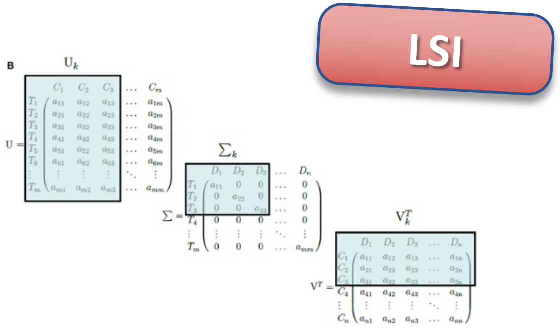
Context-GPU

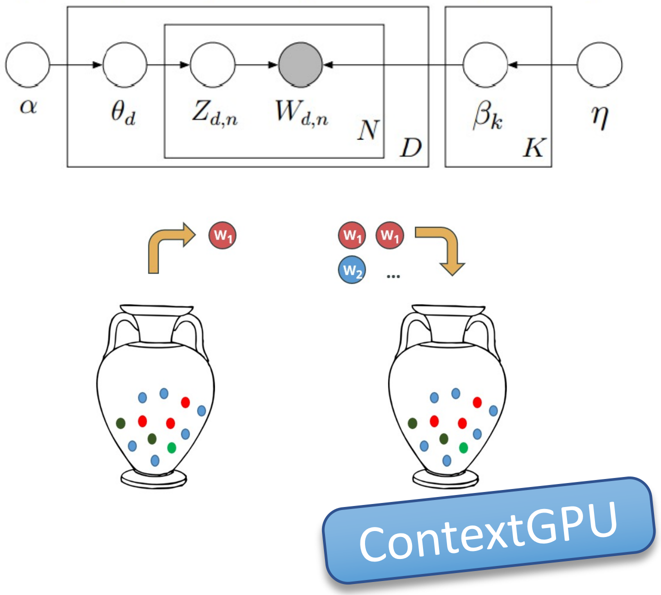
[...] white_blood_cell [...]
[...] shortness_of_breath [...]
. . .



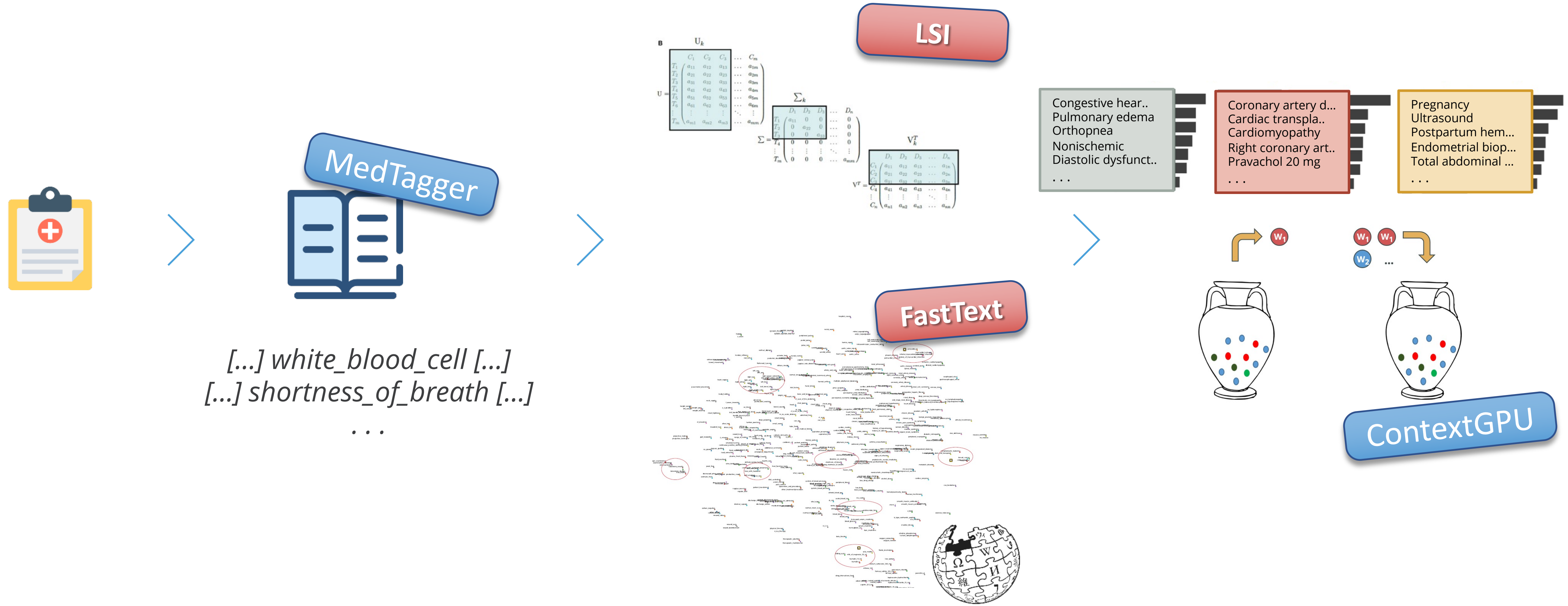
Context-GPU

[...] *white_blood_cell* [...]
[...] *shortness_of_breath* [...]
. . .





Context-GPU



MedTagger
[...] white_blood_cell [...]
[...] shortness_of_breath [...]
. . .
LSI
FastText
Congestive hear..
Pulmonary edema
Orthopnea
Nonischemic
Diastolic dysfunct..
. . .
Coronary artery d...
Cardiac transpla..
Cardiomyopathy
Right coronary art..
Pravachol 20 mg
. . .
Pregnancy
Ultrasound
Postpartum hem...
Endometrial biop...
Total abdominal ...
. . .
ContextGPU

Experimental assessment - Dataset

I2b2 dataset

- 1,243 de-identified discharge summaries
- 7,883 unique terms with “bag of words”
- 9,932 unique terms with “bag of phrases”
- Patient history, discharge, medications and treatments, etc.

