Recent Developments in Sentiment Analysis

Professor Yulan He

University of Warwick, UK

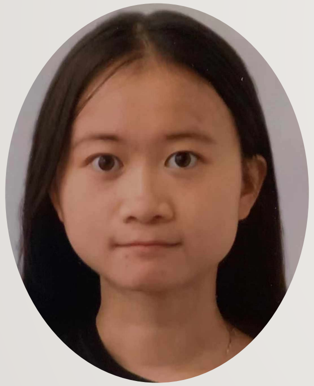
King’s College London, UK (Starting from October 2022)

Lin Gui

Teaching Assistant

Junru Lu

Gabriele Pergola



Runcong Zhao

Lixing Zhu

Hanqi Yan

Acknowledgements

Warwick NLP Group: https://warwick.ac.uk/fac/sci/dcs/research/nlp

Sentiment Classification

★★★★★ **I won't say much about the books...**
Reviewed in the United Kingdom on 15 June 2020
Verified Purchase
I won't say much about the books apart from that they are an incredible read, and some wil
they are some of the best books ever written.

I will
bette
look r

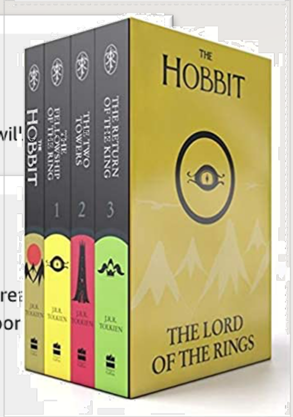
Any L

★★★★☆ **Not 50th anniversary text - Wonky Tolkien**
Reviewed in the United Kingdom on 29 March 2020
Verified Purchase
This leatherette bound Deluxe Pocket Set has stylish classical looks and feels grea
box on my copy has the marks from the shrink wrap that many others have repor
embossing on my copy is also not straight.

The Hobbit text is based on the 1995 Harper Collins edition.

The Lord of the Rings text is however not the 50th anniversary edition. It does not include the introductory text from that edition and it still has some of the mistakes which have since been corrected. For example, Gandalf's inconsistent knowledge of the palantír has been corrected. However, certain dates such as the addition of Westmarch to the Shire and the Mirror of Galadriel reference have not been corrected.

The use of the uncorrected text and the quality issues, unfortunately distract slightly from what is otherwise a really nice set.



Training data

Learn a function to map from input text to a sentiment category

Tolkein's world of Hobbit and Lord of The Rings does not need a new review. So restricting this review to Alan Lee illustrated Boxed Set Hardback edition by Harper Collins (ISBN 978-0008376109).

Slipcase is quite strongly built and is beautifully designed. Hardcovers of the books are thick and strong. Built of each book is quite sturdy and so the books are heavy. There are no illustrations or name on the front hardback (there is gold color embossed name on spine). Font size is acceptable, page quality is good. Alan Lee's illustrations is what makes the collection very attractive. Attached with the review are snaps of the sample illustrations of dark, brown, blue, green and multicolor shades without any photo editing.

★★★★★

Aspect-Based Sentiment Analysis

Booking.com

"Very comfy bed, we stayed in a suite and it was lovely, very big and clean - had microwave, toaster and fridge. Good location right near the center and has restaurants and shops across the road."

A Alix
United Kingdom

"we had a double room, but was to cold when we complaint about the heating not been working they move us to a suite, it was very nice and confortable."

J Jose
United Kingdom

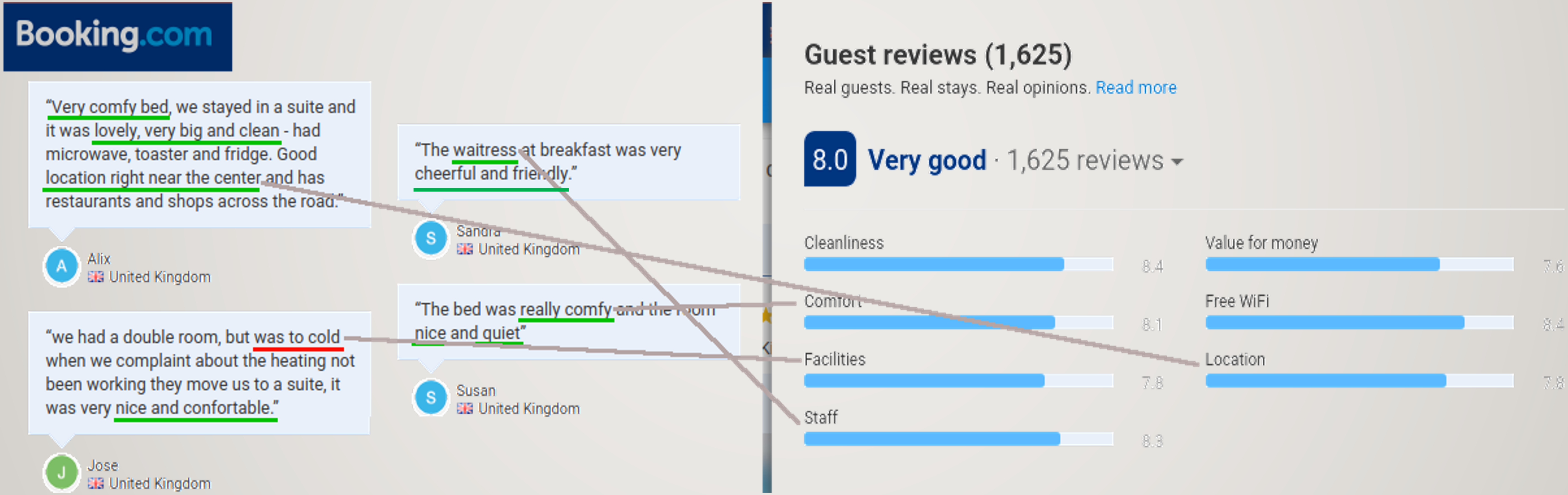
"The waitress at breakfast was very cheerful and friendly."

S Sandra
United Kingdom

"The bed was really comfy and the room nice and quiet"

S Susan
United Kingdom

Aspect-Based Sentiment Analysis



Sentiment-Topic Extraction

Make-up

Skin Care

Shaving & Hair Removal

Johnboy

Not impressed

john bristow

marvellous

A. Shopper

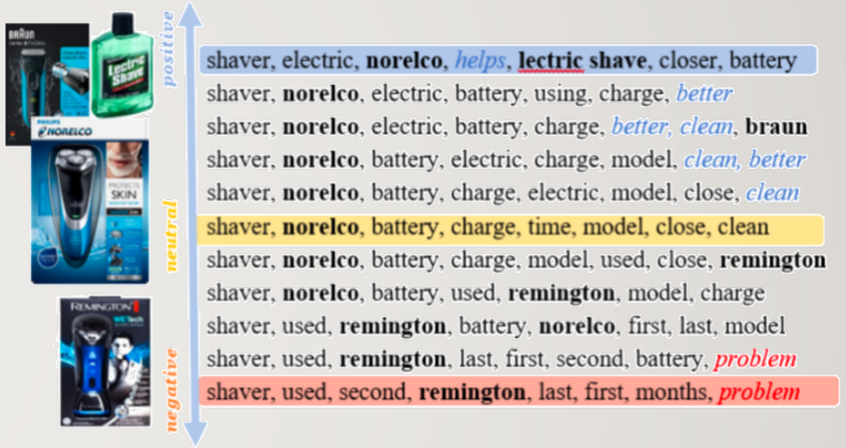
Reasonably effective lady shaver

Reviewed in the United Kingdom on 29 October 2018

Package Quantity: 1 | Verified Purchase

I love all my Remington hair tools, but this shaver is not the best. It is beautifully ergonomically designed and nice looking, but my problem is more with the double foils and the cutting edge. Shaving with this hurt my underarms (and I fully accept I may have used it incorrectly) and after one use, I decided to go back to the Philips lady shave, which I bought this to replace. So back to the Philips which I bought a week after this. By contrast with the Philips I discovered that the cutting edge is specifically designed with microscopic rounded tips, so it does not graze the skin and I think this makes all the difference.

Topics transition with varying polarity strength



Brand Score

Emotion Cause Extraction

News article

1. The crime that ten people were killed shocked the whole country.
2. This was due to personal grievances.
3. Qiu had arguments with the management staff,
4. and thought the Taoist temple host had molested his wife.
5. He became **angry**,
6. and killed the host and destroyed the temple.

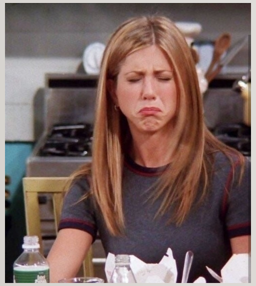
Angry

emotion cause

Dialogue Emotion Detection

Maybe this weekend we could go to the beach?

I can't go, I have to work!



That's too bad.

Sentiment-Aware Natural Language Generation

Response:

Joy

Tea is my favourite.

Query:

Why do you avoid Starbucks?

Fear

I am not sure. It is just a weird feeling.

Anger

I hate the taste of coffee.

Review Question-Answering

Top reviews from the United States

Marsha S. Auster

★☆☆☆☆ **leaky batteries**

Reviewed in the United States on December 22, 2017

Size: 36 Count (Pack of 1) | Style: AAA | **Verified Purchase**

Not only did a few batteries leak they ruined one of my Gideon Flameless LED candles. I am so disappointed. I thought that with the Amazon name I would have a quality product. I purchase many products from the Amazon website.
I would like a refund and a replacement for the ruined Gideion Flameless LED candle. I purchased three candles and the 6 inch candle is ruined.

858 people found this helpful

Helpful | Comment | Report abuse

B. Bourque

★☆☆☆☆ **These batteries leak and will destroy electronic devices.**

Reviewed in the United States on June 24, 2018

Size: 36 Count (Pack of 1) | Style: AAA | **Verified Purchase**

I usually love Amazon products, but hate these batteries. A pack of 36 AAA were purchased in 2015 and used in a number of items, ranging from clocks, multi-meters, IR temperature gauges, etc. Every single one of these devices that I have placed the Amazon batteries in, and left them for a couple years, has leaked causing serious damage to the product. Duracell and Energizer batteries have never leaked on me. Spend more money and get a quality battery that will not leak. Amazon should not be selling such a poor product that has caused many times the batteries original purchase price to my electronic devices.

555 people found this helpful

Helpful | Comment | Report abuse

D. Ross

★☆☆☆☆ **Leaked and ruined a vehicle alarm remote control!**

Reviewed in the United States on October 27, 2018

Size: 36 Count (Pack of 1) | Style: AAA | **Verified Purchase**

Unfortunately these batteries are not a quality product. I purchased a carton of them with expiration dates of 2027 yet they go dead quickly and LEAK! The only reason I tried this brand is due to many Duracell

Customer questions & answers

Have a question? Search for answers

92 votes

Question: How do they compare to Duracell?

Answer: The Amazon brand of alkaline battery meets expectations. I have used at least 50 i over the past 2-3 years. None have leaked and only one has been DOA. Several Dur leaked on me and a couple of Dollar General brand alkalines have as well. The Ama are either cheaper or competi… see more

By Amazon Customer on December 12, 2018

See more answers (29)

44 votes

Question: How long do these last on the bose QC 25s and do you think they would easily last go through 2-3 rechargeable aa battery chargesa week?

Answer: Bose qc uses AAA size not AA. Bose says alkaline battery last about 35 hours. 48 ba provide about 1600 hours of use but likely less if headphones are left on when not

By Sui Lin Chee on June 18, 2016

See more answers (3)

39 votes

Question: what is the guarantee date on these batteries

Answer: My AAs bought in about July 2017 are stamped 12-2026. The AAs I bought in late stamped 8-2025. The 9Vs I bought in early 2017 are stamped 8-2021. The AAAs I b to mid 2017 are stamped 6-2027. The D cells I bought in December 2017 are stam All are shrink wrapped so they can't… see more

By Amazon Customer on February 4, 2018

See more answers (9)

15 votes

Question: I am less interested in which battery last 5 minutes longer and more interested in w leak the least. which leak least?

Answer: Decent batteries should not leak even IF you put something away. I've had things w items left for 4-5-6 years. Opened, batteries dead...no leakage!!!! I have used Durac Walmart, IKEA, CVS brand batteries. In 40 years NEVER had a battery leak!!! Won't thst people claim they … see more

By Amazon Customer on October 13, 2020

See more answers (14)

Opinion Summarisation

Booking.com

"Very comfy bed, we stayed in a suite and it was lovely, very big and clean - had microwave, toaster and fridge. Good location right near the center,and has restaurants and shops across the road."

Alix
United Kingdom

"we had a double room, but was to cold when we complaint about the heating not been working they move us to a suite, it was very nice and confortable."

Jose
United Kingdom

"The waitress at breakfast was very cheerful and friendly."

Sandra
United Kingdom

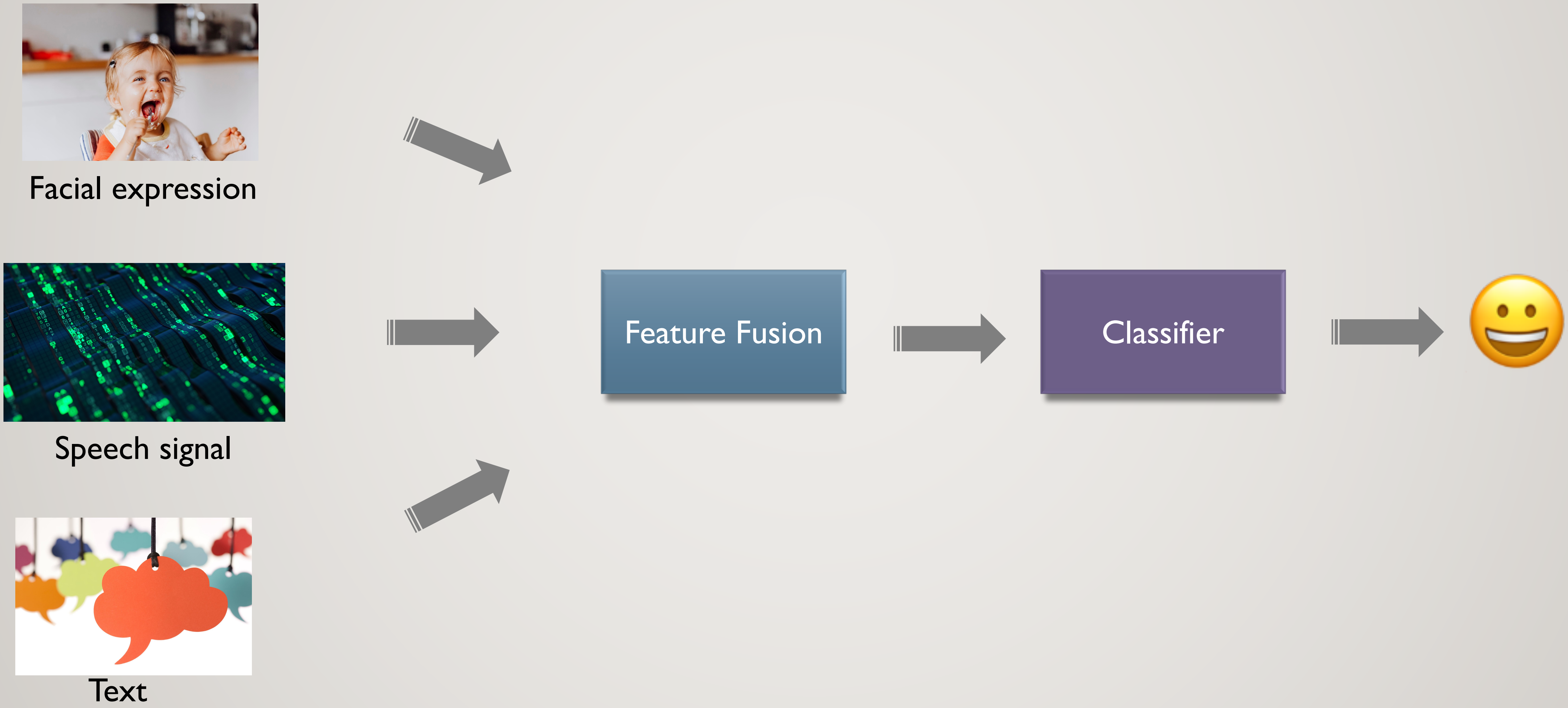
"The bed was really comfy and the room nice and quiet"

Susan
United Kingdom

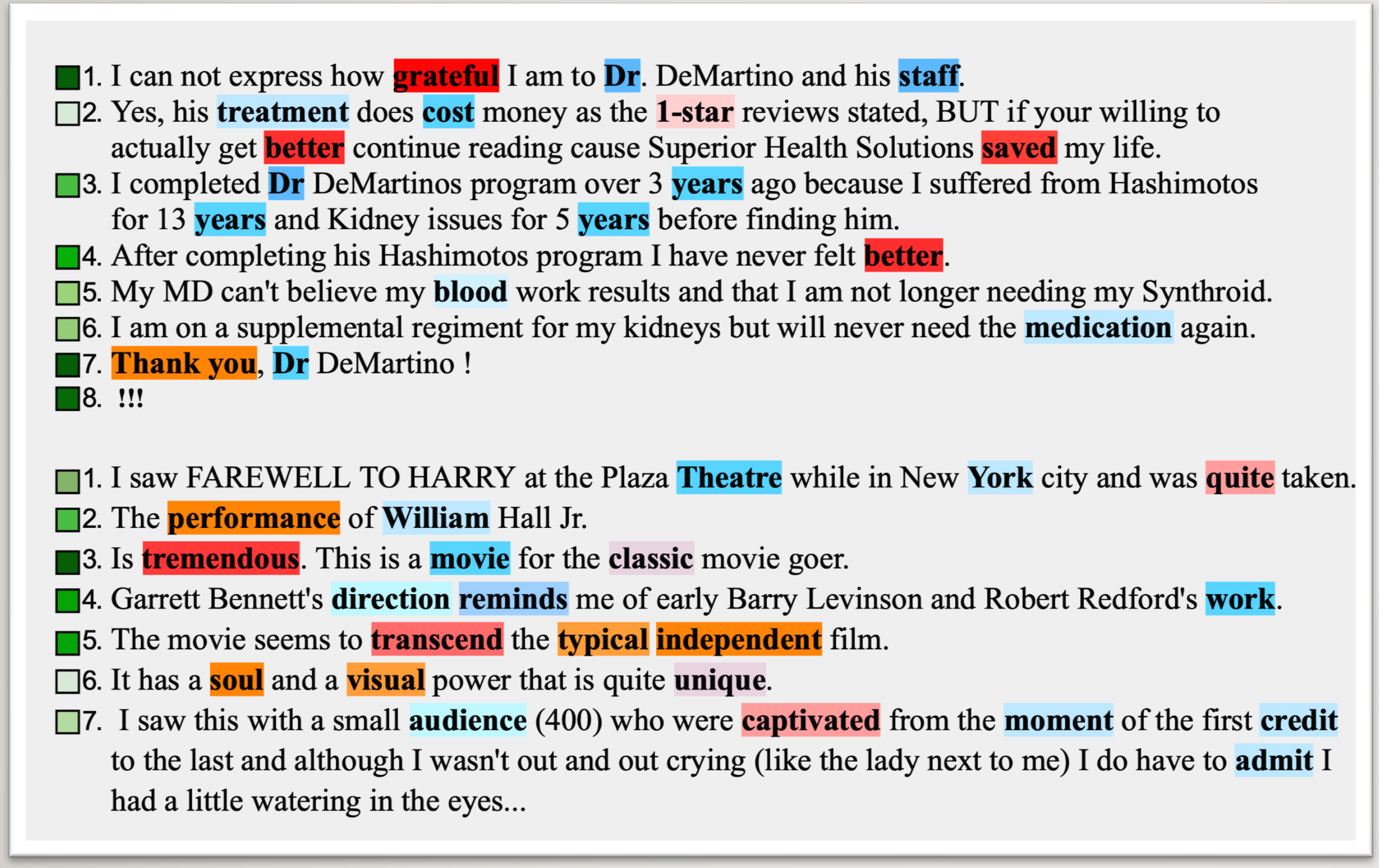
Summary:

The hotel is located near to the city centre. The room was nice with a comfy bed. The waitress was friendly.