“Community annotation experiment for ground truth generation for the i2b2 medication challenge”, *Journal of the American Medical Informatics Association, Uzuner et al. 2010*

Qualitative assessment

	LDA	
chemotherapy	fever	congestive heart failure
dilantin	urinalysis	diuresis
oncology	culture	ejection fraction
xrt	bacteria	approximately
oncologist	white blood cell	felt
cycle	polys	orthopnea
breast cancer	urinary tract infection	digoxin
cancer	band	dyspnea
seizure	fluid	weight
left breast	white	pressure

	Contex-GPU	
pregnancy	coronary artery disease	congestive heart failure
ultrasound	cardiac transplant	pulmonary edema
postpartum hemorrhage	cardiomyopathy	orthopnea
endometrial biopsy	right coronary artery	nonischemic
total abdominal hysterectomy	pravachol 20 mg	diastolic dysfunction
postpartum	paroxysmal atrial fibrillation	cardiomyopathy
vomiting	cyclosporine	heart failure
salpingo oophorectomy	herpes zoster	shortness of breath
physical examination	fenofibrate tricor	cardiac catheterization
fibroid	right coronary artery	atrial fibrillation

Qualitative assessment

	LDA	
chemotherapy	fever	congestive heart failure
dilantin	urinalysis	diuresis
oncology	culture	ejection fraction
xrt	bacteria	approximately
oncologist	white blood cell	felt
cycle	polys	orthopnea
breast cancer	urinary tract infection	digoxin
cancer	band	dyspnea
seizure	fluid	weight
left breast	white	pressure

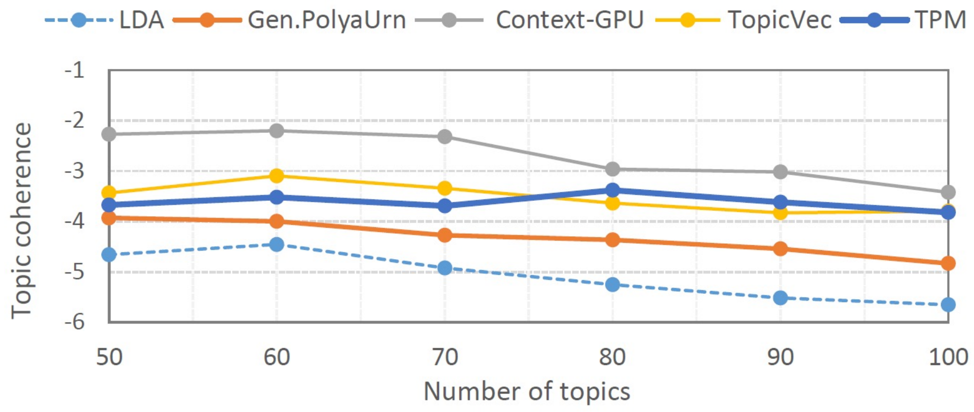
	Contex-GPU	
pregnancy	coronary artery disease	congestive heart failure
ultrasound	cardiac transplant	pulmonary edema
postpartum hemorrhage	cardiomyopathy	orthopnea
endometrial biopsy	right coronary artery	nonischemic
total abdominal hysterectomy	pravachol 20 mg	diastolic dysfunction
postpartum	paroxysmal atrial fibrillation	cardiomyopathy
vomiting	cyclosporine	heart failure
salpingo oophorectomy	herpes zoster	shortness of breath
physical examination	fenofibrate tricor	cardiac catheterization
fibroid	right coronary artery	atrial fibrillation

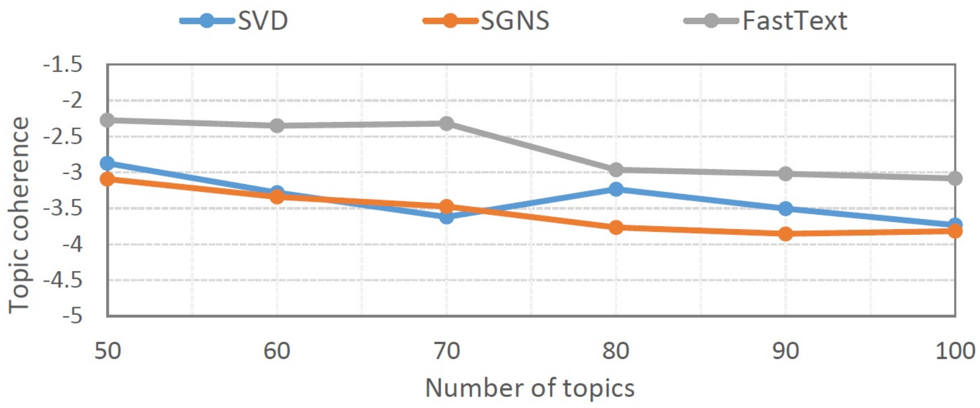
Qualitative assessment

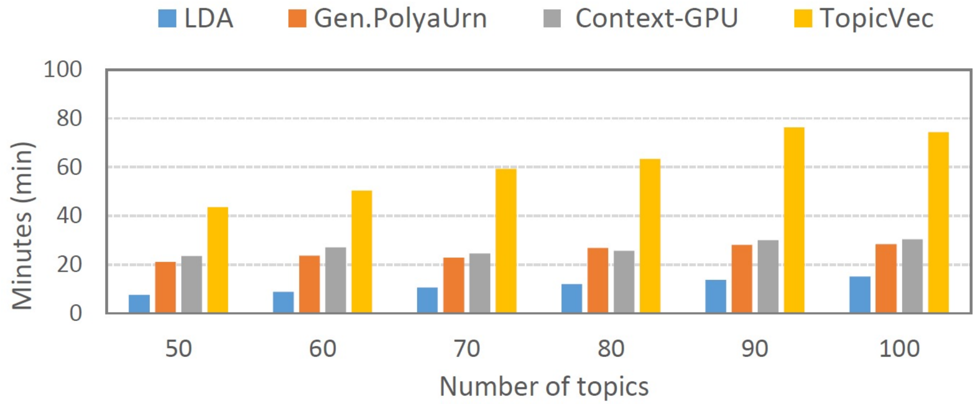
LDA		
chemotherapy	fever	congestive heart failure
dilantin	urinalysis	diuresis
oncology	culture	ejection fraction
xrt	bacteria	approximately
oncologist	white blood cell	felt
cycle	polys	orthopnea
breast cancer	urinary tract infection	digoxin
cancer	band	dyspnea
seizure	fluid	weight
left breast	white	pressure

Contex-GPU		
pregnancy	coronary artery disease	congestive heart failure
ultrasound	cardiac transplant	pulmonary edema
postpartum hemorrhage	cardiomyopathy	orthopnea
endometrial biopsy	right coronary artery	nonischemic
total abdominal hysterectomy	pravachol 20 mg	diastolic dysfunction
postpartum	paroxysmal atrial fibrillation	cardiomyopathy
vomiting	cyclosporine	heart failure
salpingo oophorectomy	herpes zoster	shortness of breath
physical examination	fenofibrate tricor	cardiac catheterization
fibroid	right coronary artery	atrial fibrillation

Context-GPU - Performance







Conclusions

- **Scale up** to bigger dataset (from the i2b2 to the Mimic III dataset) or documents in a new domain (e.g. legal documents)

- Substitute more context-depend representations using **Contextualized Language Models** within the inference process. In addition, **WordPiece tokenizer** combining sub-tokens could better fit the topical phrase encoding.

- **Promotion of medical entities** by means of a biomedical NER (e.g. SciSpacy).

- Not only topic, but **medical coherence**

Boosting Low-Resource Biomedical QA via Entity-Aware Masking Strategies

Gabriele Pergola[†], Elena Kochkina[§], Lin Gui [†], Maria Liakata[*] and Yulan He[†]

[†]University of Warwick
[§] Alan Turing Institute
[*]Queen Mary University Of London

European Chapter of the Association for Computational Linguistics (EACL)

2021

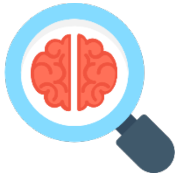
Outline

- Brief Introduction to the QA task
- Why is Biomedical QA hard?
- Transfer Learning
- Biomedical Entity-Aware Masking (BEM) Strategy
- Experimental assessment
- Conclusion and Future Work

Why Biomedical Question Answering?

Find me an answer among **scientific papers** of the current literature

I need more **information** on my health problem



Find me relevant **information** on this issue!

Question-Answering

Questions: Who patented a steam engine in 1781?

Answer: James Watt

QA: Machine Reading Comprehension

Questions: Who patented a steam engine in 1781?

Passage: In 1781 James Watt patented a steam engine that produced continuous rotary motion. Watt's ten-horsepower engines enabled a wide range of manufacturing machinery to be powered. The engines could be sited anywhere that water and coal or wood fuel could be obtained. By 1883, engines that could provide 10,000 hp had become feasible. The stationary steam engine was a key component of the Industrial Revolution, allowing factories to locate where water power was unavailable. The atmospheric engines of Newcomen and Watt were large compared to the amount of power they produced, but high pressure steam engines were light enough to be applied to vehicles such as traction engines and the railway locomotives.

[1]: "SQuAD: 100,000+ questions for machine comprehension of text.", Rajpurkar et al. EMNLP 2016

Biomedical Question-Answering

Questions: What is the incubation period of the virus?

Passage: The mean incubation period was 5.6 days (95% CI: 4.4, 7.4) when excluding Wuhan residents—slightly larger than the estimate without right truncation. The mean estimate for illness onset to hospital admission was 9.7 days (95% CI: 5.4, 17.0) for living cases and 6.6 days (95% CI: 5.2, 8.8) for deceased cases, with the former nearly 2.5 times the length of its untruncated version. Illness onset to death and hospital admission to death were likewise longer than their non-truncated counterparts, at 20.2 days (95% CI: 15.1, 29.5) and 13.0 days (95% CI: 8.7, 20.9), respectively.

[1]: *"Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data." Linton, N.M. et al. 2020, Journal of Clinical Medicine.*

Why is Biomedical QA hard?

- **Compositional Semantic** ‘*spike protein sequence*’ and ‘*white blood cell*’.

Why is Biomedical QA hard?

- **COMPOSITIONAL SEMANTIC** ‘*spike protein sequence’* and ‘*white blood cell’.*
- **TECHNICAL JARGON** ‘*basophilia’*, ‘*synechococcus elongatus’.*

Why is Biomedical QA hard?

- **Compositional Semantic** — ‘*spike protein sequence*’ and ‘*white blood cell*’.
- **Technical jargon** — ‘*basophilia*’, ‘*synechococcus elongatus*’.
- **Abbreviations** — *S.B.P.* expanded both as:
  - ‘*spontaneous bacterial peritonitis*’
  - ‘*systolic blood pressure*’

Why is Biomedical QA hard?

- **Compositional Semantic** — '*spike protein sequence*' and '*white blood cell*'.
- **Technical jargon** — '*basophilia*', '*synechococcus elongatus*'.
- **Abbreviations** — *S.B.P.* expanded both as:
  - '*spontaneous bacterial peritonitis*'
  - '*systolic blood pressure*'
- **Polysemy** — '*Inflammation*' up to 5 different meanings[1].

[1]: "Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies", Pisanelli et al. 2004

Why is Biomedical QA hard?

- **Compositional Semantic** — '*spike protein sequence*' and '*white blood cell*'.
- **Technical jargon** — '*basophilia*', '*synechococcus elongatus*'.
- **Abbreviations** — *S.B.P.* expanded both as: - '*spontaneous bacterial peritonitis*'
  - '*systolic blood pressure*'
- **Polysemy** — '*Inflammation*' up to 5 different meanings[1].
- **Lack of Structure**

[1]: "Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies", Pisanelli et al. 2004

Why is Biomedical QA hard?

- **Compositional Semantic** — ‘*spike protein sequence*’ and ‘*white blood cell*’.
- **Technical jargon** — ‘*basophilia*’, ‘*synechococcus elongatus*’.
- **Abbreviations** — *S.B.P.* expanded both as: - ‘*spontaneous bacterial peritonitis*’
  - ‘*systolic blood pressure*’
- **Polysemy** — ‘*Inflammation*’ up to 5 different meanings[1].
- **Lack of Structure**
- **Data Availability**

CovidQA [1]: 127 pairs
BioASQ [2]: ~ 2000 pairs
SQuAD [3]: ~ 150.000 pairs

[1]: “Coping with Medical Polysemy in the Semantic Web: the Role of Ontologies”, Pisanelli et al. 2004

Transfer Learning: Sequential Adaptation

Pre-Training

Wikipedia
BookCorpus
…

→

Masked
Language Model
(**MLM**)

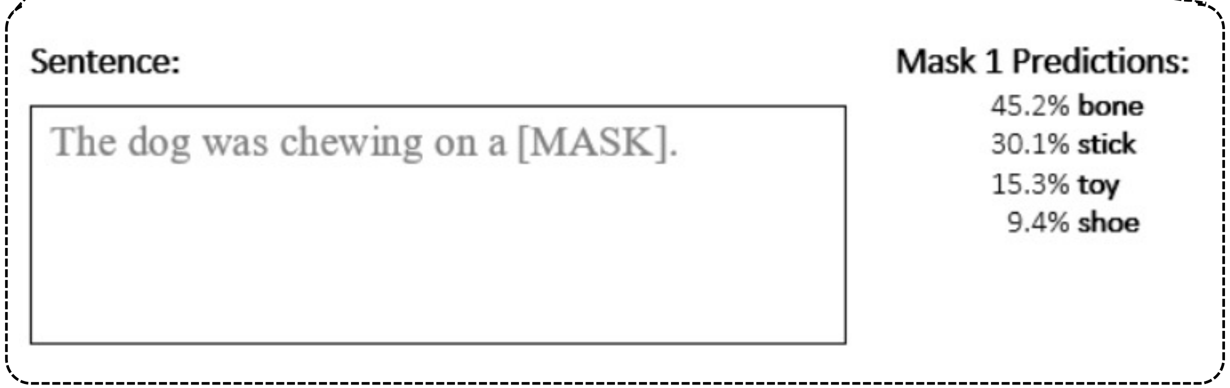
Transfer Learning: **Sequential Adaptation**

Pre-Training

Wikipedia
BookCorpus
...