Multimodal Sentiment Analysis



Interpretability in Sentiment Analysis



- Visualisation of attention weights of two example reviews.
- Words coloured in the red family are those with high sentiment-associated weights, while words coloured in the blue family are those with high aspect-associated weights.
- Each sentence is also preceded with a green-shaded box with varying intensities indicating different *sentence-level attention weights.*

Outline

- Aspect-Based Sentiment Analysis
- Sentiment-Topic Extraction
- Emotion Cause Detection
- **Break**
- Dialogue Emotion Detection
- Review Question-Answering
- Interpretability in Sentiment Analysis

Outline

- Aspect-Based Sentiment Analysis
- Sentiment-Topic Extraction
- Emotion Cause Detection
- Dialogue Emotion Detection
- Review Question-Answering
- Interpretability in Sentiment Analysis

Problem Definition

- **Task:** Given a sentence (and a pre-defined aspect list), aspect-based sentiment analysis (ABSA) aims at inferring the sentiment polarity of an aspect expressed in a sentence.
- **Related task**: Aspect Sentiment Triplet Extraction (ASTE)
  - text → (aspect/target, polarity, sentiment expression)

Waiters are friendly and the tuna sushi seemed pretty fresh.

↓

Aspect	Aspect Term	Sentiment Expression	Polarity
SERVICE	*waiters*	*friendly*	Positive
FOOD	*tuna sushi*	*pretty fresh*	Positive

Problem Definition

- **Task:** Given a sentence (and a pre-defined aspect list), aspect-based sentiment analysis (ABSA) aims at inferring the sentiment polarity of an aspect expressed in a sentence.
- **Related task**: Aspect Sentiment Triplet Extraction (ASTE)
  - text → (aspect/target, polarity, sentiment expression)
- **Related task**: Aspect Category Sentiment Analysis (ACSA)
  - text → (aspect term, aspect category, polarity)

The sushi seemed pretty fresh and was proportioned.

Aspect Term	Aspect Category	Polarity
sushi	FOOD#QUALITY	Positive
sushi	FOOD#STYLE_OPTION	Positive

Tagging Scheme

- B – begin, I – inside, E – end, S – single, O – outside

Waiters are friendly and the tuna sushi seemed pretty fresh.

Aspect tag:	S	O	O	O	O	B	E	O	O	O
Polarity tag:	POS	O	O	O	O	POS	POS	O	O	O
Aspect+polarity tag:	S+POS	O	O	O	O	B+POS	E+POS	O	O	O
Opinion expression tag:	O	O	S	O	O	O	O	O	B	E

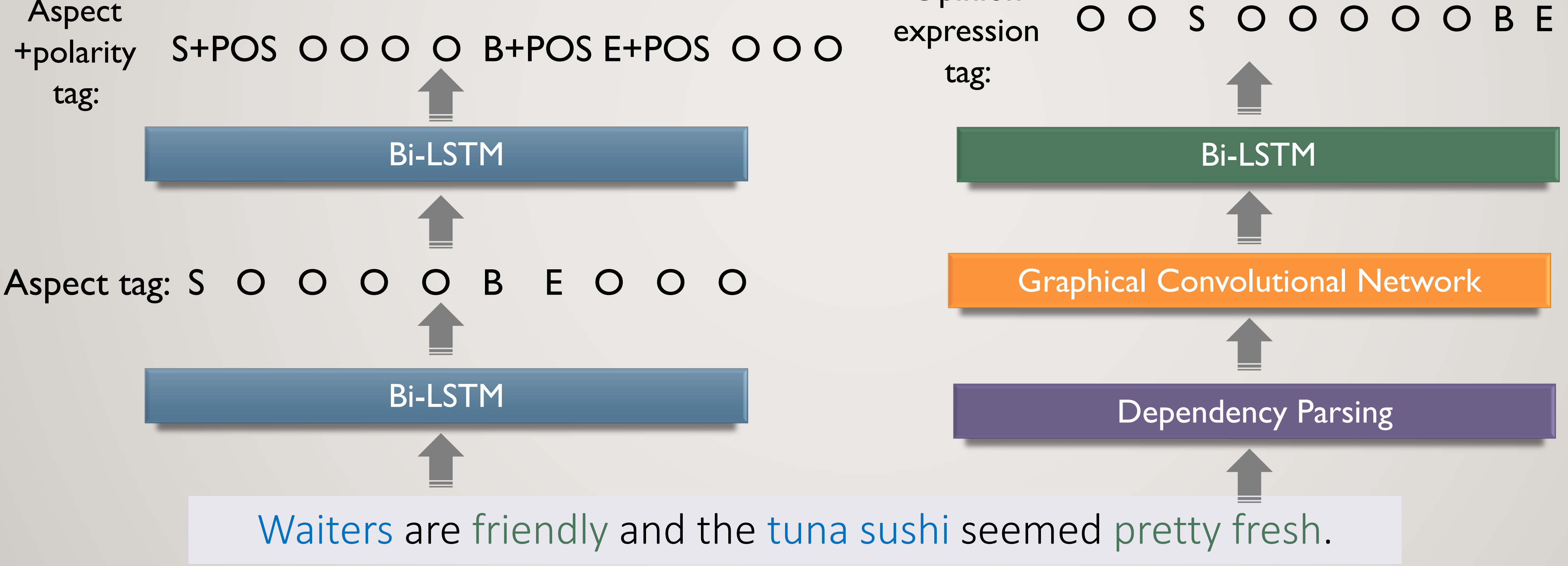
Two-Stage Framework

Waiters are friendly and the sushi seemed pretty fresh.

- Stage 1: given a sentence, extract two label sequence –
  - aspect terms and their polarities – ("waiters", positive); ("tuna sushi", positive)
  - opinion terms – "*friendly*"; "*pretty fresh*"
- Stage 2: pair up aspect terms with their corresponding opinion expressions
  - ("waiters", positive, "*friendly*")
  - ("tuna sushi", positive, "*pretty fresh*")

Two-Stage Framework: Stage 1

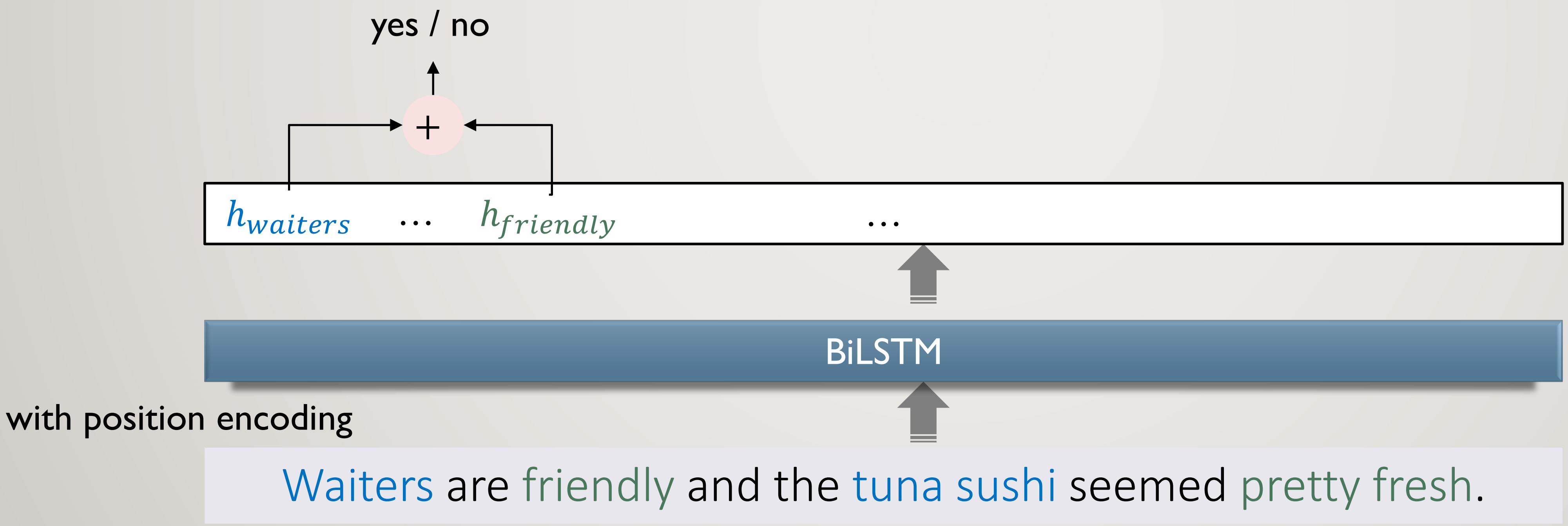
- Stage 1: given a sentence, extract two label sequence –
  - aspect terms and their polarities – (“waiters”, positive); (“tuna sushi”, positive)
  - opinion terms – ”*friendly*”; “*pretty fresh*”



Peng et al. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis, AAAI 2020.

Two-Stage Framework: Stage 2

- Stage 2: pair up aspect terms with their corresponding opinion expressions
  - candidate aspect terms and their polarities – (“waiters”, positive); (“tuna sushi”, positive)
  - candidate opinion terms – ”*friendly*”; “*pretty fresh*”
  - possible aspect-opinion pairs: (“waiters”, positive, ”*friendly*”) (“waiters”, positive, ” *pretty fresh*”) (“tuna sushi”, positive, ”*friendly*”) (“tuna sushi”, positive, “*pretty fresh*”)



Peng et al. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis, AAAI 2020.

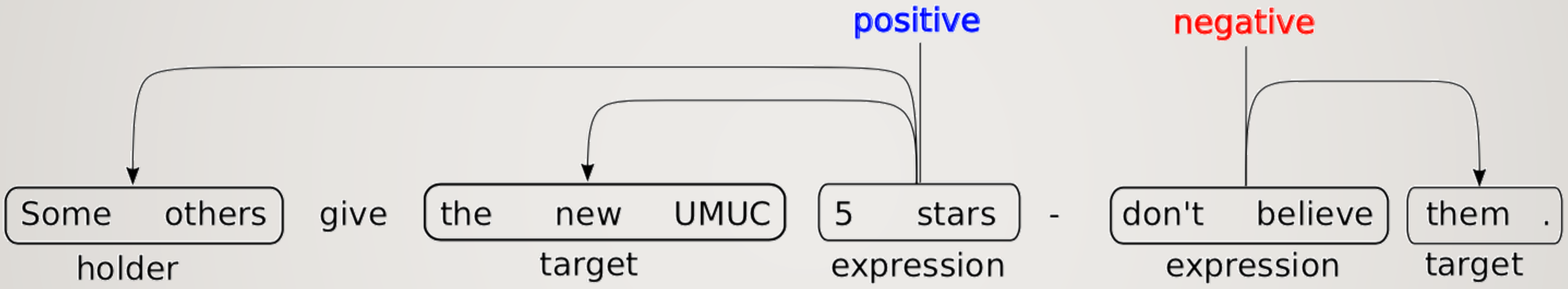
Example Output

Review text	Ground Truth	Model Output
Rice is too dry, tuna wasn't so fresh either.	(Rice, too dry, NEG) (tuna, wasn't so fresh, NEG)	(Rice, too dry, NEG) (tuna, wasn't so fresh, NEG) (Rice, wasn't so fresh, NEG) ✗ (tuna, too dry, NEG) ✗
I am pleased with the fast log on, speedy WiFi connection and the long battery life.	(log on, pleased, POS) (log on, fast, POS) (WiFi connection, speedy, POS) (battery life, long, POS)	(log, pleased, POS) ✗ (log, fast, POS) ✗ (WiFi connection, speedy, POS) (battery life, long, POS)
The service was exceptional – sometime there was a feeling that we were served by the army of friendly waiters.	(service, exceptional, POS), (waiters, friendly, POS)	(service, exceptional, POS) (waiters, friendly, POS)

Peng et al. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis, AAAI 2020.

Structured Sentiment Analysis

- **Problem definition**: to predict a holder, target, sentiment expression, their relationships and a polarity attribute.



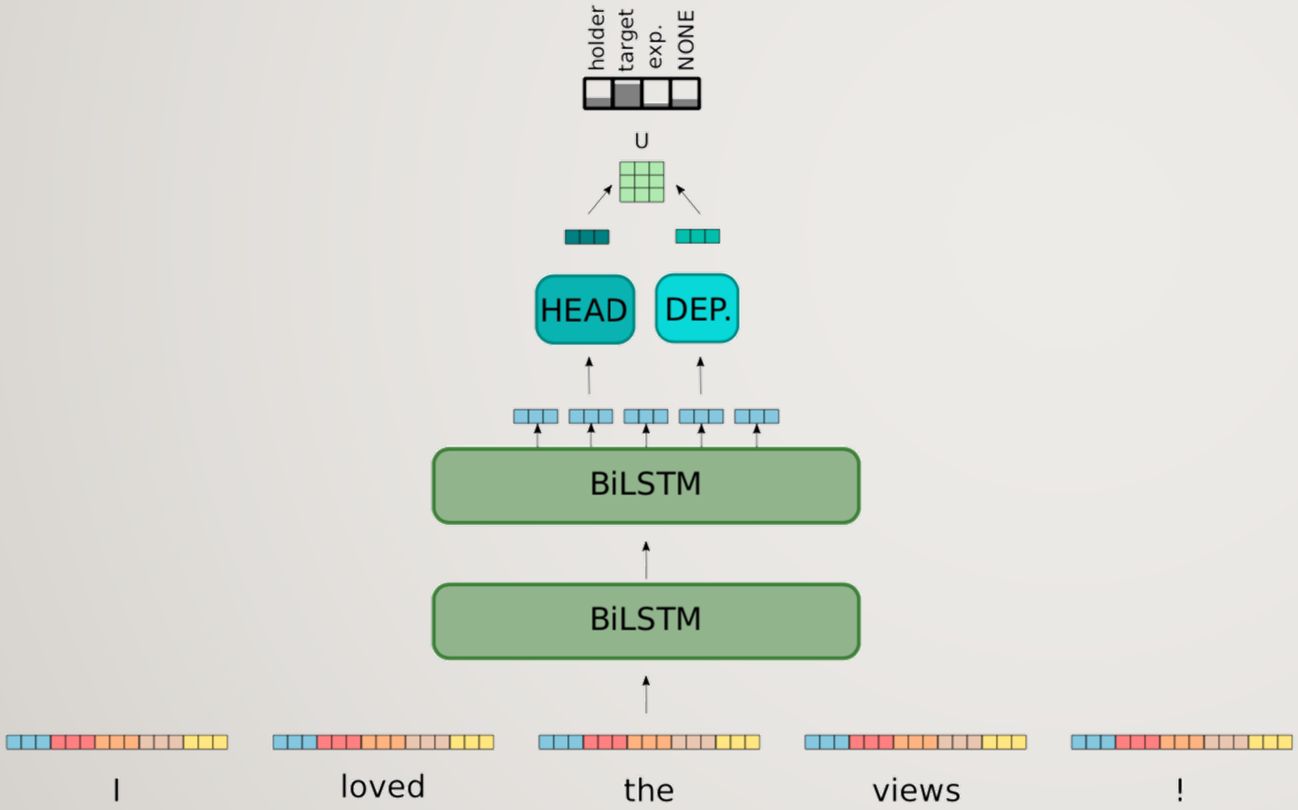
Holder: *some others*
Target: *the new UMUC*
Senti. expr.: *5 stars*
Polarity: *positive*

Holder: *null*
Target: *them*
Senti. expr.: *don't believe*
Polarity: *negative*

Barnes et al., [Structured Sentiment Analysis as Dependency Graph Parsing](), ACL 2021.

Structured Sentiment Analysis

- Convert the sentiment graph into the parsing graph
- Adapt the neural dependency parser by (Dozat and Manning, 2018).



- An input sentence is encoded by BiLSTM.
- The contextualised embeddings are fed into two NNs to create representations for potential *heads* and *dependents*.
- A bilinear transformation using the tensor U is used to score possible arc labels.

Challenges

ABSA heavily relies on annotated training data.

Some aspects may have limited annotated data.

Can we leverage aspect-invariant sentiment expressions to automatically generate synthetic training data?

Aspect-Dependent vs. Aspect-Invariant Sentiment Expressions

- Aspect-dependent sentiment expressions

The ***food quality*** is high.

FOOD#QUALITY [positive]

The ***food price*** is high.

FOOD#PRICE [negative]

- Aspect-invariant sentiment expressions

The ***service*** was good!

SERVICE#GENERAL [positive]

The ***food quality*** was good!

FOOD#QUALITY [positive]

Creating Synthetic Training Examples

- Aspect-invariant sentiment expressions

The ***service*** was good!

SERVICE#GENERAL [positive]

synthetic
example

The ***location*** was good!

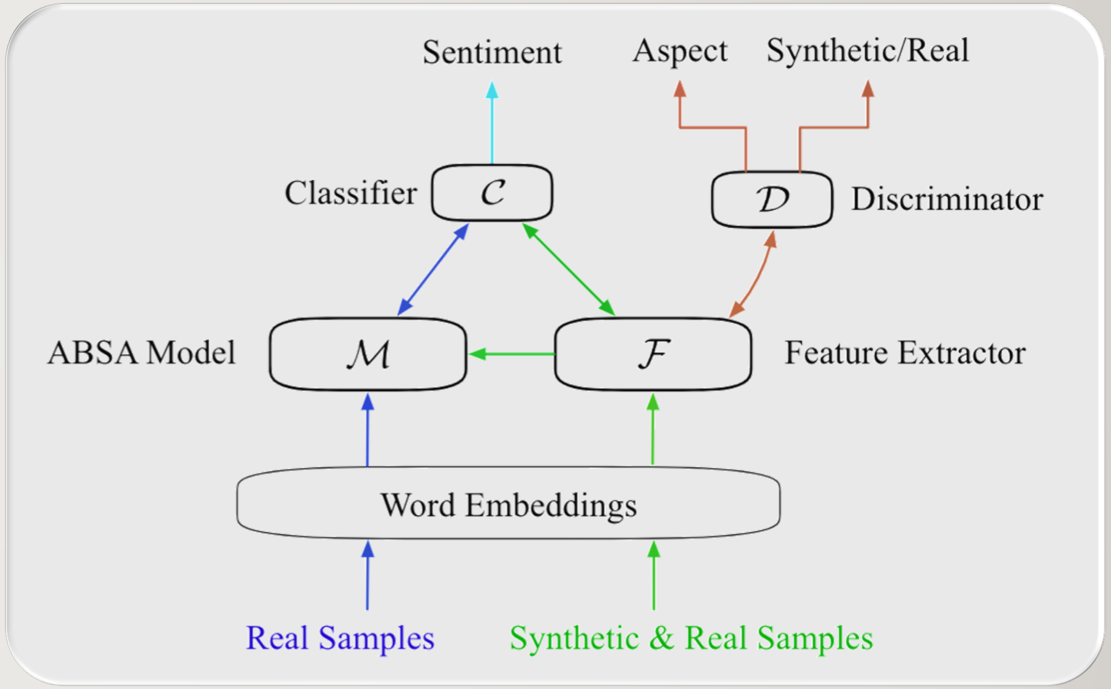
LOCATION#GENERAL [positive]

Adversarial Multi-Task Learning

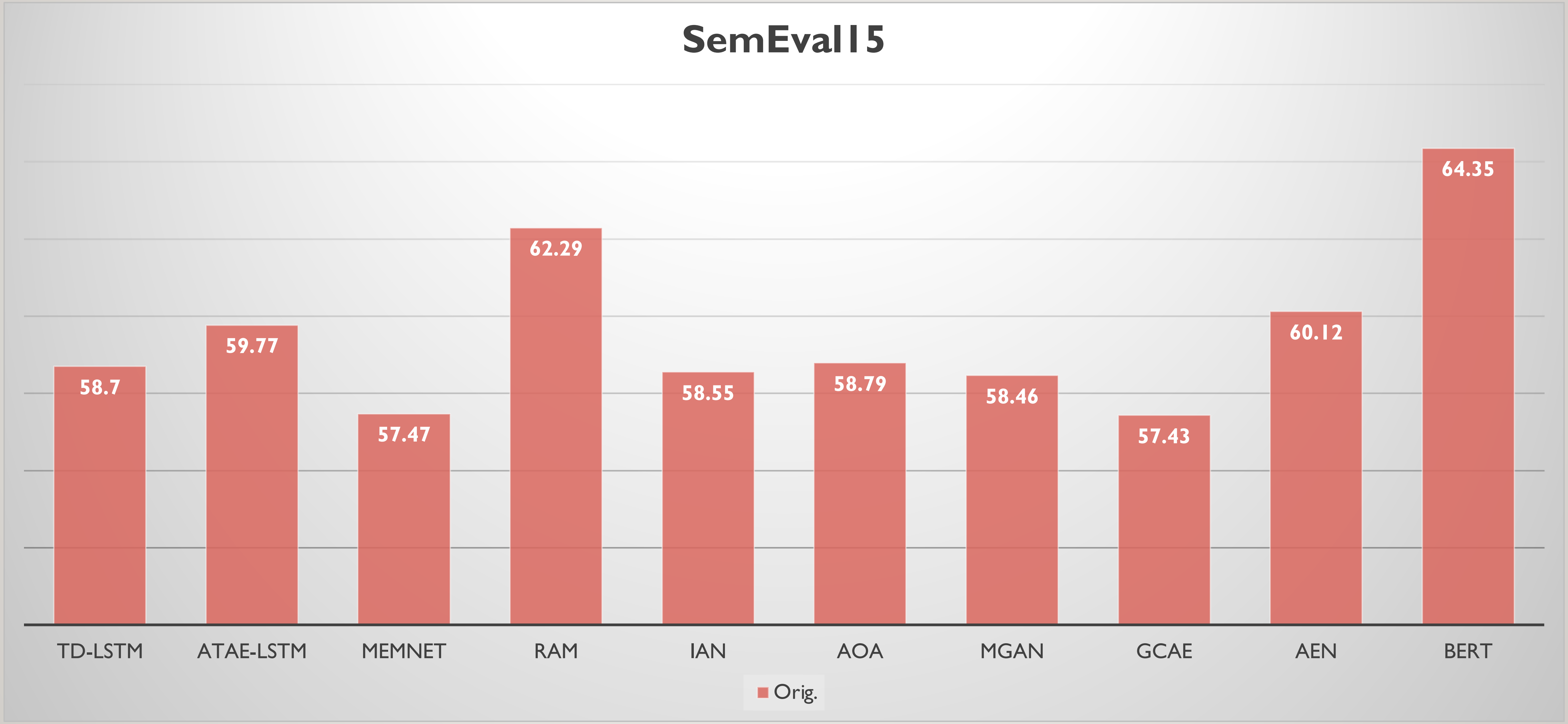
- First generate synthetic cross-aspect samples by replacing the aspect entities in the original training samples with other aspect entities.