→

Masked
Language Model
(**MLM**)



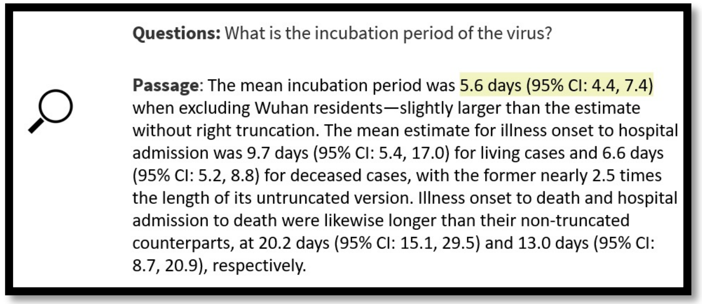
Transfer Learning: **Sequential Adaptation**

Pre-Training

Fine-Tuning

Wikipedia
BookCorpus
...

MLM

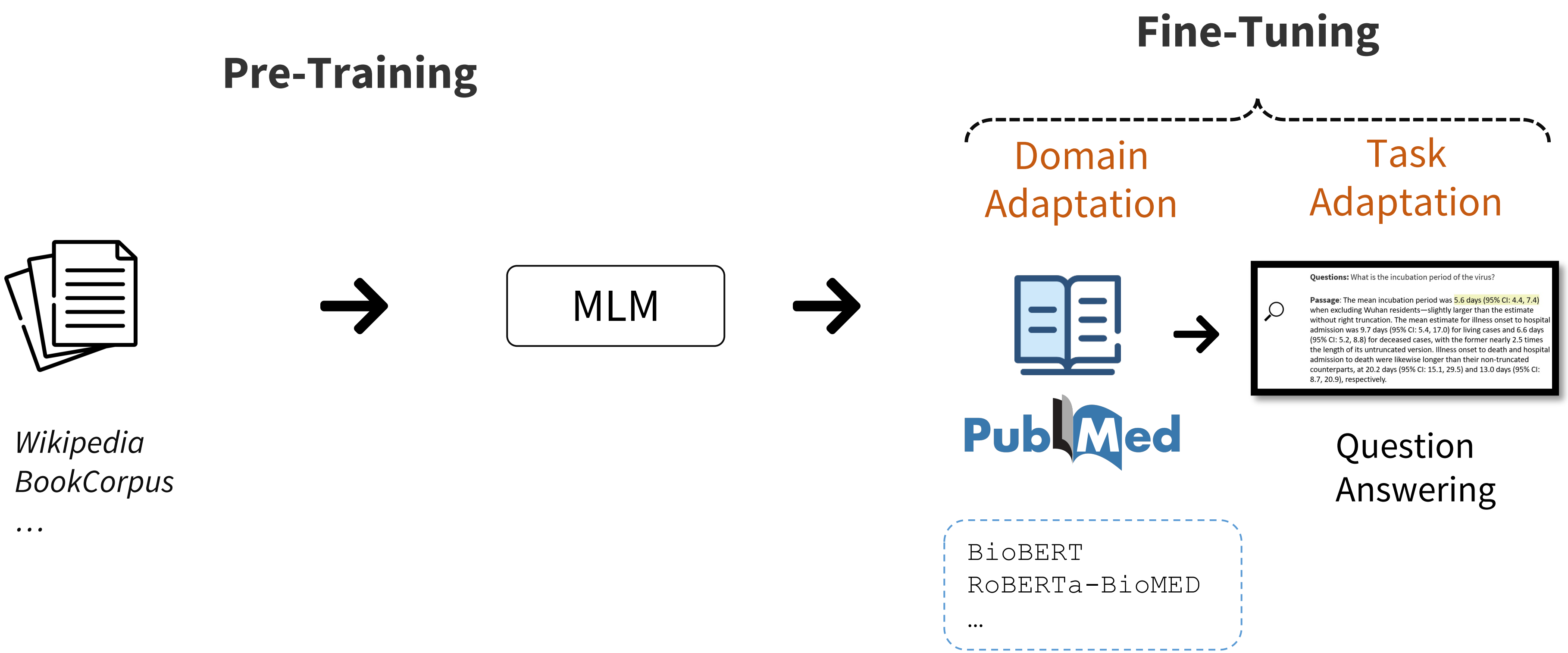


Questions: What is the incubation period of the virus?

Passage: The mean incubation period was 5.6 days (95% CI: 4.4, 7.4) when excluding Wuhan residents—slightly larger than the estimate without right truncation. The mean estimate for illness onset to hospital admission was 9.7 days (95% CI: 5.4, 17.0) for living cases and 6.6 days (95% CI: 5.2, 8.8) for deceased cases, with the former nearly 2.5 times the length of its untruncated version. Illness onset to death and hospital admission to death were likewise longer than their non-truncated counterparts, at 20.2 days (95% CI: 15.1, 29.5) and 13.0 days (95% CI: 8.7, 20.9), respectively.

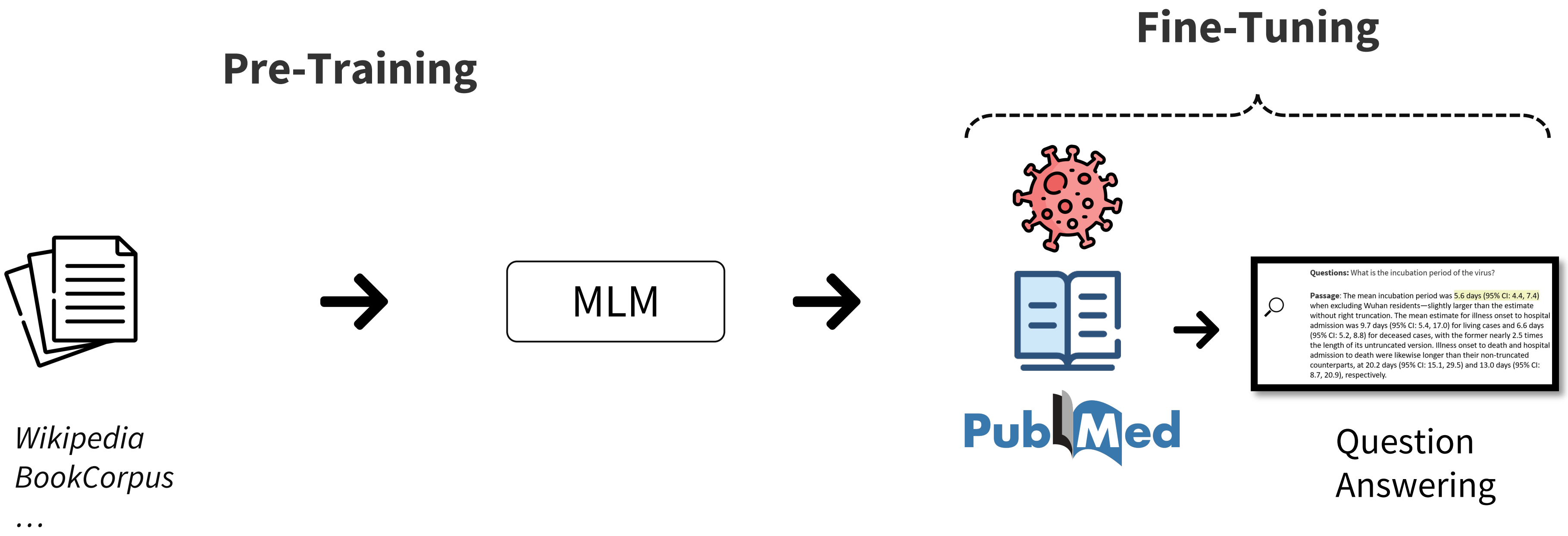
BioASQ
CovidQA
...