- If the original training sample contain *aspect-invariant sentiment expressions*, then the discriminator would not be able to distinguish real/synthetic samples.

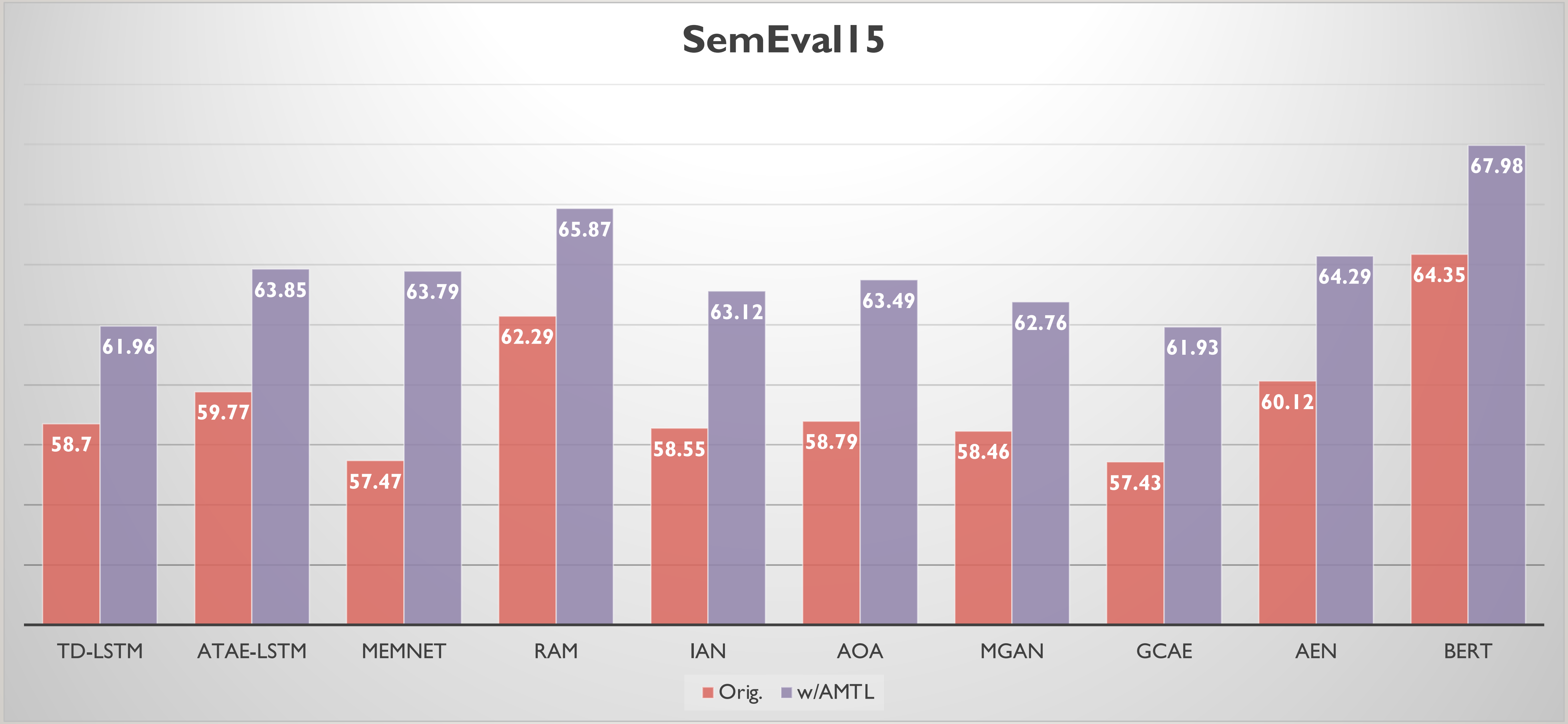
- If the original sample contains *aspect-dependent sentiment expressions*, then it is likely that the polarity label of the synthetic example would be wrong, which can be easily identified by the discriminator.



Aspect-Based Sentiment Classification Results (Micro-F1)



Aspect-Based Sentiment Classification Results (Micro-F1)



Open Challenges

Dealing with implicit aspects and/or implicit sentiments

Looks *nice*, and the surface is smooth, but certain apps takes seconds to respond.

Aspect Term	Aspect Category	Sentiment Expression	Polarity
NULL	Design	*nice*	Positive
surface	Design	*smooth*	Positive
apps	Software	*NULL*	Negative

(Cai et al., ACL 2021)

Limited labelled data – meta learning and transfer learning

	Training set Sentences	Test set Sentences	Aspect Categories	Attribute Categories
SemEval15-Laptop	1739	761	22	9
SemEval15-Restaurant	1315	685	6	5

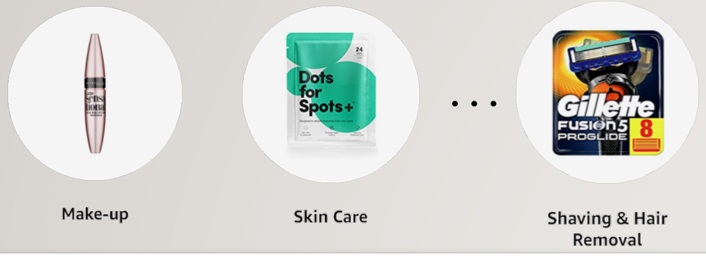
ABSA in domains beyond product reviews, e.g., patient reviews

I have had Lateral Flow tests. A lovely lady did the swab in a matter of think they have the right of way, nothing can be done about this I know, but worth noting if you are a first moments and I was advised what to do next. *Everything was clear and carefully explained.* The only downside - where it is situated, *the signage is clear but the road is narrow and windy* and there are some very big lorries entering and leaving the adjacent building site who seem to time visitor to the site.

Outline

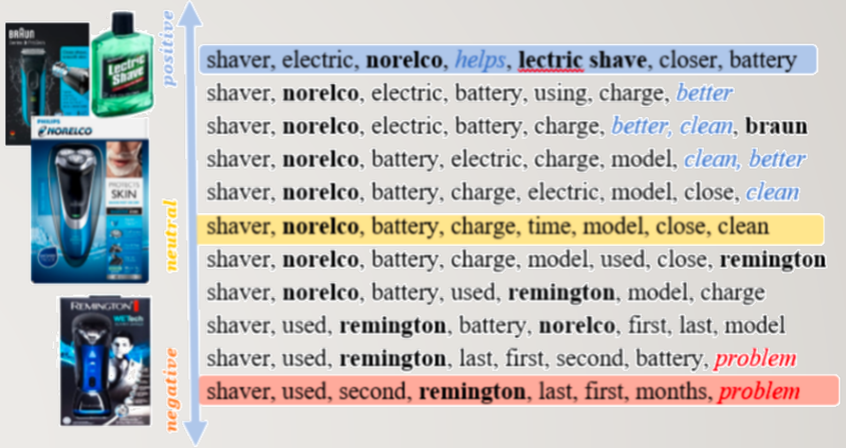
- Aspect-Based Sentiment Analysis
- Sentiment-Topic Extraction
- Emotion Cause Detection
- Dialogue Emotion Detection
- Review Question-Answering
- Interpretability in Sentiment Analysis

Brand Sentiment Analysis




Johnboy
Not impressed
john bristow
marvellous
A. Shopper
Reasonably effective lady shaver
Reviewed in the United Kingdom on 29 October 2018
Package Quantity: 1 | Verified Purchase
I love all my Remington hair tools, but this shaver is not the best. It is beautifully ergonomically designed and nice looking, but my problem is more with the double foils and the cutting edge. Shaving with this hurt my underarms (and I fully accept I may have used it incorrectly) and after one use, I decided to go back to the Philips lady shave, which I bought this to replace. So back to the Philips which I bought a week after this. By contrast with the Philips I discovered that the cutting edge is specifically designed with microscopic rounded tips, so it does not graze the skin and I think this makes all the difference.

Topics transition with varying polarity strength



Brand Score

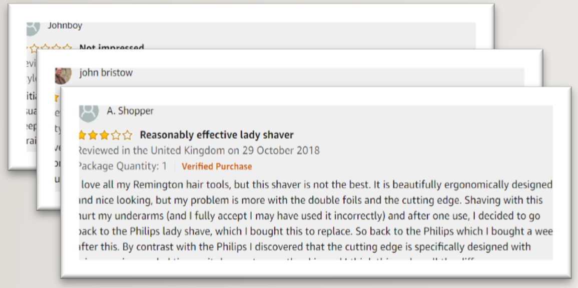
- **Brand-Topic Model (BTM)** aims to detect polarity-bearing topics associated to different brands discussed in product reviews.
- BTM automatically infers **real-valued brand-associated sentiment scores** and generate fine-grained sentiment-topics in which we can observe **continuous** changes of words under a certain topic (e.g., 'shaver' or 'cream') while its associated sentiment varies gradually.

Zhao et al., [Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews](), EACL 2021.

Problem Setup

Input

Collection of Reviews



+

Reviews Score

+

Product Brand

REMINGTON

PHILIPS NORELCO

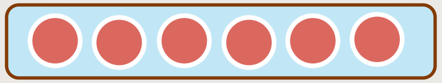
BRAUN

BTM

Brand Score $\in [-1, 1]$

+

Learned topic distribution



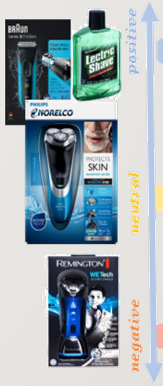
Generate

Brand Ranking

4.3 PHILIPS NORELCO

4.1 REMINGTON

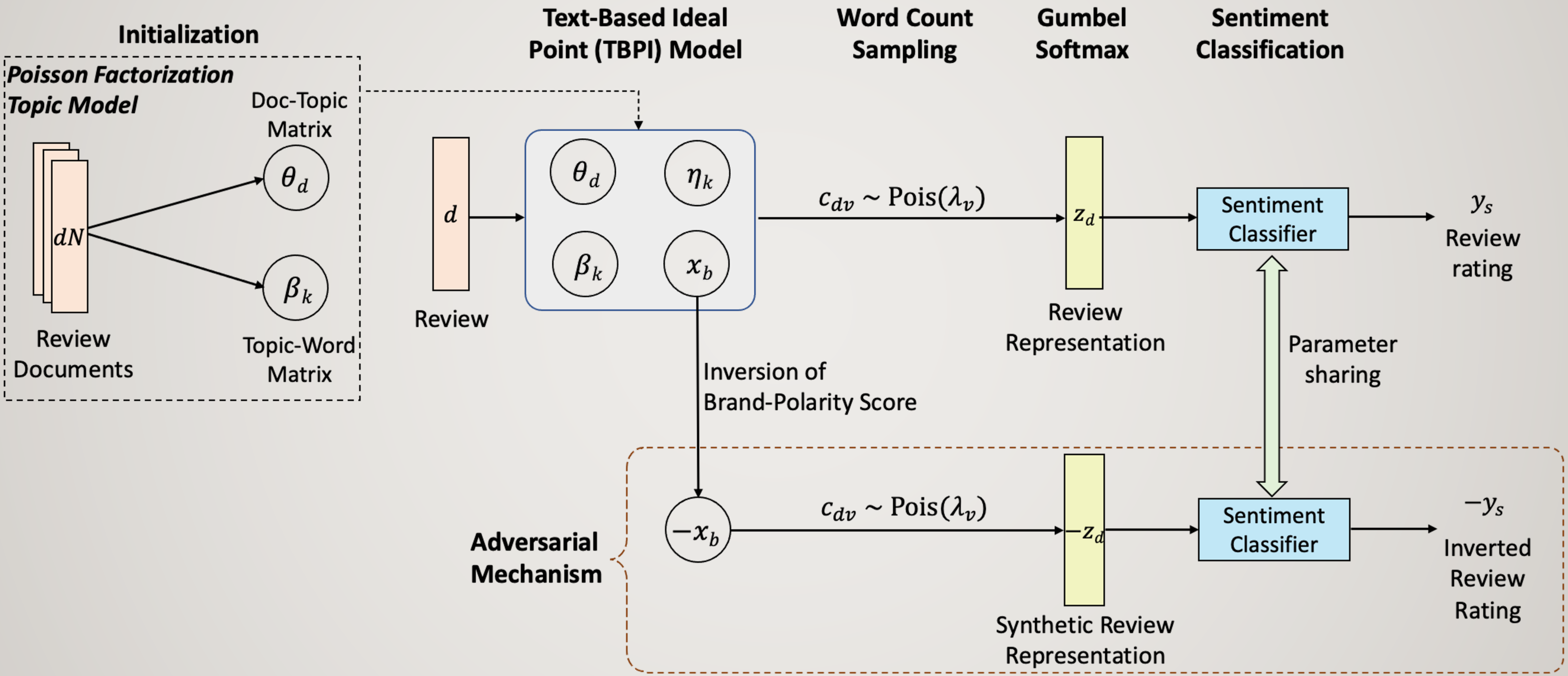
3.9 Brand Score $\in [-1, 1]$



shaver, electric, **norelco**, *helps*, **lectric shave**, closer, battery
shaver, **norelco**, electric, battery, using, charge, *better*
shaver, **norelco**, electric, battery, charge, *better*, *clean*, **braun**
shaver, **norelco**, battery, electric, charge, model, *clean*, *better*
shaver, **norelco**, battery, charge, electric, model, close, *clean*
shaver, **norelco**, battery, charge, time, model, close, clean
shaver, **norelco**, battery, charge, model, used, close, **remington**
shaver, **norelco**, battery, used, **remington**, model, charge
shaver, used, **remington**, battery, **norelco**, first, last, model
shaver, used, **remington**, last, first, second, battery, *problem*
shaver, used, second, **remington**, last, first, months, *problem*

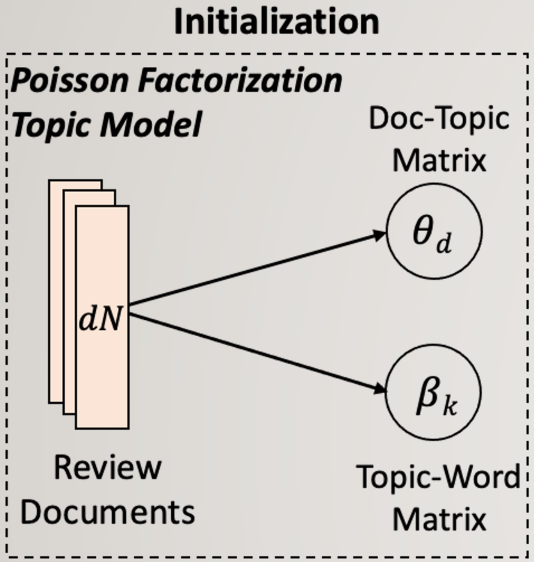
Controllable fine-grained topics generation

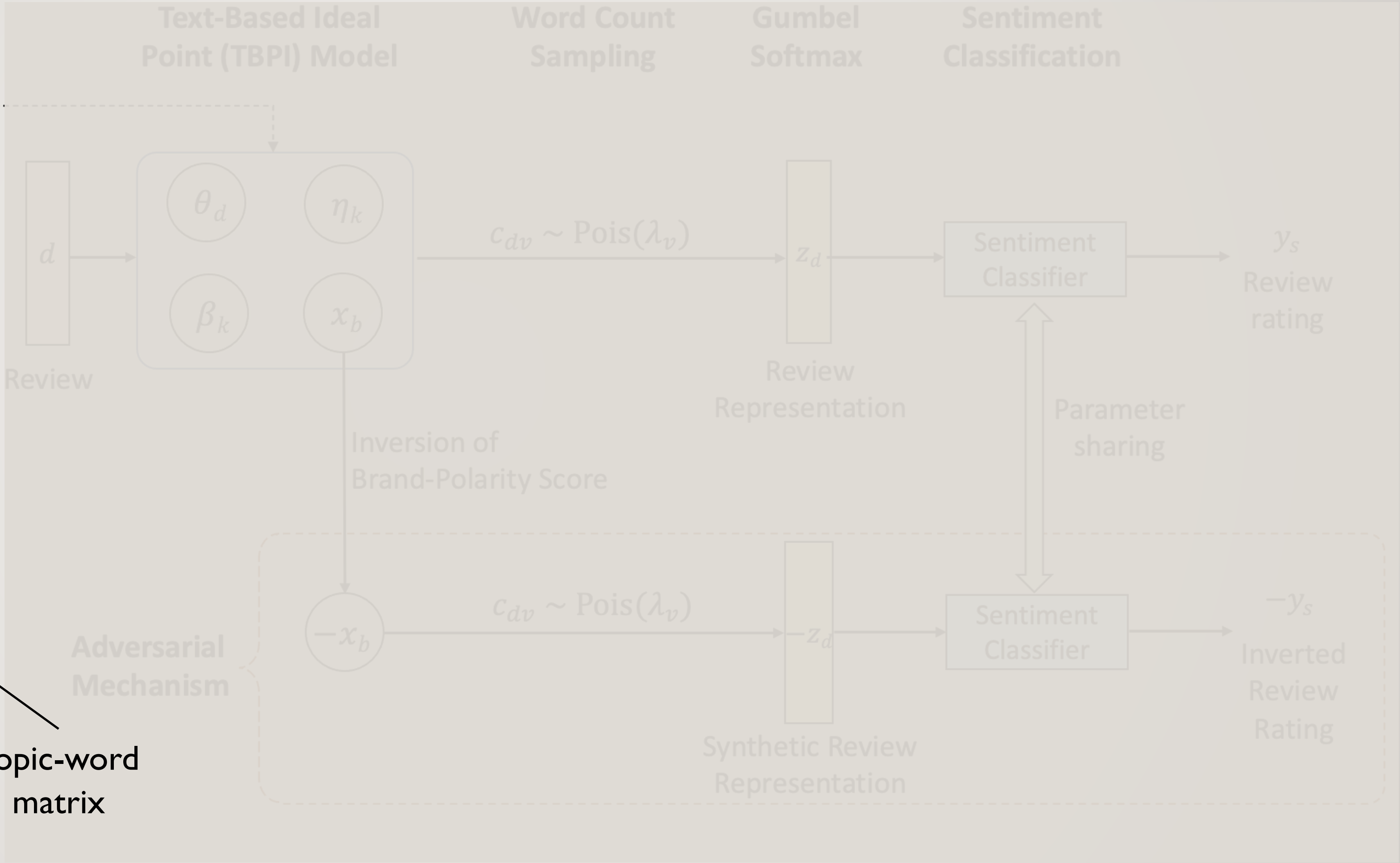
Brand-Topic Model (BTM)



Zhao et al., [Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews](), EACL 2021.