Transfer Learning: **Sequential Adaptation**

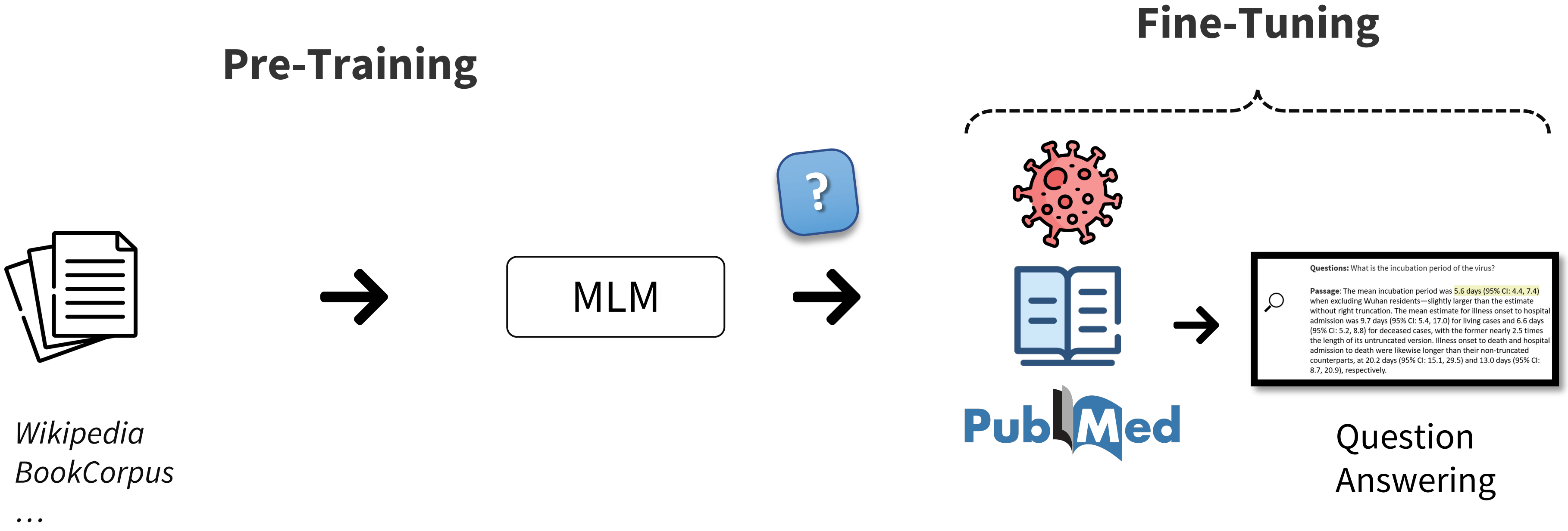


[1]: "BioBERT: a pre-trained biomedical language representation model for biomedical text mining", Lee et al., Bioinformatics 2019
[2]: "Don't stop pretraining: Adapt language models to domains and tasks", Gururangan et al., ACL20

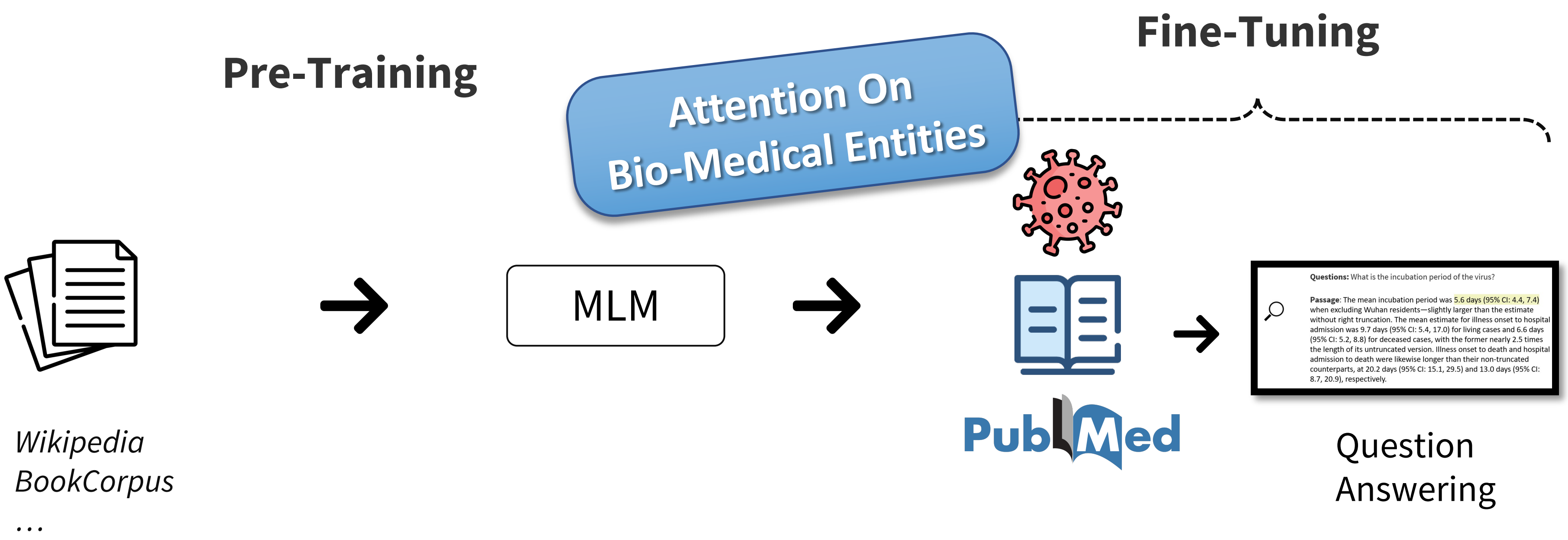
Transfer Learning: **Sequential Adaptation**



Transfer Learning: **Sequential Adaptation**



Transfer Learning: **Sequential Adaptation**



Attention On Medical Entities

Masked Language Models

Masked Language Model
Generates word representations that can be used to predict the missing tokens of an input text.

```
                        pub           glass
                         ↑              ↑
the man went to the [MASK] to buy a [MASK] of wine
```

[1]: *"BERT: Pre-training of deep bidirectional transformers for language understanding", Devlin et al., NAACL19*

Masked Language Models

Masked Language Model

Generates word representations that can be used to predict the missing tokens of an input text.

```
                         pub                     glass
                          ↑                        ↑
the man went to the [MASK] to buy a [MASK] of wine
```

BERT

Pre-trained language model (LM) using multi-layer bidirectional transformer networks to encode contextualised language representations.