Brand-Topic Model (BTM)





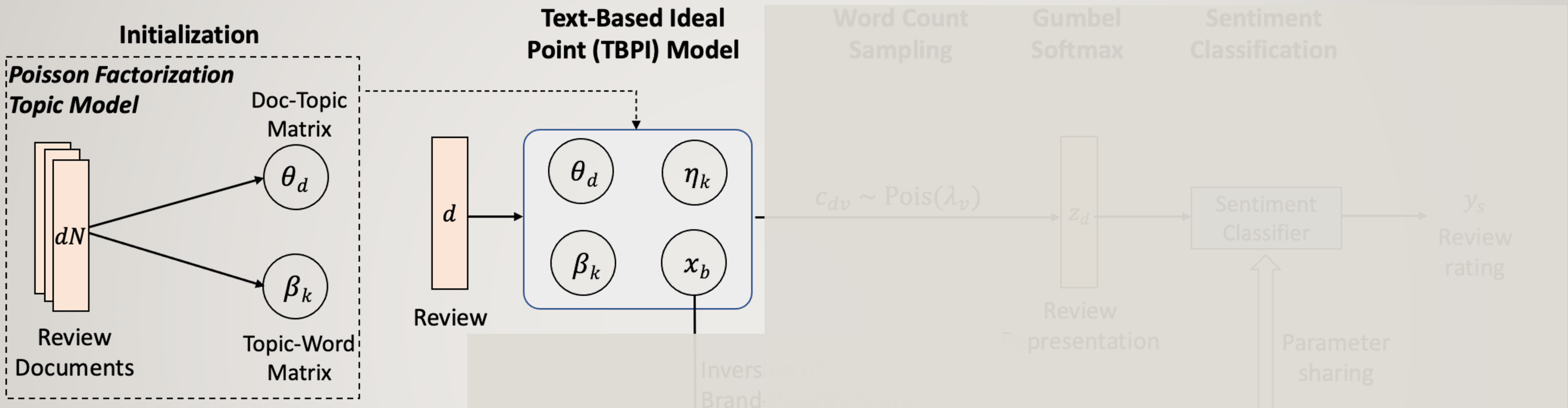
$$c_{dv} \sim \text{Poisson}\left(\sum_k \theta_{dk}\beta_{kv}\right)$$

Word count for a term v in doc d

Per-doc topic intensity

Topic-word matrix

Brand-Topic Model (BTM)



$$c_{dv} \sim \text{Poisson}\left(\sum_k \theta_{dk} \exp\{\log \beta_{kv} + x_{b_d}\eta_{kv}\}\right)$$

Word count for a term v in doc d

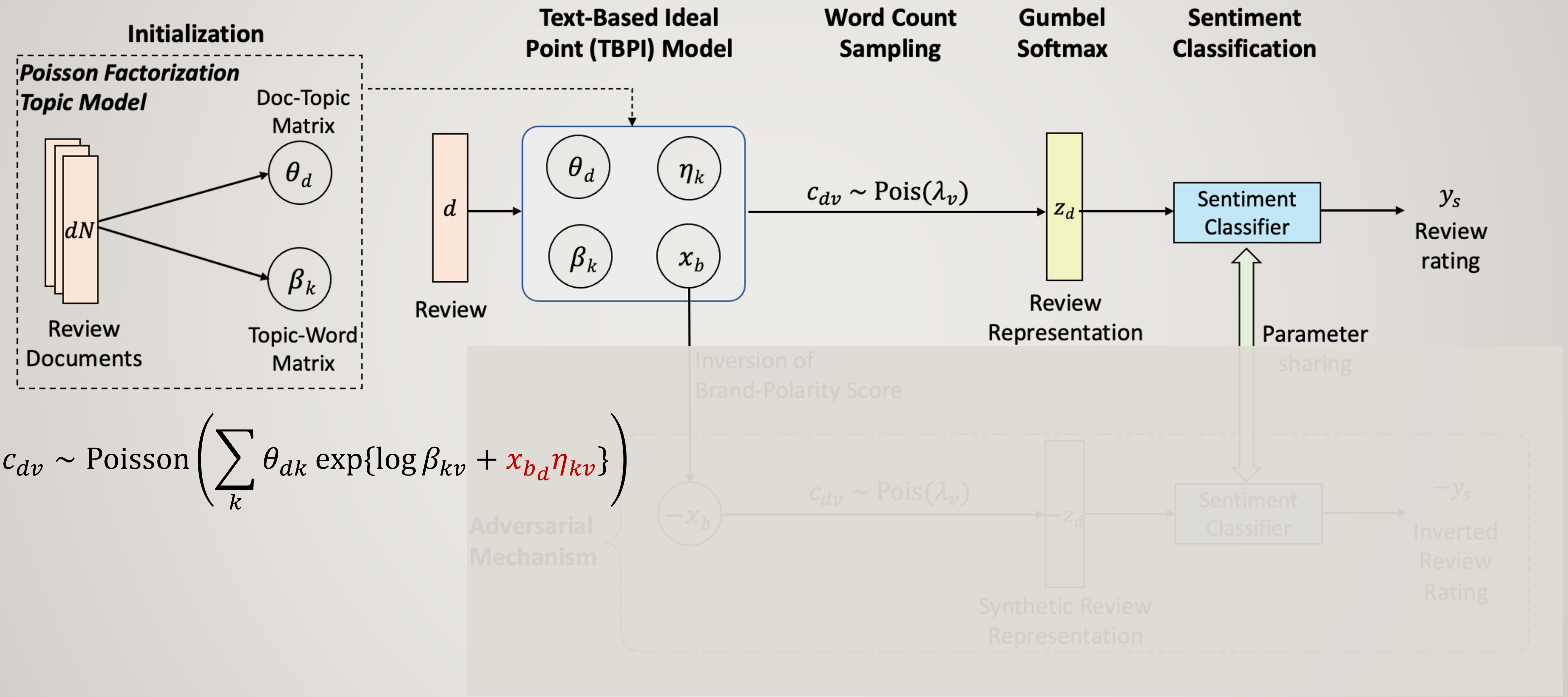
Per-doc topic intensity

Topic-word matrix

Doc brand-polarity score

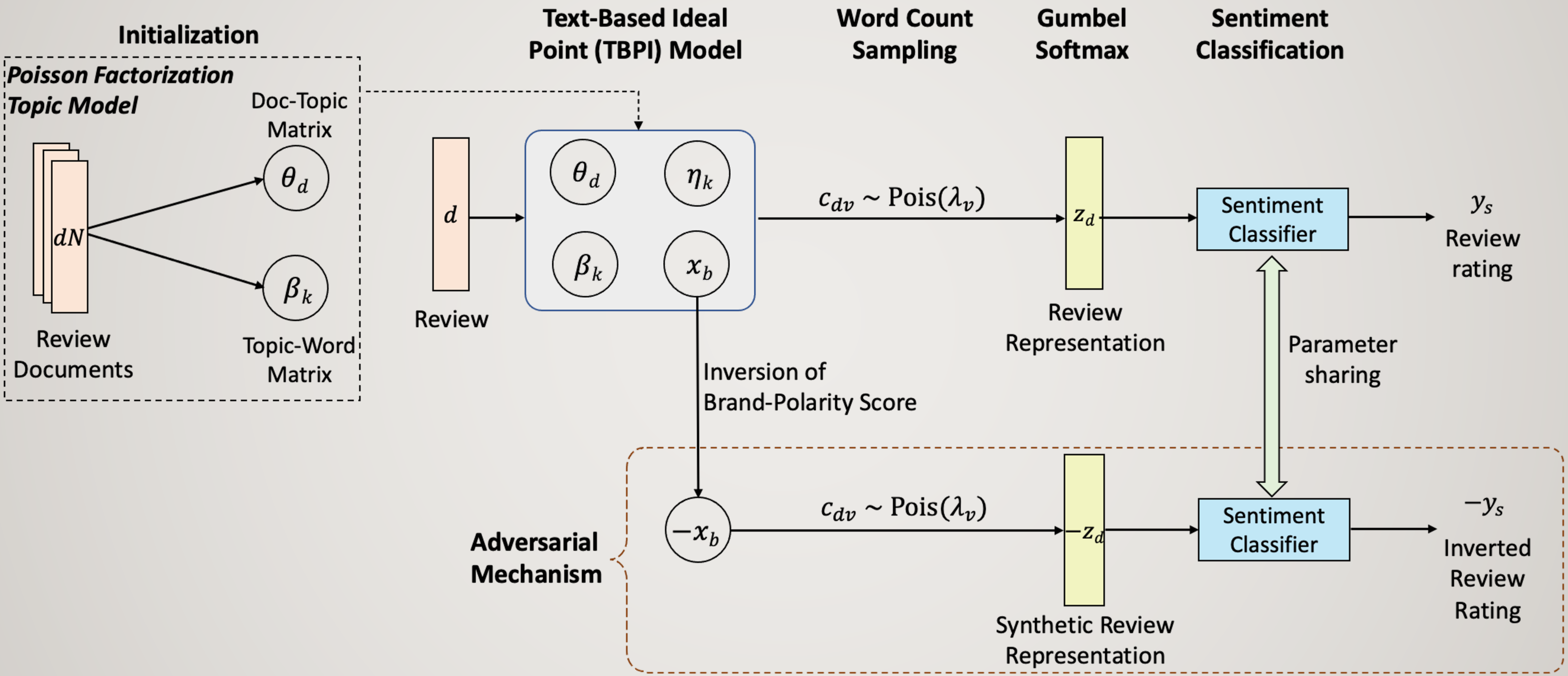
Topic-word offset

Brand-Topic Model (BTM)



$$c_{dv} \sim \text{Poisson}\left(\sum_k \theta_{dk} \exp\{\log \beta_{kv} + x_{b_d}\eta_{kv}\}\right)$$

Brand-Topic Model (BTM)



Zhao et al., [Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews](), EACL 2021.

Example Sentiment Topics

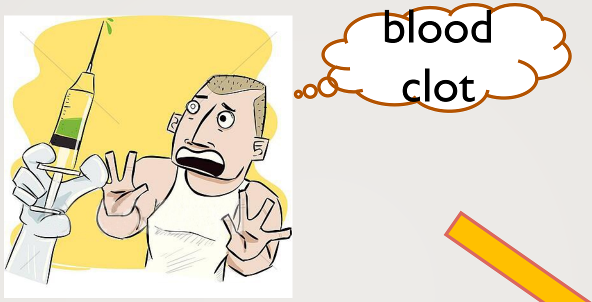
Topic Label	Sentiment Topics	Top Words
		BTM
Brush	Positive	brushes, cheap, came, pay, *pretty*, brush, *okay*, case, glue, *soft*
	Neutral	cheap, feel, set, buy, *cheaply made*, feels, made, worth, spend, bucks
	Negative	plastic, made, cheap, parts, feels, *flimsy*, money, *break*, metal, bucks
Oral Care	Positive	teeth, taste, mouth, strips, crest, mouthwash, tongue, using, *white*, rinse
	Neutral	teeth, *pain*, mouth, strips, using, taste, used, crest, mouthwash, *white*
	Negative	*pain*, *issues*, causing, teeth, caused, removing, wore, *burn*, little, cause
Duration	Positive	stay, pillow, *comfortable*, string, tub, mirror, stick, back, months
	Neutral	months, year, *lasted*, *stopped working*, *sorry*, n, worked, working, u, last
	Negative	months, year, last, *lasted*, battery, warranty, *stopped working*, *died*, *less*
		TBIP
Brush	Positive	love, *favorite*, products, *definitely recommend*, forever, carry, brushes
	Neutral	love, brushes, *cute*, *favorite*, *definitely recommend*, soft, *cheap*
	Negative	love, brushes, cute, *soft*, *cheap*, set, case, quality price, buy, bag
Oral Care	Positive	teeth, strips, crest, mouth, mouthwash, taste, *white*, *whitening*, sensitivity
	Neutral	teeth, strips, mouth, crest, taste, work, *pain*, using, *white*, mouthwash
	Negative	teeth, strips, mouth, crest, taste, work, *pain*, using, *white*, mouthwash
Duration	Positive	great, *love shampoo*, *great price*, *great product*, *lasts long time*
	Neutral	*great*, *great price*, *lasts long time*, *great product*, price, *works expected*
	Negative	quality, *great*, *fast shipping*, *great price*, *low price*, price quality, hoped

Zhao et al., [Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews](), EACL 2021.

Vaccine Attitude Detection

Vaccine-related tweets

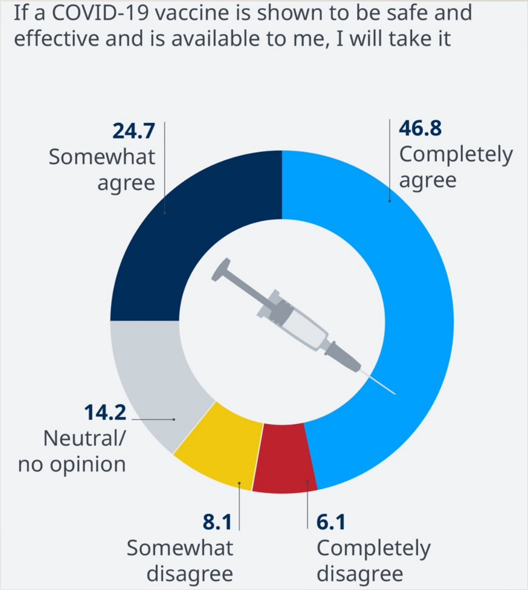
Aspect discovery



Stance classification

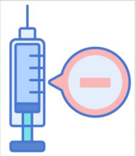
:-(:-| :-)

Aspect-based stance classification



Survey: Majority of people would get vaccinated. Nature (June 2020) global survey of 13,000+ people across 35 countries most affected by COVID-19. https://www.nature.com/articles/s41591-020-1124-9

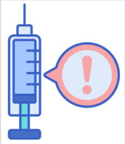
Vaccine attitude detection: why is it hard?



Galileo Galilei
@TheREALGalileo
Follow

Have felt for the past 24 hours that I've been run over by three double decker buses after the AstraZeneca vaccine yesterday morning. Starting to feel a little normal now but it's not been nice!

10 189 964



Nicolaus Copernicus
@TheREALCopernicus
Follow

There are some very interesting ties between this vaccines creators and the eugenics movement which is concerning considering it's mainly been promoted as a vaccine for poor folks in the third world

Vaccine attitude detection: why is it hard?

Vaccine attitude detection aims to **extract people's opinions towards vaccines** by analyzing their online posts.

- User attitude emerges from positive or negative **stances** related to a wide spectrum of **aspects**:

Individual and group influences	Vaccine safety
	Lack of information
	Low risk/severity of disease
	Vaccines not effective
	Mistrust in health institutions
	Healthy bodies belief
	Social norms
	Vaccination not a priority
	Against vaccination in general
	Alternative prevention methods
	Diseases are beneficial
	Fear of injection
	Previous negative experiences
	Humans too weak to fight vaccines
	Responsibility if something bad happens

Vaccine and vaccination specific issues	No medical need
	Access
	Financial cost
	Lack of recommendation from providers
	Vaccine novelty
	Inconsistent advice from providers

Contextual influences	Conspiracy theories
	Religious fatalism
	Negative exposure to media
	Violation of human rights

2

2. **Rapid literature review on motivating hesitant population groups in Europe to vaccinate**. https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/vaccina

Vaccine attitude detection: why is it hard?

Vaccine attitude detection aims to **extract people's opinions towards vaccines** by analyzing their online posts.

- User attitude emerges from positive or negative **stances** related to a wide spectrum of **aspects.**

A high (and growing) number of aspects

Limited availability of annotated data

VADet: Vaccine-Attitude Detection Model

- The goal of our work is to :
  - **detect the stance** expressed in a tweet (i.e., ‘pro-vaccination’, ‘antivaccination’, or ‘neutral’),
  - **identify a text span** that indicates the concerning **aspect** of vaccination,
  - **cluster** tweets into groups that share similar aspects.

Vaccine Attitude Detection (**VADET**) Model

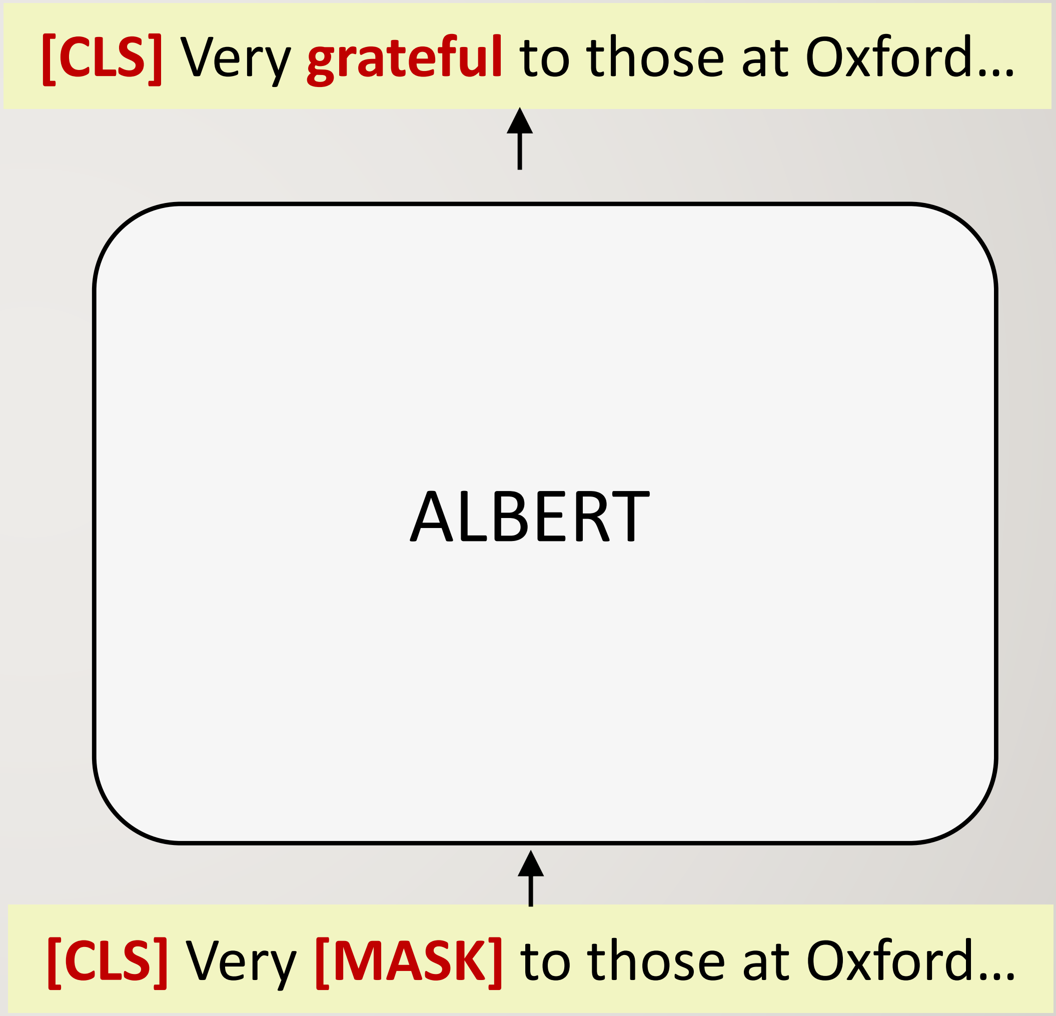
1. Initially trained on a large amount of **unannotated** Twitter data to learn latent **topics**
2. Then, it is fine-tuned on a small amount of Twitter data **annotated with stance labels** and **aspect text spans**

Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

VADet: Vaccine-Attitude Detection Model

1. Unsupervised Step

Language model fine-tuned on vaccine-related tweets

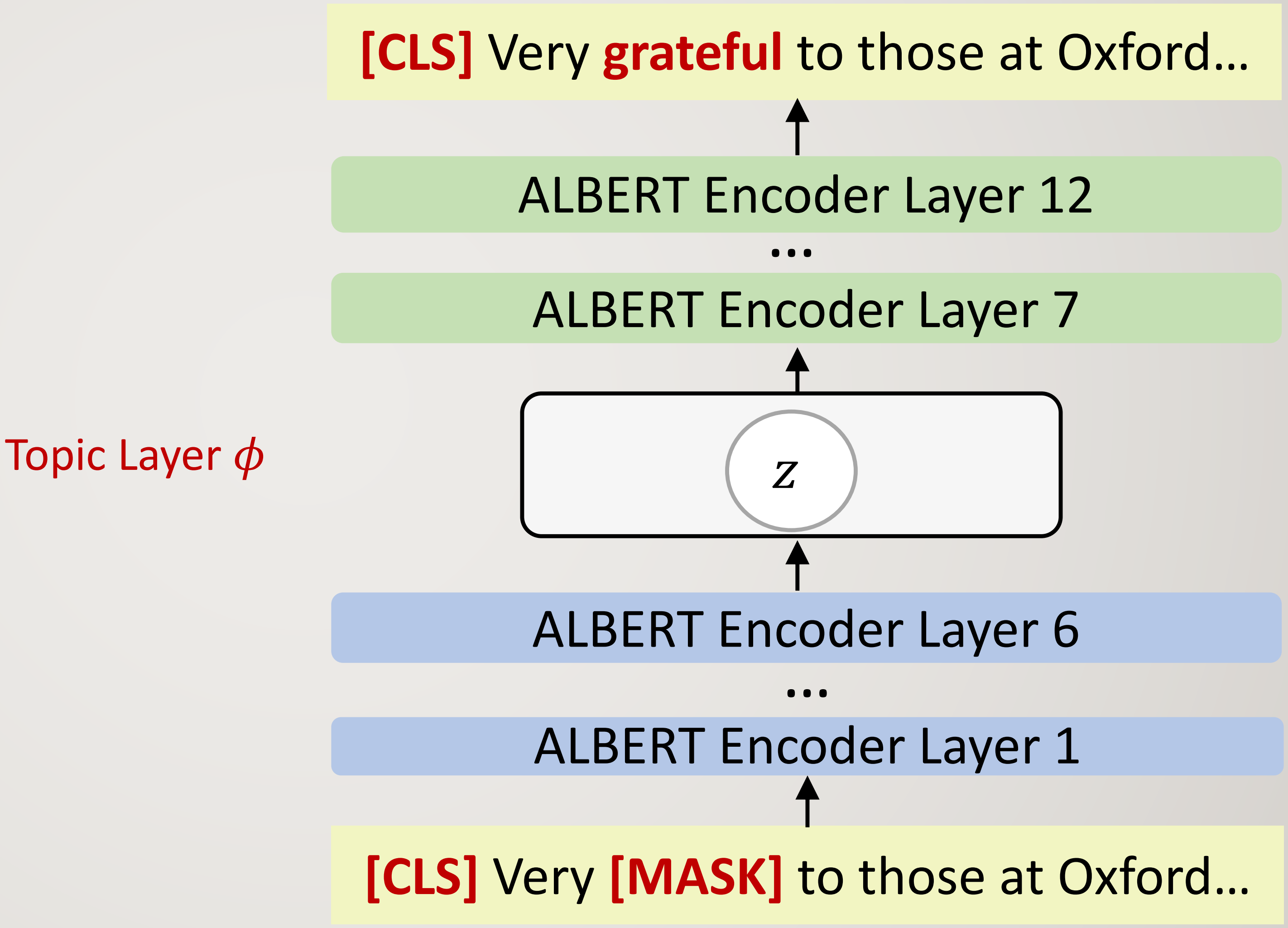


Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

VADet: Vaccine-Attitude Detection Model

1. Unsupervised Step

Language model enrichment via a Topic Layer

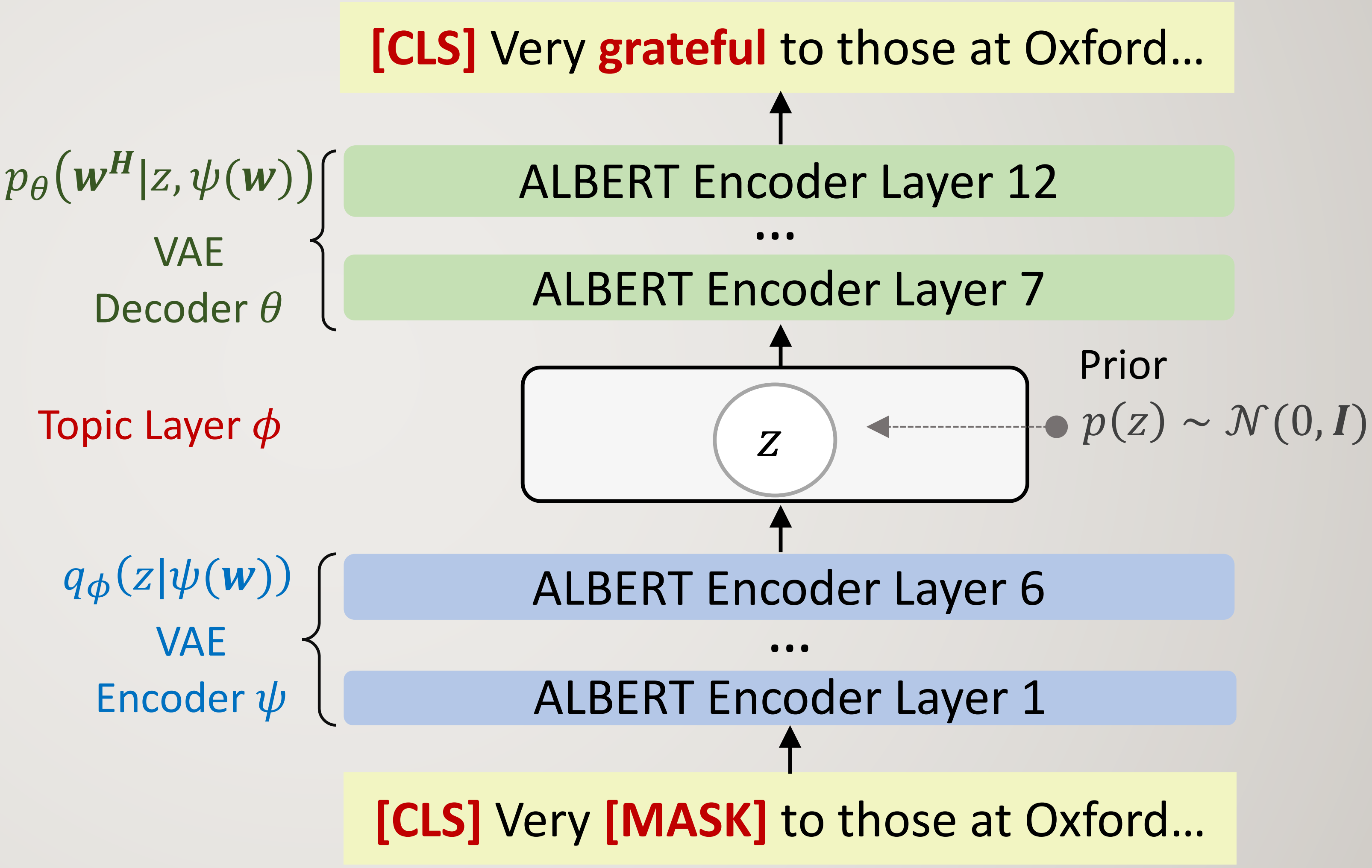


Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

VADet: Vaccine-Attitude Detection Model

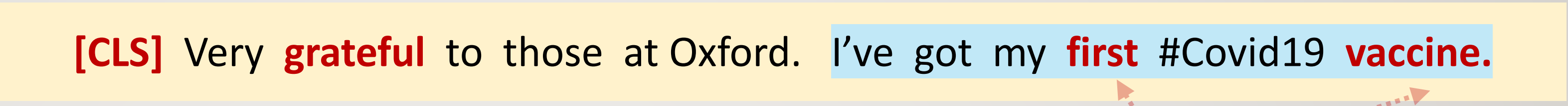
1. Unsupervised Step

Language model enrichment via a Topic Layer



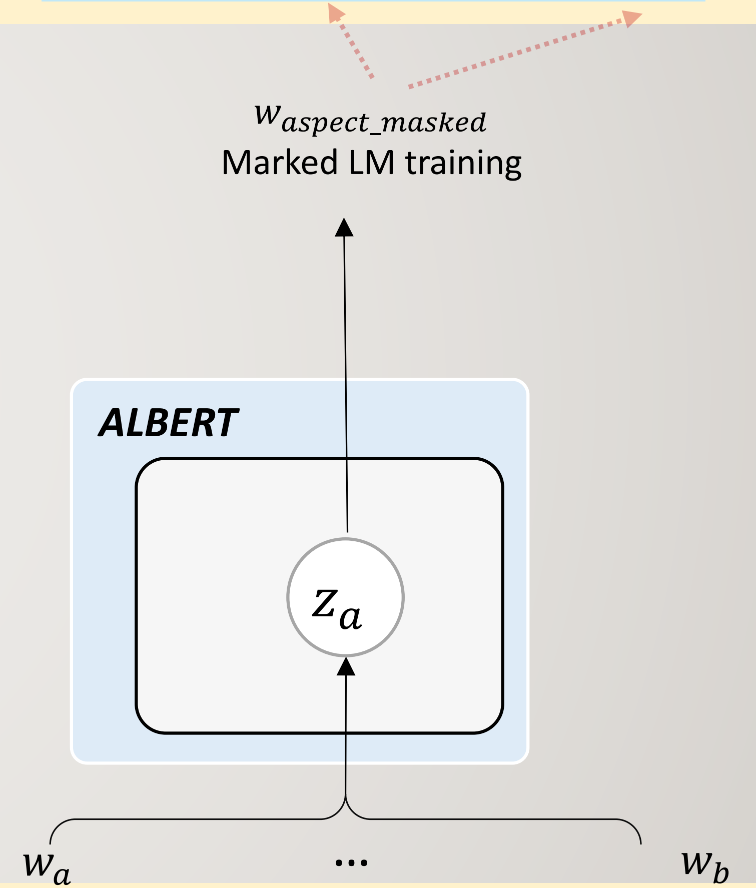
Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

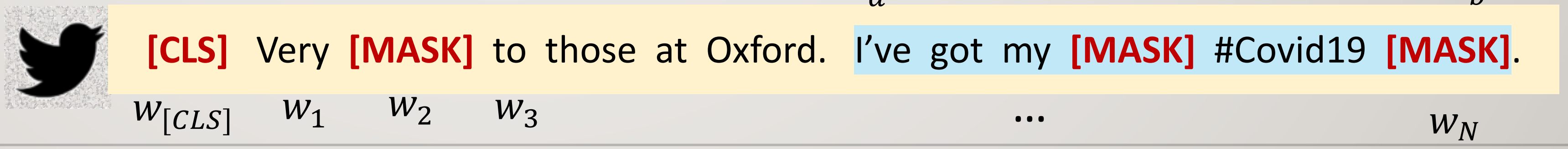
VADet: Vaccine-Attitude Detection Model



2. Supervised step

- A pre-trained language model is applied to the aspect span.

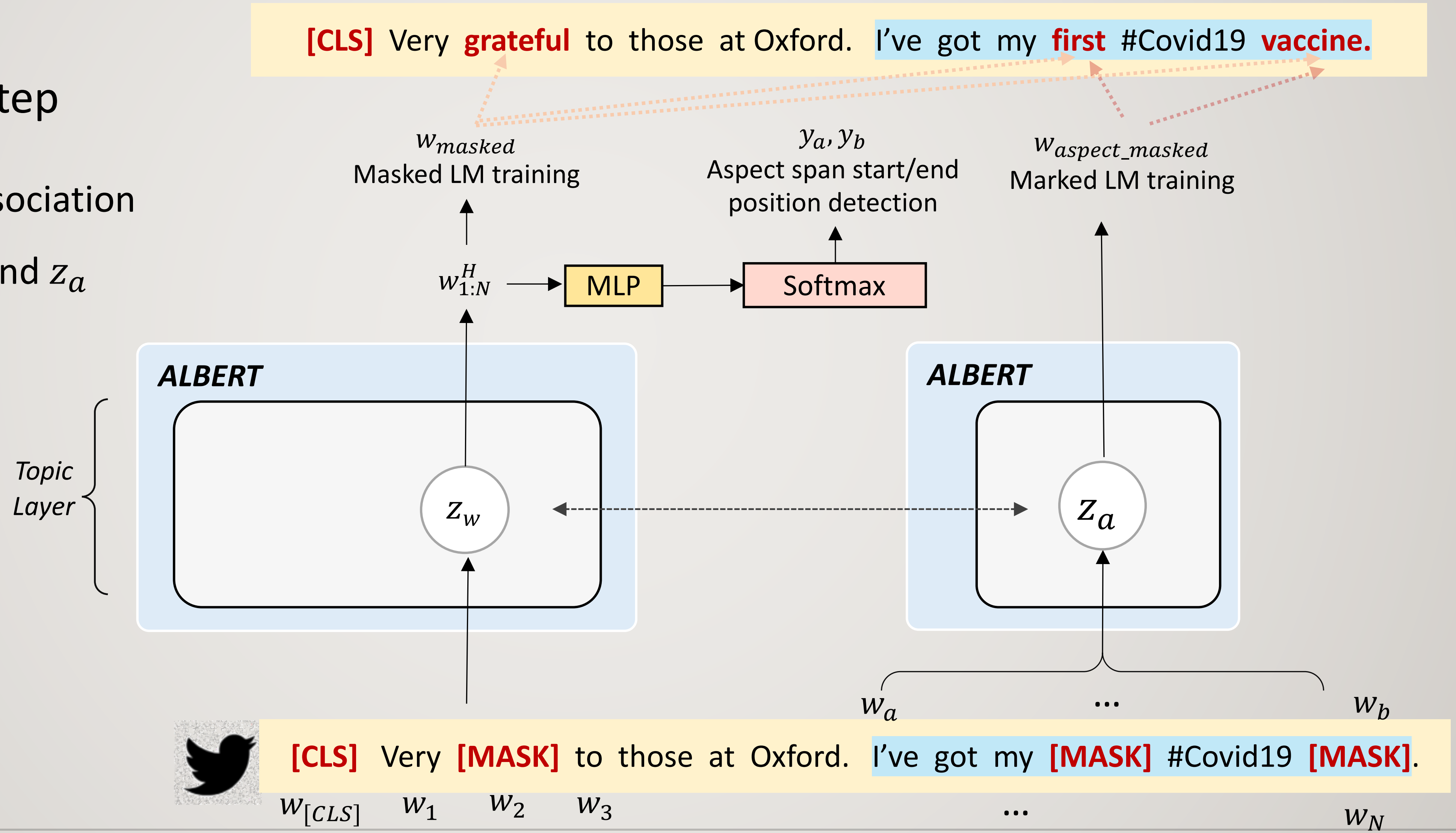




VADet: Vaccine-Attitude Detection Model

2. Supervised step

- Enforce an association between z_w and z_a

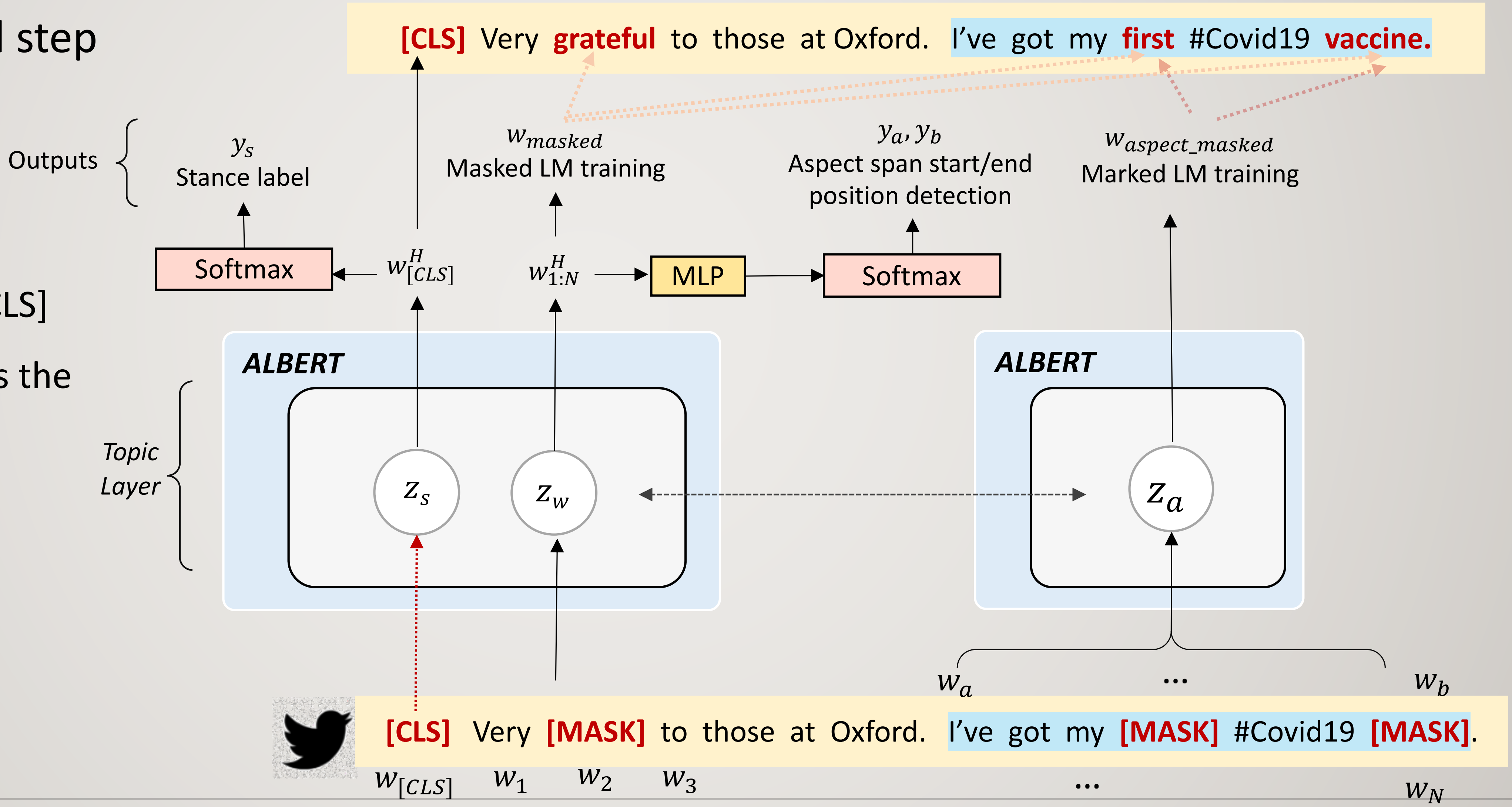


Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

VADet: Vaccine-Attitude Detection Model

2. Supervised step

- z_s is used to reconstruct [CLS] which predicts the stance label



Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

Dataset

- **VAD: V**accine **A**ttitude **D**ataset**:**
  - *1.9 million English tweets* collected between February 7th and April 3rd 2021 using 60 predefined keywords relating to COVID-19 vaccines
  - *2,800 tweets have been annotated* of which 2,000 are used for training and the remaining 800 are used for testing.
    - Each tweet is annotated with a stance label and an aspect span.

TAKE HEED Norwegian Experts Say Deadly Blood Clots WERE Caused by the AstraZeneca Covid Vaccine "Nothing but the vaccine can explain why these individuals have had this immune response", says Professor and Chief Physician Pal Andre Holme.