BERT's Masking Strategy

15% tokens randomly chosen:

- 80% replaced with `[MASK]`
- 10% randomly swapped
- 10% kept the same

Biomedical Entity-Aware Masking Strategy

BEM masks a proportion of the medical entities in text.
Thus, realigning the word representation to predict the missing medical entities.

```
Patients with diabetes (HR 1.59) were more likely to
reach to the composite endpoints than those without.
```

Biomedical Entity-Aware Masking Strategy

BEM masks a proportion of the medical entities in text.
Thus, realigning the word representation to predict the missing medical entities.

Patients with **diabetes** (**HR** 1.59) were more likely to reach to the **composite endpoints** than those without.

Biomedical Entity-Aware Masking Strategy

BEM masks a proportion of the medical entities in text.
Thus, realigning the word representation to predict the missing medical entities.

Patients diabetes HR

[MASK] with [MASK] ([MASK] 1.59) were more likely to
reach to the [MASK] [MASK] than those without.

composite endpoints

Detect Medical Entities

Sci spaCy

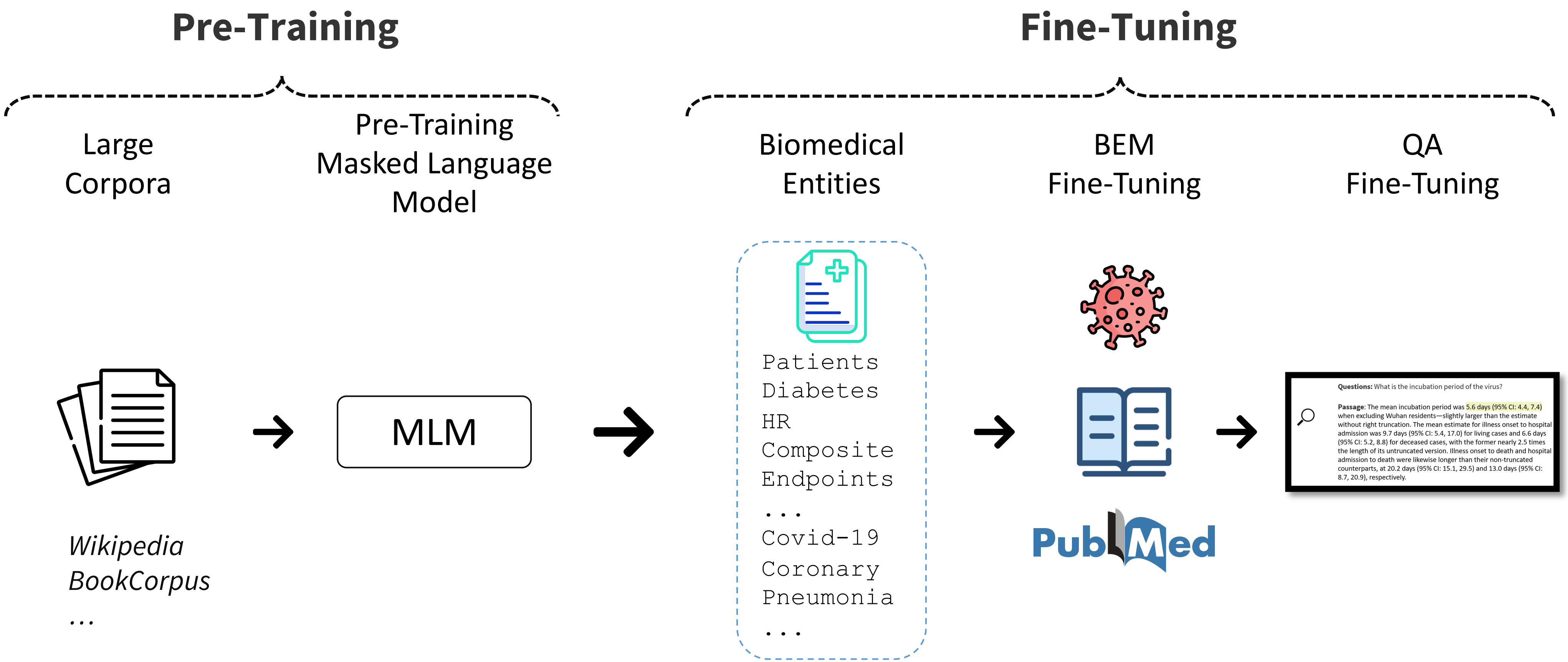
http://scispacy.apps.allenai.org/

Patients with diabetes (HR 1.59) were more likely to reach to the composite endpoints than those without.

↓

Patients ENTITY with diabetes ENTITY (HR ENTITY 1.59) were more likely to reach to the composite ENTITY endpoints ENTITY than those without.

Transfer Learning: **Sequential Adaptation**



Results: CovidQA and BioASQ dataset

- LM fine-tuning on passages provided as context among the PubMed articles referred in the BioASQ [1] and AI2's COVID-19 Open Research [2] datasets.

- **CovidQA** [3] is a manually curated dataset based on the AI2's COVID-19 Open Re-search dataset, and consists of 127 question-answer pairs with 27 questions and 85 unique related articles.

What is the **incubation period** of the virus?
What is the length of viral **shedding** after illness onset?
What is the **incubation period** across different age groups?
What is the proportion of patients who were **asymptomatic**?
What is the **asymptomatic transmission** during incubation?

[1]: "An overview of the BioASQ largescale biomedical semantic indexing and question answering competition", Tsatsaronis et al., BMC Bioinformatics 2015
[2]: "The Covid-19 open research dataset", L. Wang et al. 2020
[3]: "Rapidly bootstrapping a question answering dataset for COVID-19" R. Tang et al. 2020

Results: CovidQA and BioASQ dataset

- LM fine-tuning on passages provided as context among the PubMed articles referred in the BioASQ [1] and AI2's COVID-19 Open Research [2] datasets.

- **CovidQA** [3] is a manually curated dataset based on the AI2's COVID-19 Open Re-search dataset, and consists of 127 question-answer pairs with 27 questions and 85 unique related articles.

```
What is the RR for severe infection in COVID-19 patients with hypertension?
What is the HR for severe infection in COVID-19 patients with hypertension?
What is the OR for severe infection in COVID-19 patients with hypertension?
```

[1]: "An overview of the BioASQ largescale biomedical semantic indexing and question answering competition", Tsatsaronis et al., BMC Bioinformatics 2015
[2]: "The Covid-19 open research dataset", L. Wang et al. 2020
[3]: "Rapidly bootstrapping a question answering dataset for COVID-19" R. Tang et al. 2020

Results: CovidQA and BioASQ dataset

- LM fine-tuning on passages provided as context among the PubMed articles referred in the BioASQ [1] and AI2's COVID-19 Open Research [2] datasets.

- **CovidQA** [3] is a manually curated dataset based on the AI2's COVID-19 Open Re-search dataset, and consists of 127 question-answer pairs with 27 questions and 85 unique related articles.

- **BioASQ** is one of the larger biomedical QA datasets available with over 2000 question-answer pairs. To use it within the extractive questions-answering framework, we convert the questions into the SQuAD dataset format.

[1]: "An overview of the BioASQ largescale biomedical semantic indexing and question answering competition", Tsatsaronis et al., BMC Bioinformatics 2015
[2]: "The Covid-19 open research dataset", L. Wang et al. 2020
[3]: "Rapidly bootstrapping a question answering dataset for COVID-19" R. Tang et al. 2020

Results: CovidQA and BioASQ dataset

- LM fine-tuning on passages provided as context among the PubMed articles referred in the BioASQ [1] and AI2's COVID-19 Open Research [2] datasets.

- **BioASQ** is one of the larger biomedical QA datasets available with over 2000 question-answer pairs. To use it within the extractive questions-answering framework, we convert the questions into the SQuAD dataset format.

What does the **pembrolizumab companion diagnostic test** assess?
What is the **combined effect** of Nfat and miR-25?
Which genomic positions are preferentially selected for **transposon insertion**?
Is **Baloxavir** effective for influenza?

Results: QA Metrics

#	Model	CovidQA			BioASQ 7b		
		P@1	R@3	MRR	SAcc	LAcc	MRR
1	**BERT**	0.081*	0.117*	0.159*	0.012	0.032	0.027
2	+ BioASQ	0.125	0.177	0.206	0.226	0.317	0.262
3	+ STM + BioASQ	0.132	0.195	0.218	0.233	0.325	0.265
4	+ BEM + BioASQ	0.145	0.278	0.269	0.241	0.341	0.288
5	**RoBERTa**	0.068	0.115	0.122	0.023	0.041	0.036
6	+ BioASQ	0.106	0.155	0.178	0.278	0.324	0.294
7	+ STM + BioASQ	0.112	0.167	0.194	0.282	0.333	0.300
8	+ BEM + BioASQ	0.125	0.198	0.236	0.323	0.374	0.325
9	**RoBERTa-Biomed**	0.104	0.163	0.192	0.028	0.044	0.037
10	+ BioASQ	0.128	0.355	0.315	0.415	0.398	0.376
11	+ STM + BioASQ	0.136	0.364	0.321	0.423	0.410	0.397
12	+ BEM + BioASQ	0.143	0.386	0.347	**0.435**	0.443	0.398
13	**BioBERT**	0.097*	0.142*	0.170*	0.031	0.046	0.039
14	+ BioASQ	0.166	0.419	0.348	0.410†	0.474†	0.409†
15	+ STM + BioASQ	0.172	0.432	0.385	0.418	0.482	0.416
16	+ BEM + BioASQ	*0.179*	**0.458**	*0.391*	0.421	**0.497**	**0.434**
17	**T5 LM**						
18	+ MS-MARCO	**0.282***	0.404*	**0.415***	—	—	—

Results: QA Metrics

#	Model	CovidQA			BioASQ 7b		
		P@1	R@3	MRR	SAcc	LAcc	MRR
1	**BERT**	0.081*	0.117*	0.159*	0.012	0.032	0.027
2	+ BioASQ	0.125	0.177	0.206	0.226	0.317	0.262
3	+ STM + BioASQ	0.132	0.195	0.218	0.233	0.325	0.265
4	+ BEM + BioASQ	0.145	0.278	0.269	0.241	0.341	0.288
5	**RoBERTa**	0.068	0.115	0.122	0.023	0.041	0.036
6	+ BioASQ	0.106	0.155	0.178	0.278	0.324	0.294
7	+ STM + BioASQ	0.112	0.167	0.194	0.282	0.333	0.300
8	+ BEM + BioASQ	0.125	0.198	0.236	0.323	0.374	0.325
9	**RoBERTa-Biomed**	0.104	0.163	0.192	0.028	0.044	0.037
10	+ BioASQ	0.128	0.355	0.315	0.415	0.398	0.376
11	+ STM + BioASQ	0.136	0.364	0.321	0.423	0.410	0.397
12	+ BEM + BioASQ	0.143	0.386	0.347	**0.435**	0.443	0.398
13	**BioBERT**	0.097*	0.142*	0.170*	0.031	0.046	0.039
14	+ BioASQ	0.166	0.419	0.348	0.410†	0.474†	0.409†
15	+ STM + BioASQ	0.172	0.432	0.385	0.418	0.482	0.416
16	+ BEM + BioASQ	*0.179*	**0.458**	*0.391*	0.421	**0.497**	**0.434**
17	**T5 LM**						
18	+ MS-MARCO	**0.282***	0.404*	**0.415***	—	—	—

Results: QA Metrics

#	Model	CovidQA			BioASQ 7b		
		P@1	R@3	MRR	SAcc	LAcc	MRR
1	**BERT**	0.081*	0.117*	0.159*	0.012	0.032	0.027
2	+ BioASQ	0.125	0.177	0.206	0.226	0.317	0.262
3	+ STM + BioASQ	0.132	0.195	0.218	0.233	0.325	0.265
4	+ BEM + BioASQ	0.145	0.278	0.269	0.241	0.341	0.288
5	**RoBERTa**	0.068	0.115	0.122	0.023	0.041	0.036
6	+ BioASQ	0.106	0.155	0.178	0.278	0.324	0.294
7	+ STM + BioASQ	0.112	0.167	0.194	0.282	0.333	0.300
8	+ BEM + BioASQ	0.125	0.198	0.236	0.323	0.374	0.325
9	**RoBERTa-Biomed**	0.104	0.163	0.192	0.028	0.044	0.037
10	+ BioASQ	0.128	0.355	0.315	0.415	0.398	0.376
11	+ STM + BioASQ	0.136	0.364	0.321	0.423	0.410	0.397
12	+ BEM + BioASQ	0.143	0.386	0.347	**0.435**	0.443	0.398
13	**BioBERT**	0.097*	0.142*	0.170*	0.031	0.046	0.039
14	+ BioASQ	0.166	0.419	0.348	0.410†	0.474†	0.409†
15	+ STM + BioASQ	0.172	0.432	0.385	0.418	0.482	0.416
16	+ BEM + BioASQ	*0.179*	**0.458**	*0.391*	0.421	**0.497**	**0.434**
17	**T5 LM**						
18	+ MS-MARCO	**0.282***	0.404*	**0.415***	—	—	—