Stance: Negative
Aspect span: Deadly Blood Clots WERE Caused by the AstraZeneca Covid Vaccine

Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

Visualisation

Personal freedom to choose in relation to vaccines

Queensland senator and leader of the One Nation party Pauline Hanson is fighting back against **mandatory vaccines** in Australia . #NoMandatoryVaccines #MyBodyMyChoice #VaccinePassports #Australia #MedicalApartheid

Immunity level

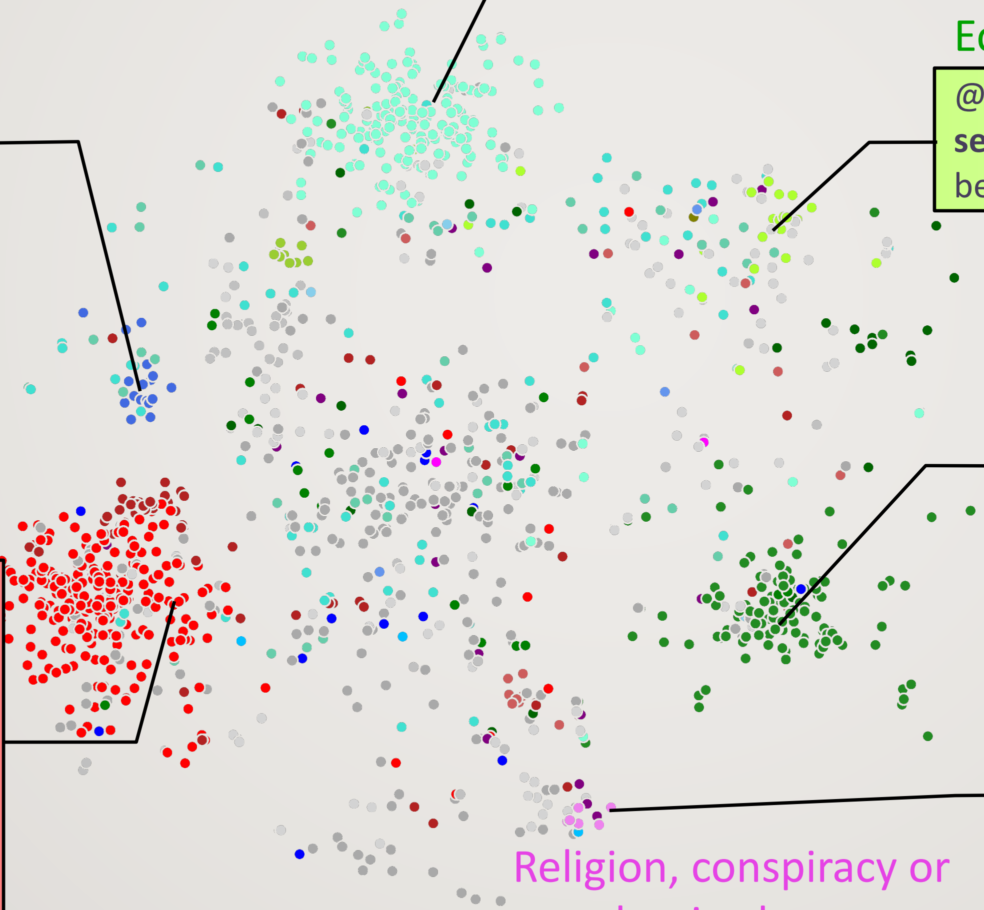
@user **AstraZeneca vaccine proven to reduce hospitalisation of over 80s by 80% after just one dose** The opposition are full of _pile_of_poo_

Economic effect

@user I am not a Putin fan at all , but this **Sputnik vaccine seems to be one of the best in the business** . Certainly better than AstraZeneca's jab .

Personal experiences

A year ago **I tried the Moderna vaccine to see if it was safe** . Spoiler : It is ! Now , on my #COVIDvaccine anniversary , I??m happy to share that I just got a 3rd dose . This booster experiment will reveal 1 if strainadapted vaccines boost immunity 2 whether they are safe .



(Adverse) side effects

The MHRA has published new vaccine adverse reaction data , so expect more in this series soon : ??Covid19 Vaccine?? Adverse Reactions; Part Two : **AstraZeneca Is Not A Safe Option** #Budget #BudgetSpeech2021 #Budget21 #RishiSunak #PMQs #COVID19Vaccine

Religion, conspiracy or moral attitudes

@user Nothing cures or protects against viruses . **This is a trial vax , gene therapy** . mRNA changes DNA . This is mass depopulation/sterilisation there??s warning on vax websites , no sex for a month after a shot incase of sterilisation . It??s a politicised 99 . 98% survival rate FLU

Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.

Open Challenges

- Sentiment-topic extraction on streaming data – dealing with sentiment/topic drift
- Interpretability of sentiment-topics
- Human-in-the-loop sentiment-topic extraction

Outline

- Aspect-Based Sentiment Analysis
- Sentiment-Topic Extraction
- Emotion Cause Detection
- Dialogue Emotion Detection
- Review Question-Answering
- Interpretability in Sentiment Analysis

Emotion Cause Extraction: Task Definition

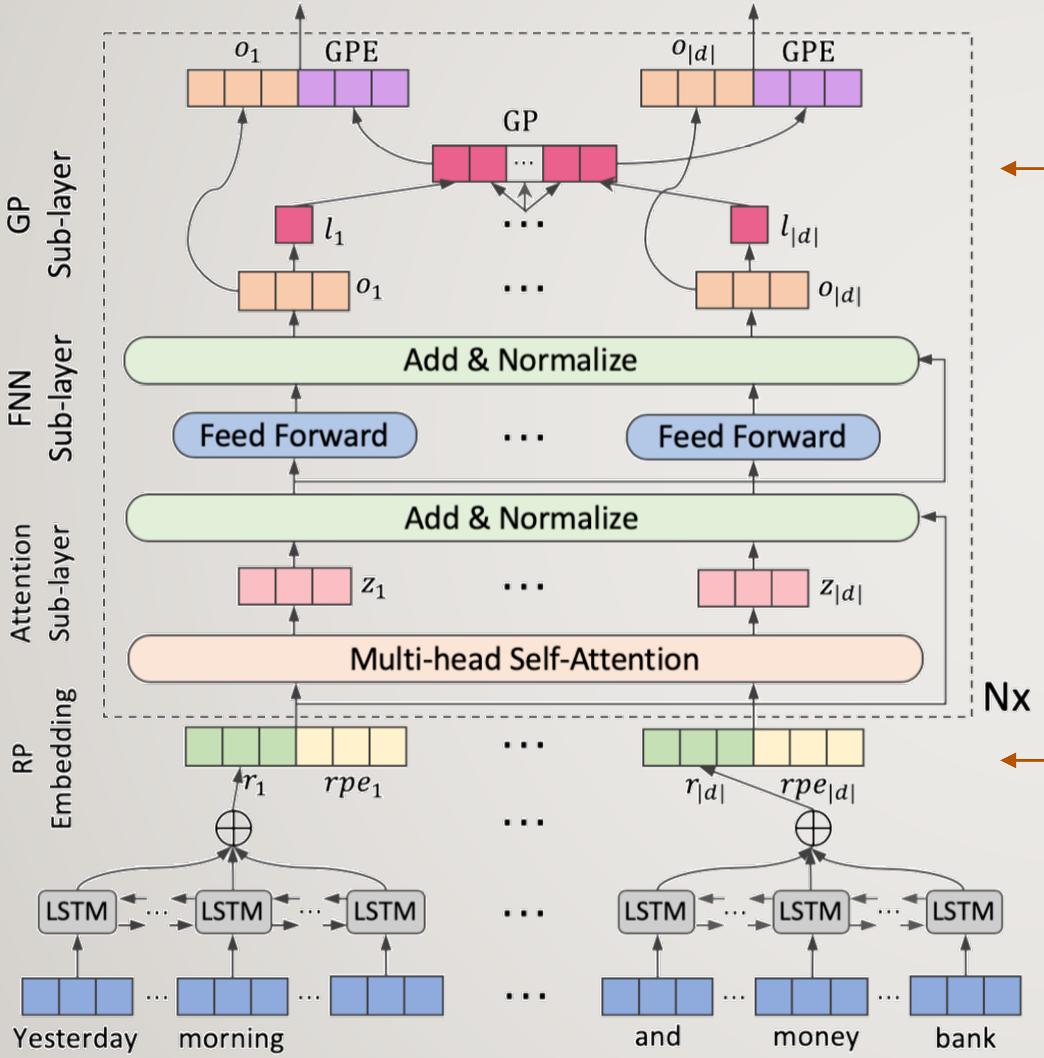
News article

1. The crime that ten people were killed shocked the whole country.
2. This was due to personal grievances.
3. Qiu had arguments with the management staff,
4. and thought the Taoist temple host had molested his wife.
5. He became **angry**,
6. and killed the host and destroyed the temple.

Angry

emotion cause

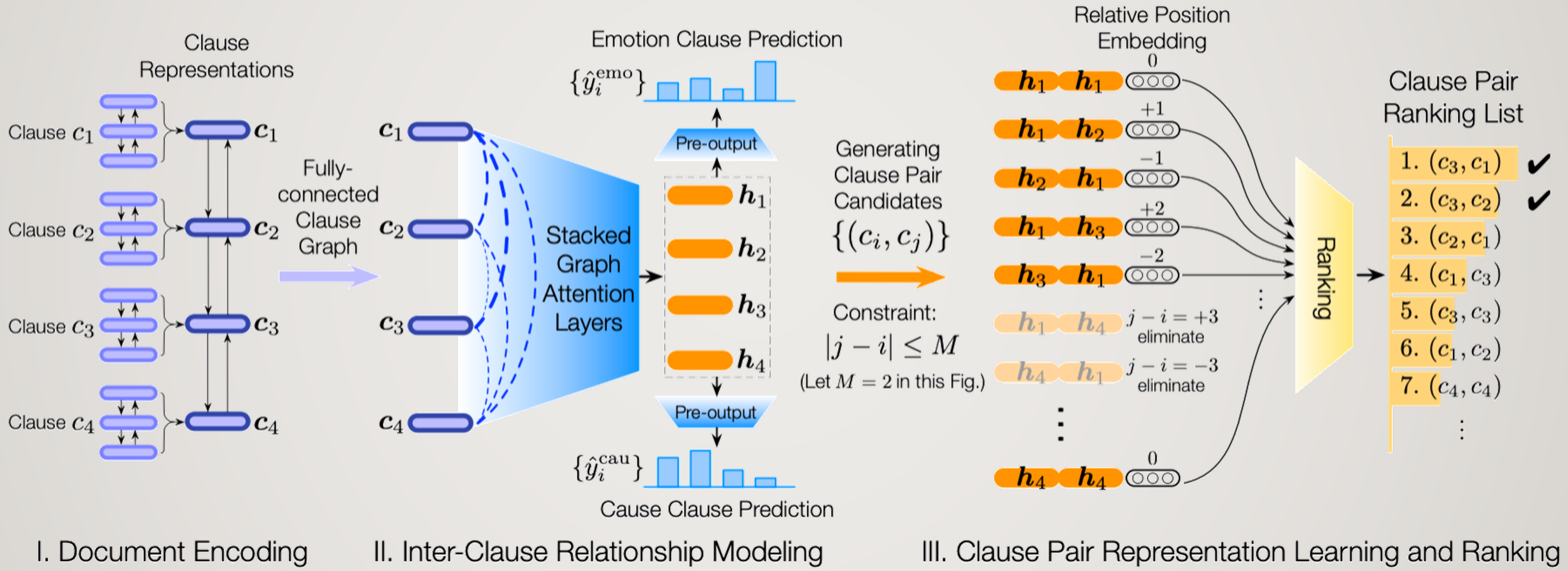
RNN-Transformer Hierarchical Network (RTHN)



- Concatenate Transformer output with GP embedding
- Form Global Prediction (GP) vector
- Predict label for each clause
- Clause-level transformer
- Add the relative position (w.r.t. the emotion clause) embedding rpe_d
- Word-level Bi-LSTM to encode clause representation r_d

Rui et al. [RTHN: an RNN-Transformer Hierarchical Network for Emotion Cause Extraction](). AAAI 2019.

End-to-End Emotion-Cause Pair extraction



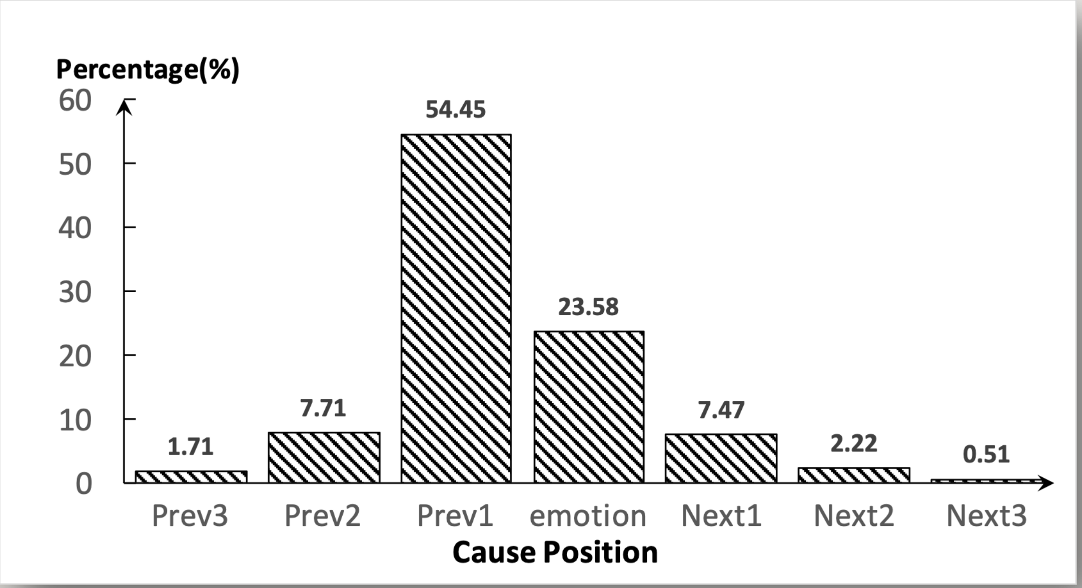
Word-level & clause-level RNN

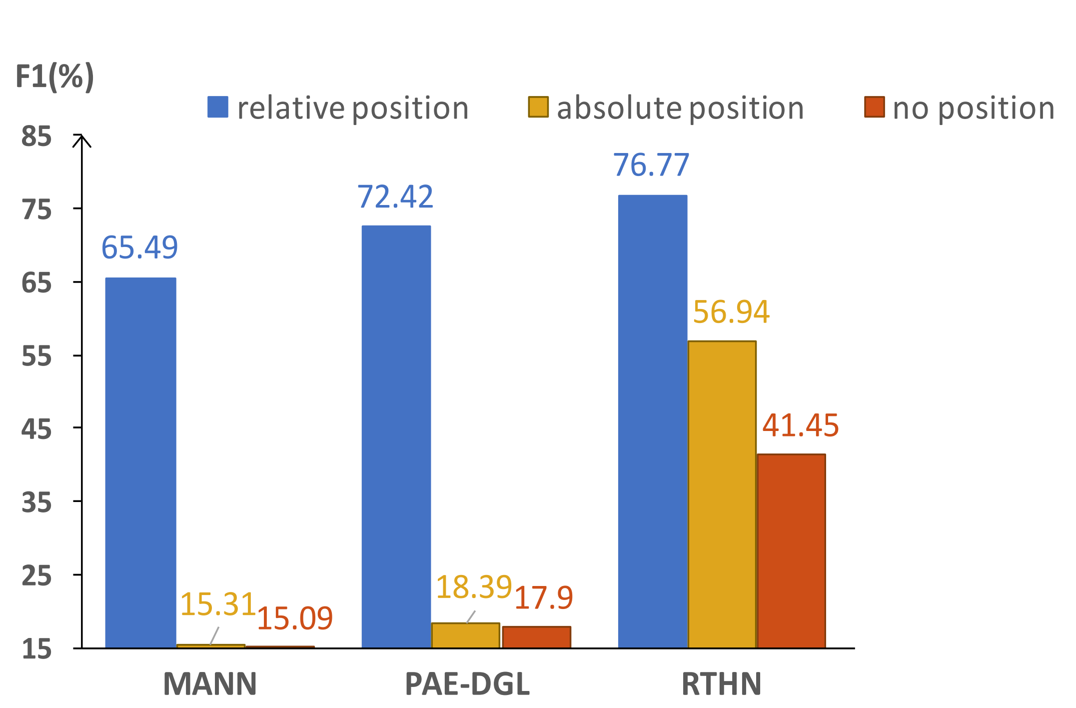
Fully-connected GATs
+
Two separate classifiers

Kernel-based
Relative Position Embedding

Position Bias Problem

- Emotion Cause Extraction (ECE) dataset (Gui et al., 2016)
  - 2,105 documents from SINA city news
  - Each document has only one emotion clause and one or more cause clauses

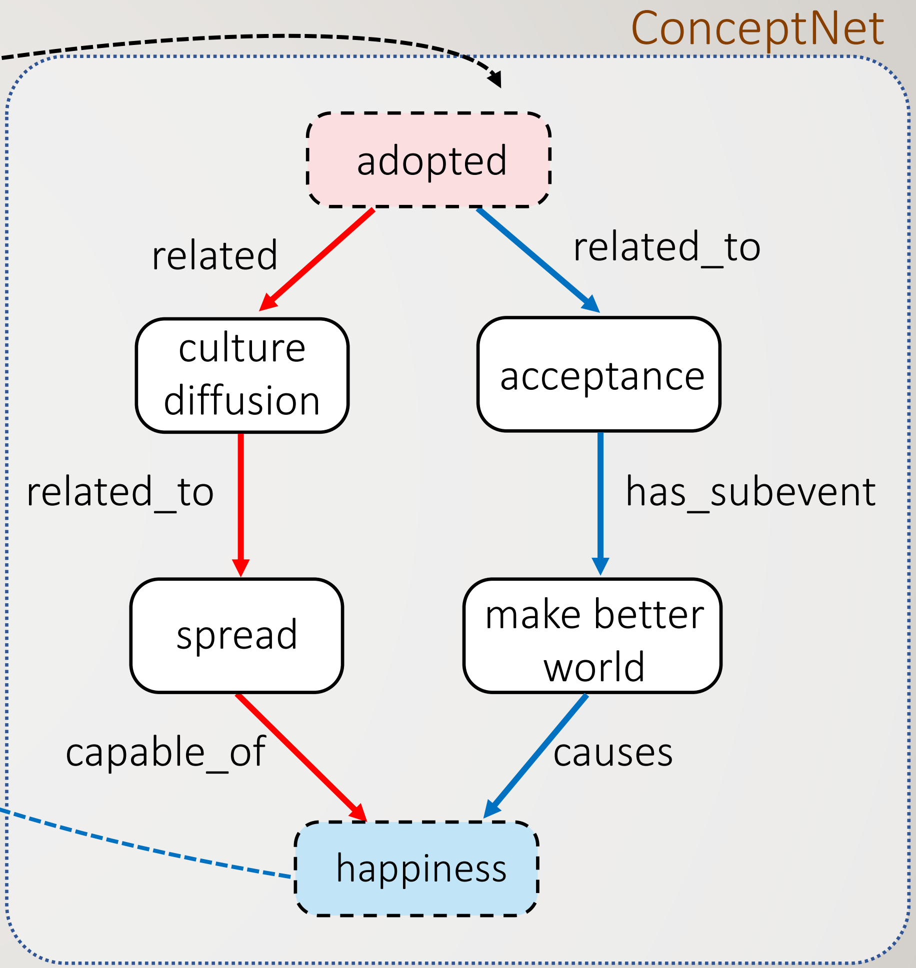


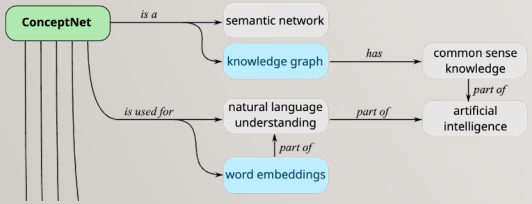


Leverage Commonsense Knowledge for Emotion Cause Extraction

Clause emotion label

1. "When I found that my advice had been adopted
2. and the department leader replied to my mail
3. I realized that I had made contributions to the country's development".
4. Bai Jinyue, an ordinary worker in XingTai Steel factory in HeBei province,
5. talked to the journalist with **excitement**, — Happiness
6. with a Thank-You letter in his hands.
7. Since 27 years ago different departments,
8. such as the public security,
9. and the Committee on Science and Technology, have accepted his advice.





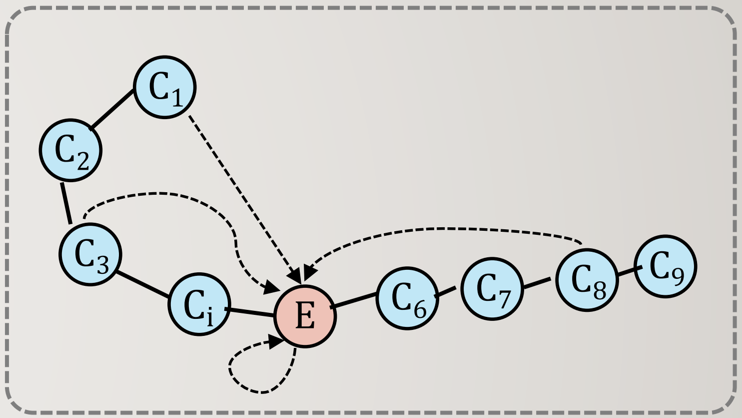
https://conceptnet.io/

Clause Graph Construction

1. "When I found that my advice had been adopted
2. and the department leader replied to my mail
3. I realized that I had made contributions to the country's development".
4. Bai Jinyue, an ordinary worker in XingTai Steel factory in HeBei province,
5. talked to the journalist with excitement

Happiness

6. with a Thank-You letter in his hands.
7. Since 27 years ago different departments,
8. such as the public security,
9. and the Committee on Science and Technology, have accepted his advice.