Results: Qualitative Results

BERT with STM	BERT with BEM
*What is the **OR** for severe infection in COVID-19 patients with hypertension?*	
- There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. **-** Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).	**-** There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. **-** Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).
*What is the **HR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	**-** After adjusting for age and smoking status, patients with COPD (HR 2.681), diabetes (HR 1.59), and malignancy (HR 3.50) were more likely to reach to the composite endpoints than those without.
*What is the **RR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	**-** In univariate analyses, factors significantly associated with severe COVID-19 were male sex (14 studies; pooled RR=1.70, ...), hyper-tension (10 studies 2.74 ...),diabetes (11 studies ...), and CVD (..).

Results: Qualitative Results

BERT with STM	BERT with BEM
*What is the **OR** for severe infection in COVID-19 patients with hypertension?*	
- There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. - Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).	- There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. - Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).
*What is the **HR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	- After adjusting for age and smoking status, patients with COPD (HR 2.681), diabetes (HR 1.59), and malignancy (HR 3.50) were more likely to reach to the composite endpoints than those without.
*What is the **RR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	- In univariate analyses, factors significantly associated with severe COVID-19 were male sex (14 studies; pooled RR=1.70, ...), hyper-tension (10 studies 2.74 ...),diabetes (11 studies ...), and CVD (..).

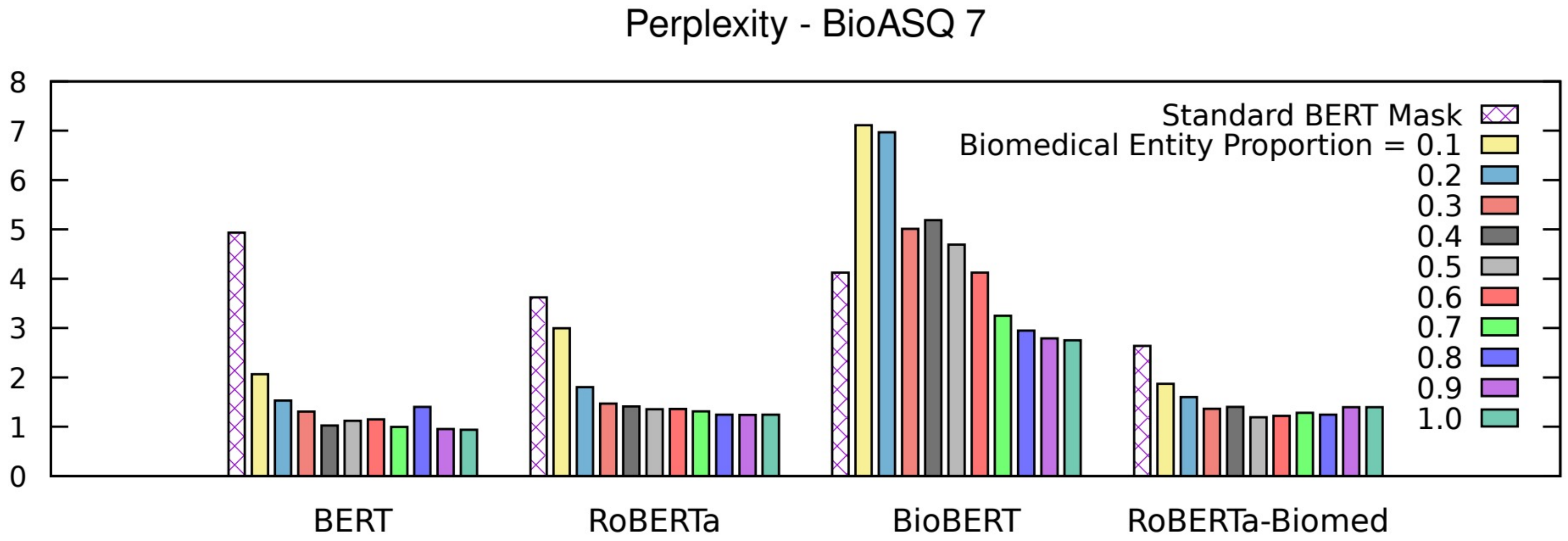
Results: Qualitative Results

BERT with STM	BERT with BEM
*What is the **OR** for severe infection in COVID-19 patients with hypertension?*	
- There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. - Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).	- There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. - Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).
*What is the **HR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	- After adjusting for age and smoking status, patients with COPD (HR 2.681), diabetes (HR 1.59), and malignancy (HR 3.50) were more likely to reach to the composite endpoints than those without.
*What is the **RR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	- In univariate analyses, factors significantly associated with severe COVID-19 were male sex (14 studies; pooled RR=1.70, ...), hyper-tension (10 studies 2.74 ...),diabetes (11 studies ...), and CVD (..).

Results: Qualitative Results

BERT with STM	BERT with BEM
*What is the **OR** for severe infection in COVID-19 patients with hypertension?*	
- There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. - Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).	- There were significant correlations between COVID-19 severity and [..], diabetes [OR=2.67], coronary heart disease [OR=2.85]. - Compared with the non-severe patient, the pooled odds ratio of hypertension, respiratory system disease, cardiovascular disease in severe patients were (OR 2.36, ..), (OR 2.46, ..) and (OR 3.42, ..).
*What is the **HR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	- After adjusting for age and smoking status, patients with COPD (HR 2.681), diabetes (HR 1.59), and malignancy (HR 3.50) were more likely to reach to the composite endpoints than those without.
*What is the **RR** for severe infection in COVID-19 patients with hypertension?*	
- - - -	- In univariate analyses, factors significantly associated with severe COVID-19 were male sex (14 studies; pooled RR=1.70, ...), hypertension (10 studies 2.74 ...),diabetes (11 studies ...), and CVD (..).

Results: Perplexity



Conclusions

Open questions:

- It remains worth investigating what would be the best strategy in choosing the medical entities to mask depending on their **types** (e.g. UMLS categories), and in what proportion doing so.

- Language model pre-training can capture world knowledge by storing it implicitly in the network parameters. More structured knowledge can be integrated to drive the masking process, adopting, for instance, a **knowledge graph** of the medical entities' dependences

- First step towards the integration of topics for a **Topic-Aware Masking Strategy**.

- Assess whether a similar technique would be applicable in other **low-resources domains**, or to drive the model towards **less represented instances** (mitigating the inherent bias).