: Clause emotion label

: Sequence edge

: Knowledge edge

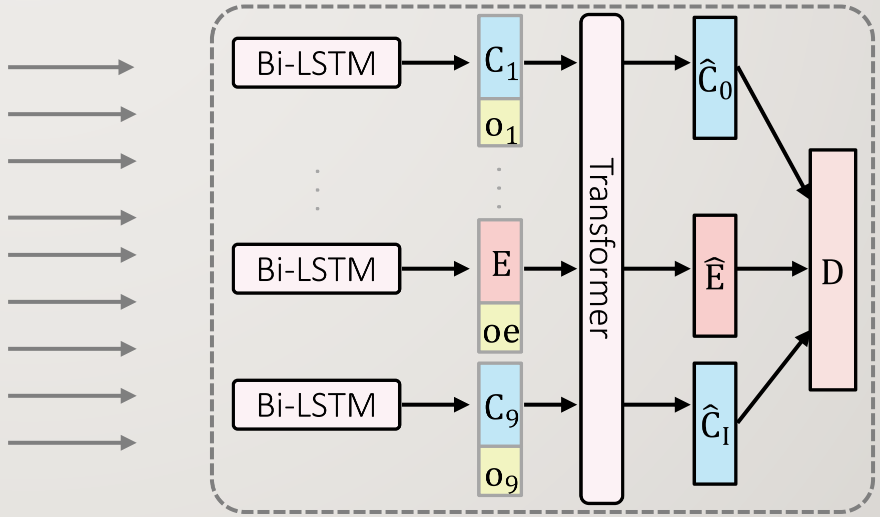
(a) Document Encoding

- Each clause is first encoded using Bi-LSTM → clause representation C_i.
- The clause sequence is fed to a Transformer to capture the sequential information (*S-Edge*).
- The clause position information (O_i) is used to enrich the clause representation.
- The emotion clause embedding is used as the query, and the candidate clause representations C_i as both the key and value vectors, to derive the document representation D

News Corpus

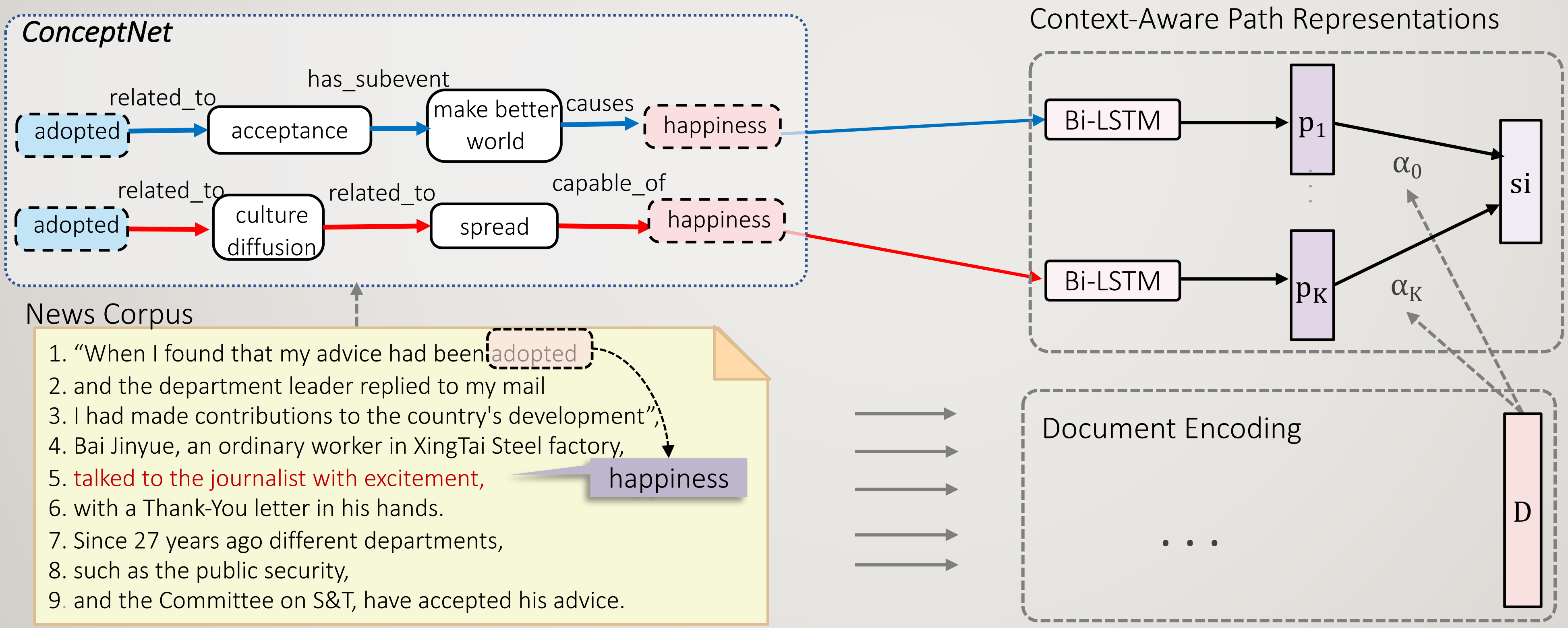
1. "When I found that my advice had been adopted
2. and the department leader replied to my mail
3. I had made contributions to the country's development",
4. Bai Jinyue, an ordinary worker in XingTai Steel factory,
5. talked to the journalist with excitement,
6. with a Thank-You letter in his hands.
7. Since 27 years ago different departments,
8. such as the public security,
9. and the Committee on S&T, have accepted his advice.

happiness



(b) Context-Aware Path Representation Learning

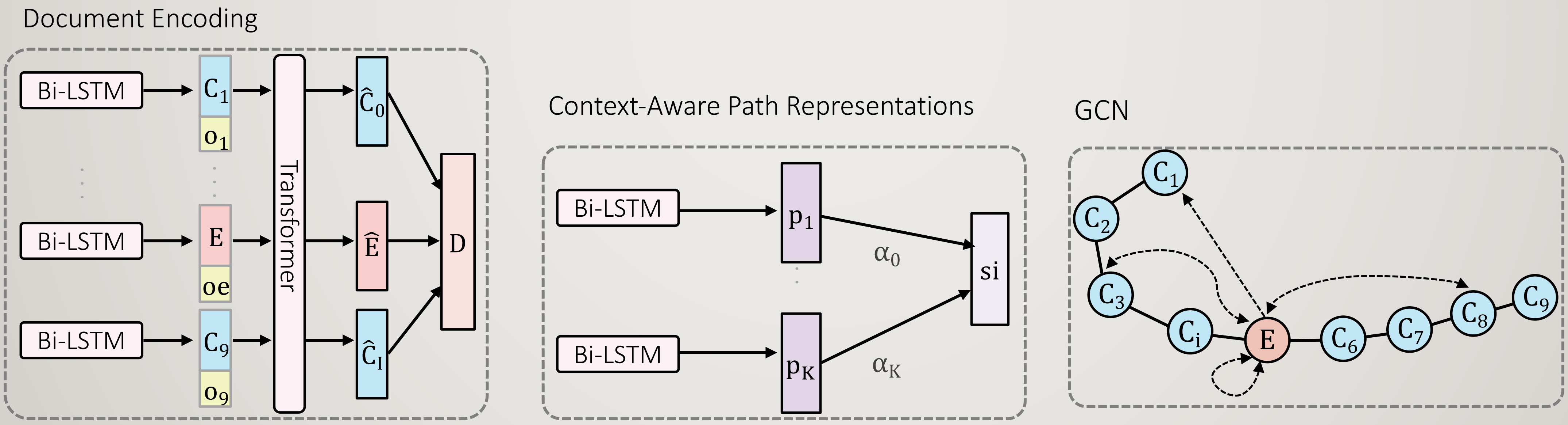
- A maximum K paths between a candidate clause and the emotion clause
- Weigh each path based on their similarity with the document context
- Aggregate the K knowledge paths to derive the context-aware path representation $\boldsymbol{s}_i$



$$\boldsymbol{s}_i = \sum_{i=1}^{K} \alpha_i p_i$$

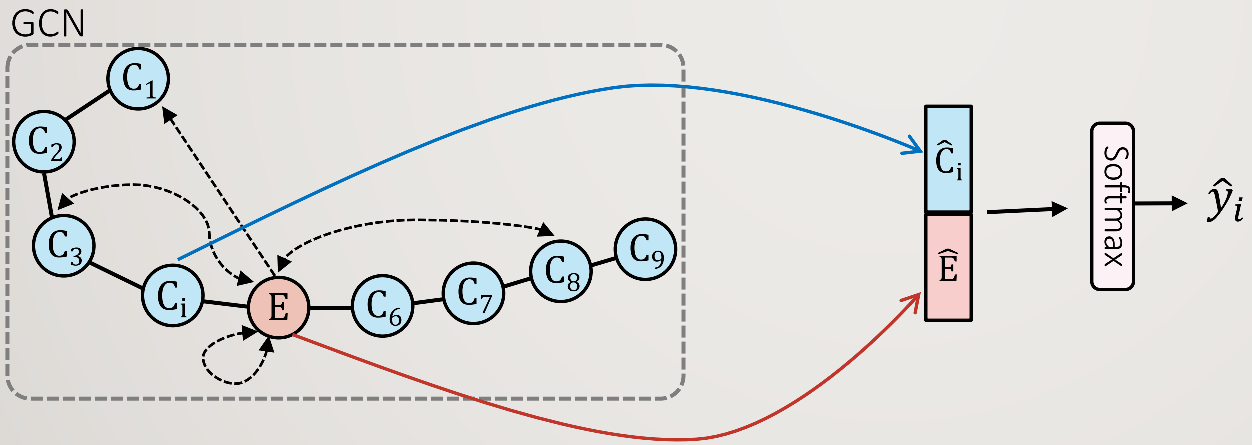
(c) GCN for Clause Representation Update

- Initialize the node (or clause) in the clause graph with $\hat{C}_i$ and the extracted knowledge path with $\boldsymbol{s}_i$
- Update clause representation using Relation-GCNs, designed for information aggregation over multiple different edges

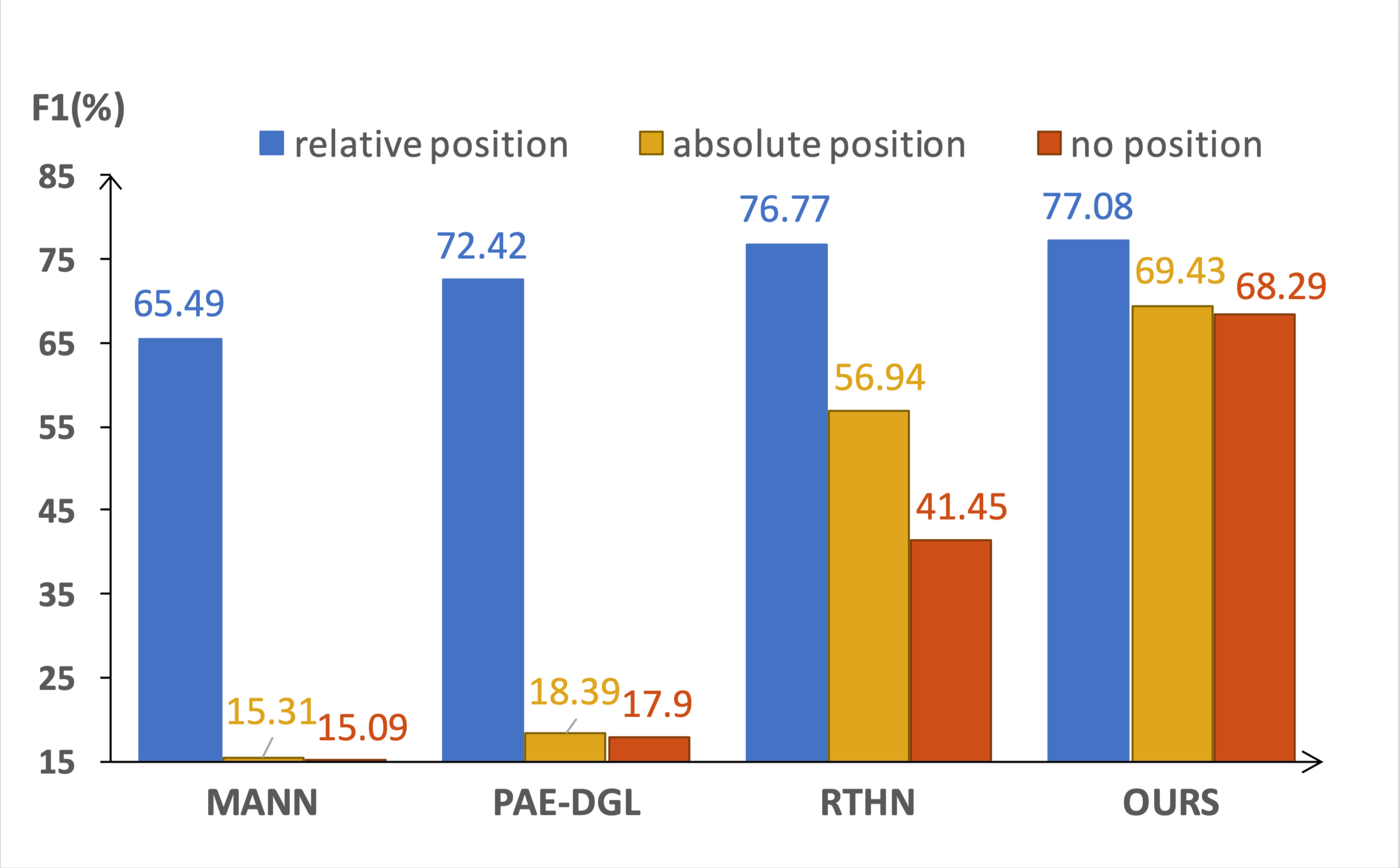


(d) Cause Clause Extraction

- The candidate clause node representation and the emotion node representation are concatenated and fed to a softmax layer to yield the predictive class distribution $\hat{y}_i$.



Impact of Position Information



Yan et al. [Position Bias Mitigation: A Knowledge-Aware Graph Model for Emotion Cause Extraction](). ACL 2021.

Open Challenges

- Multiple emotion-cause pairs extraction from document
- Dealing with implicit emotions
- Better ways in acquiring and incorporating commonsense knowledge
- Need a large-scale annotated dataset[1]

[1]RECCON dataset contains 10k utterance cause-effect pairs, Poria et al. [Recognizing Emotion Cause in Conversations](), 2020.

BREAK

Outline

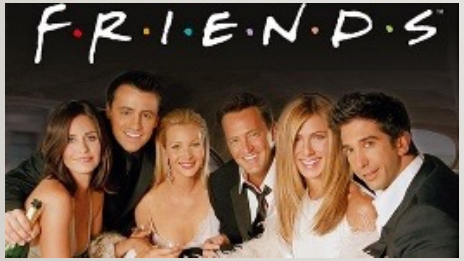
- Aspect-Based Sentiment Analysis
- Sentiment-Topic Extraction
- Emotion Cause Detection
- Dialogue Emotion Detection
- Review Question-Answering
- Interpretability in Sentiment Analysis

Dialogue Emotion Detection

Daily dialogues

Social media platforms

TV sitcoms



Emotion Detection

… Happiness

Surprise …

… ?

Emotion Prediction

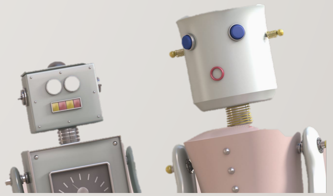
… Happiness

Surprise …

Happiness or Neutral ?

Counseling services

Chatbots

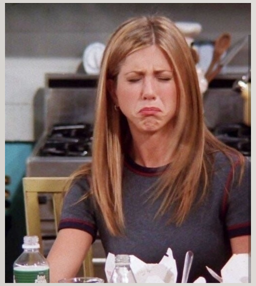


Dialogue Emotion Detection

Need to consider historical dialogue context

Maybe this weekend we could go to the beach?

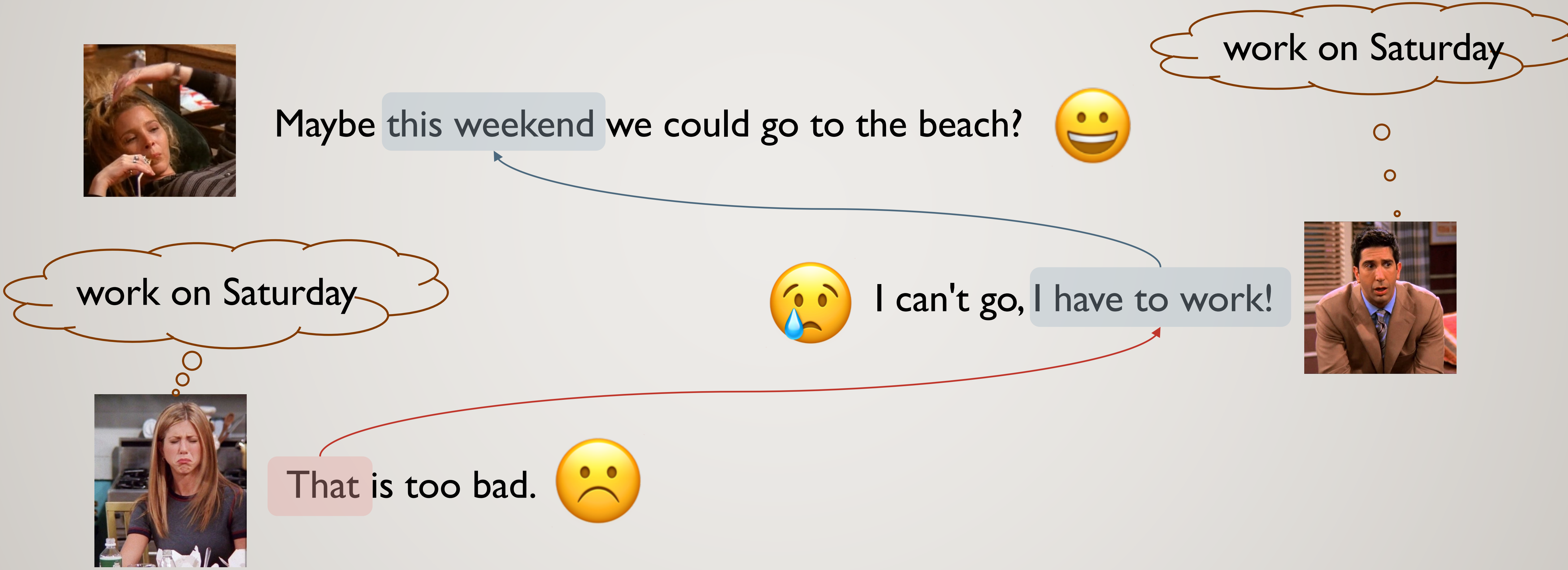
I can't go, I have to work!



That is too bad.

Dialogue Emotion Detection

Commonsense knowledge is required for inferring emotion.

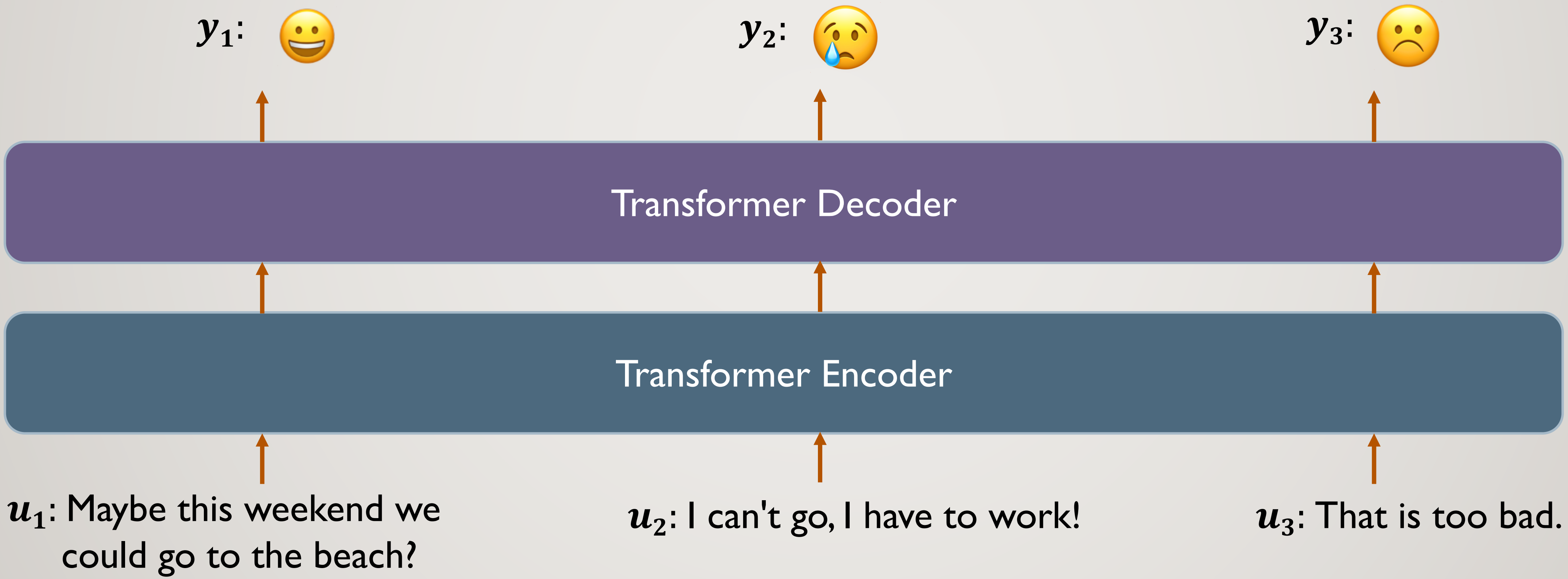


Knowledge-Enriched Transformer (KET)

- By leveraging an external knowledge base, ConceptNet, the meaning of "***friends***" in the last utterance is enriched by associated knowledge entities, "*socialize*", "*party*", and "*movie*".
- The implicit "*Happiness*" emotion in the last utterance can be inferred more easily via its enriched meaning.

Capture the Context

- Framing dialogue emotion detection as a sequence labelling problem



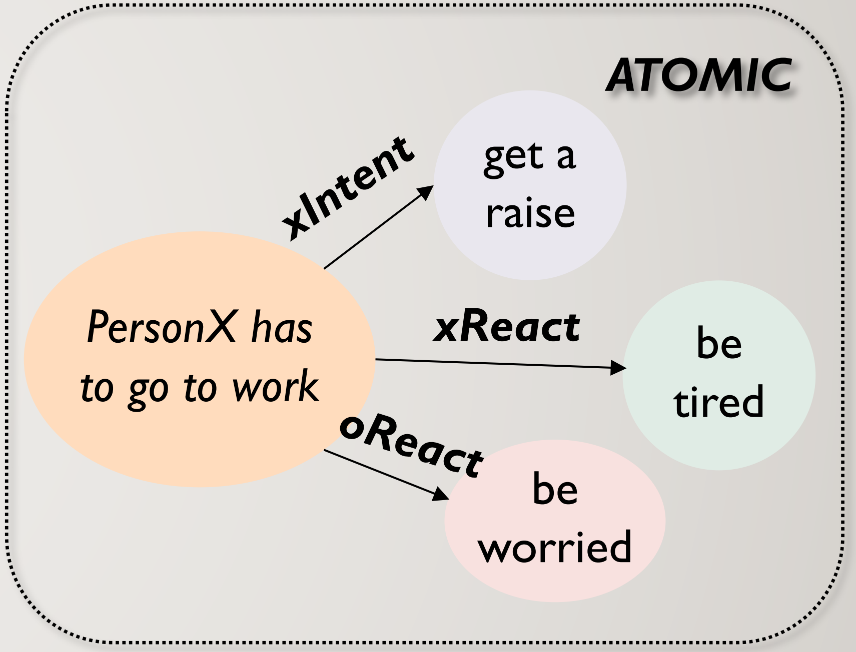
Zhu et al., Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection, ACL 2021.

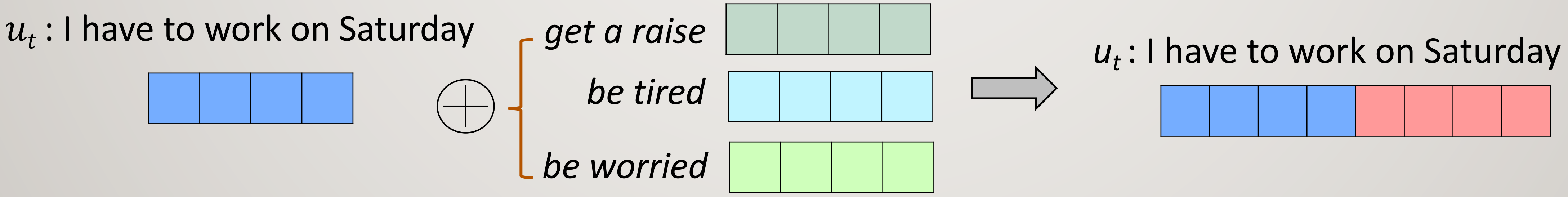
Enrich utterances with knowledge

Route 1: Knowledge retrieval from a commonsense knowledge base – ATOMIC

$\boldsymbol{u_t}$**:** I have to work on Saturday

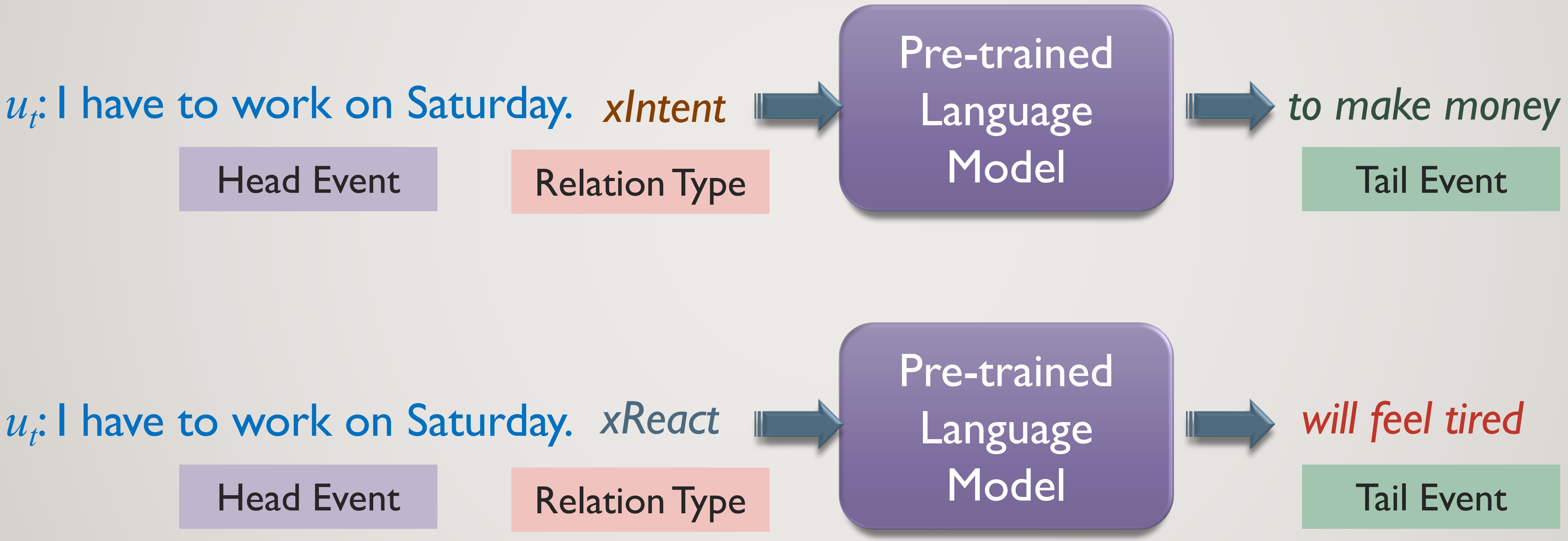
PersonX has to go to work	*xIntent*	*get a raise*
PersonX has to go to work	*xReact*	*be tired*
PersonX has to go to work	*oReact*	*be worried*





Enrich utterance with knowledge

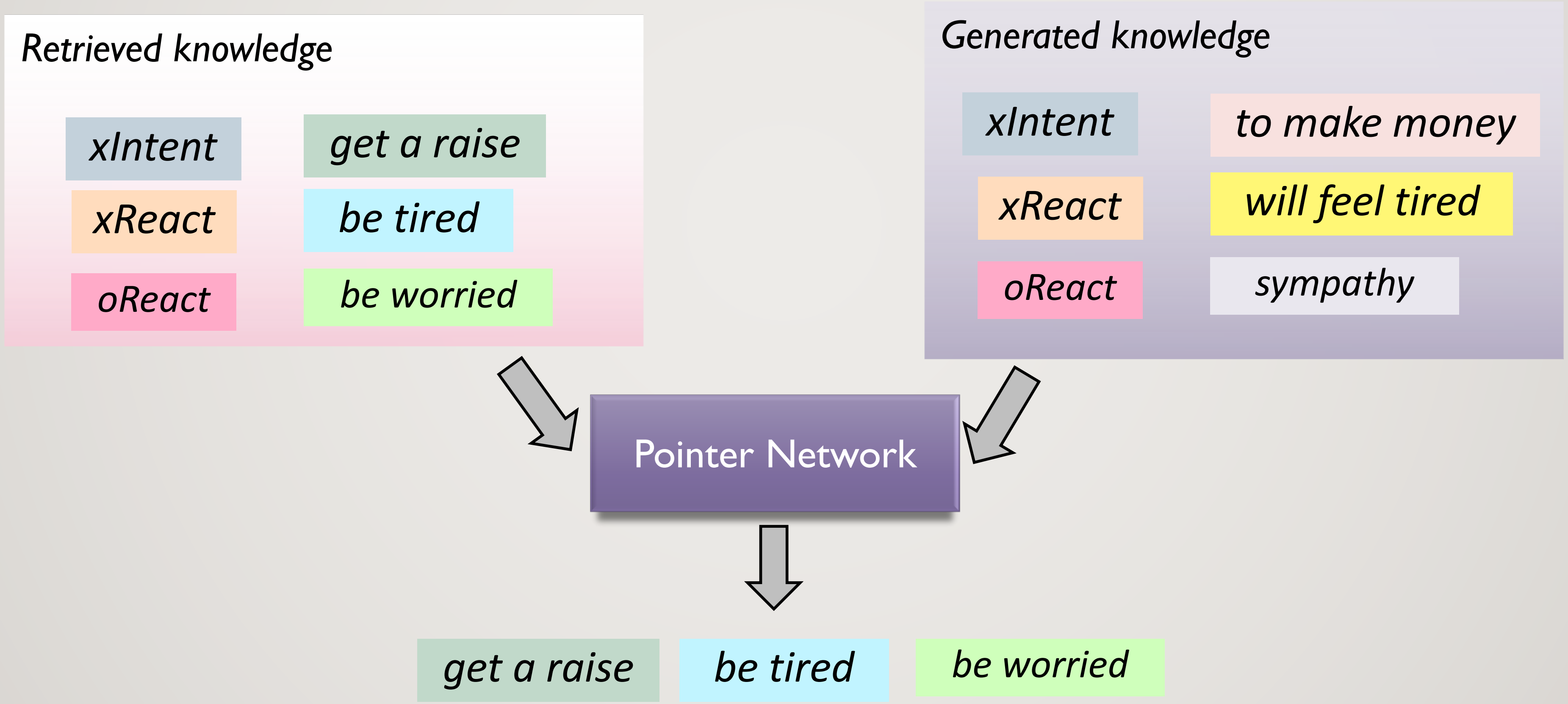
Route 2: knowledge generation using a pre-trained language model on ATOMIC



Bosselut et al., *COMET: Commonsense Transformers for Automatic Knowledge Graph Construction*, ACL 2019

Knowledge Fusion

- Use a pointer network to select knowledge obtained from different routes



Zhu et al., [Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection](), ACL 2021.

Incorporation of Commonsense Knowledge

A: Alright, go on.
B: OK, I have to sleep on the west side because I grew up in California and otherwise the ocean would be on the wrong side.
A: Oh, my god, you are a freak.

→ Mad ✗

Incorporation of Commonsense Knowledge

- Effect of attention mechanism in commonsense knowledge retrieval

A: Alright, go on.
B: OK, I have to sleep on the west side because I grew up in California and otherwise the ocean would be on the wrong side.
A: Oh, my god, you are a freak.

→ Joyful ✓

Commonsense knowledge:

A wants to be liked.
A wants to be accepted.
A wants to be a freak.
A will feel satisfied.
A will feel ashamed.
A will feel happy.
B will feel impressed.
B will feel disgusted.
B will feel surprised.

Consideration of Topic Information

A: Johnny died yesterday, we knew that it was coming, but…

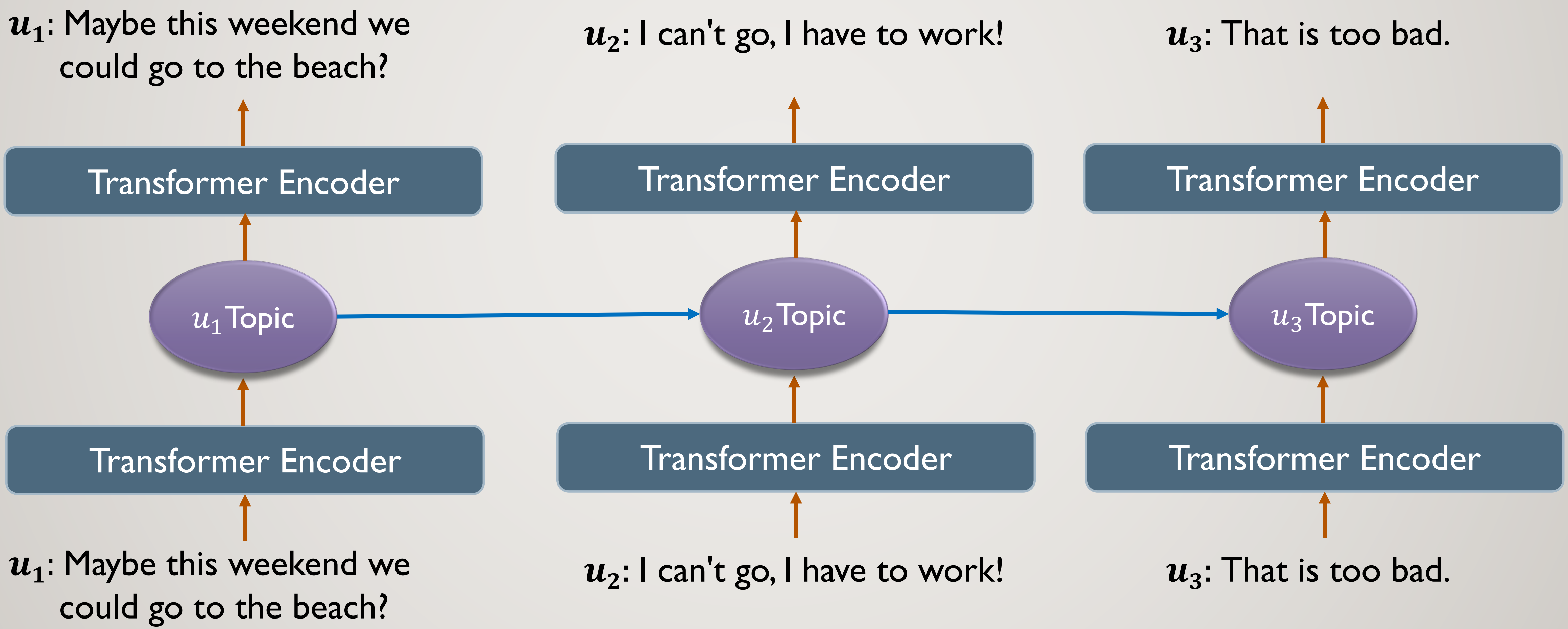
Like just last week, he was doing so well. :B

A: Then all of a sudden they gave him a microphone, he asked me to marry him, like onstage.

He was doing so well. :B

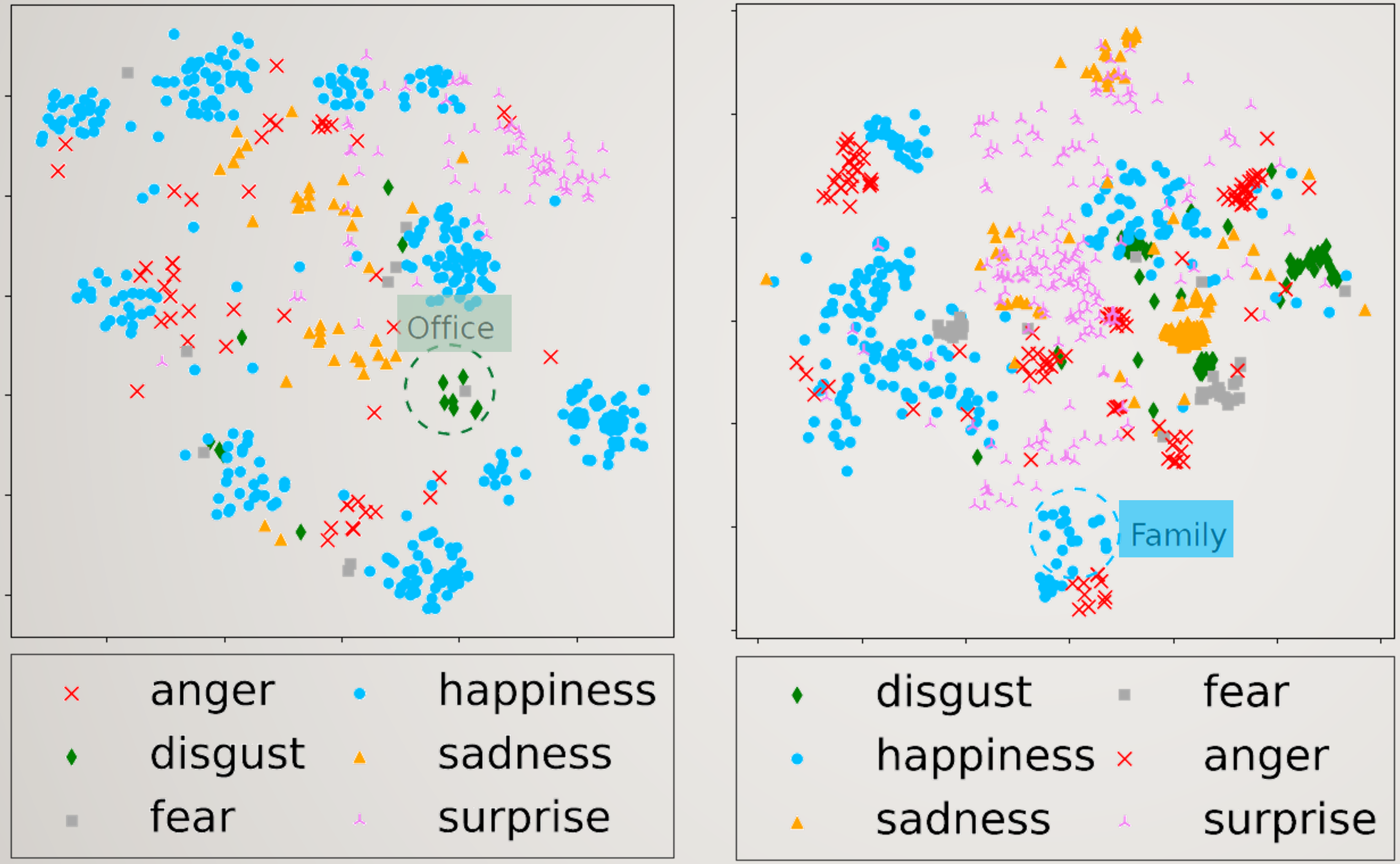
Topic-Driven and Knowledge-Aware Transformer (TODKAT)

- Learn Topic Representation



Zhu et al., Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection, ACL 2021.

Concordance between Topics and Emotion



(a) DailyDialog (b) MELD

- T-SNE visualization of topic representations of utterances from DailyDialog and MELD.
- Colours indicate the ground-truth emotion labels
- Utterances are clustered into polarised groups.

Zhu et al., [Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection](), ACL 2021.

Concordance between Topics and Emotion

Topic	Utterances	Emotion
Office	A: How are you doing, Christopher? B: To be honest, I'm really fed up with work at the moment. I need a break! A: Are you doing anything this weekend? B: I have to work on Saturday all day! I really hate my job!	disgust
Family	A: Yeah, I-I heard. I think it's great! Ohh, I'm so happy for you! B: I can't believe you're getting married! C: Yeah. D: Monica and Rachel made out.	happy

(c) Representative utterances and their topics

- Each cluster is exemplified by a group of utterances.

Zhu et al., Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection, ACL 2021.

Open Challenges

Multi-label dialogue emotion detection

Different emotions perceived by speakers/listeners (self vs. inter-personal dependencies)

Modelling topics, personalities and situational context

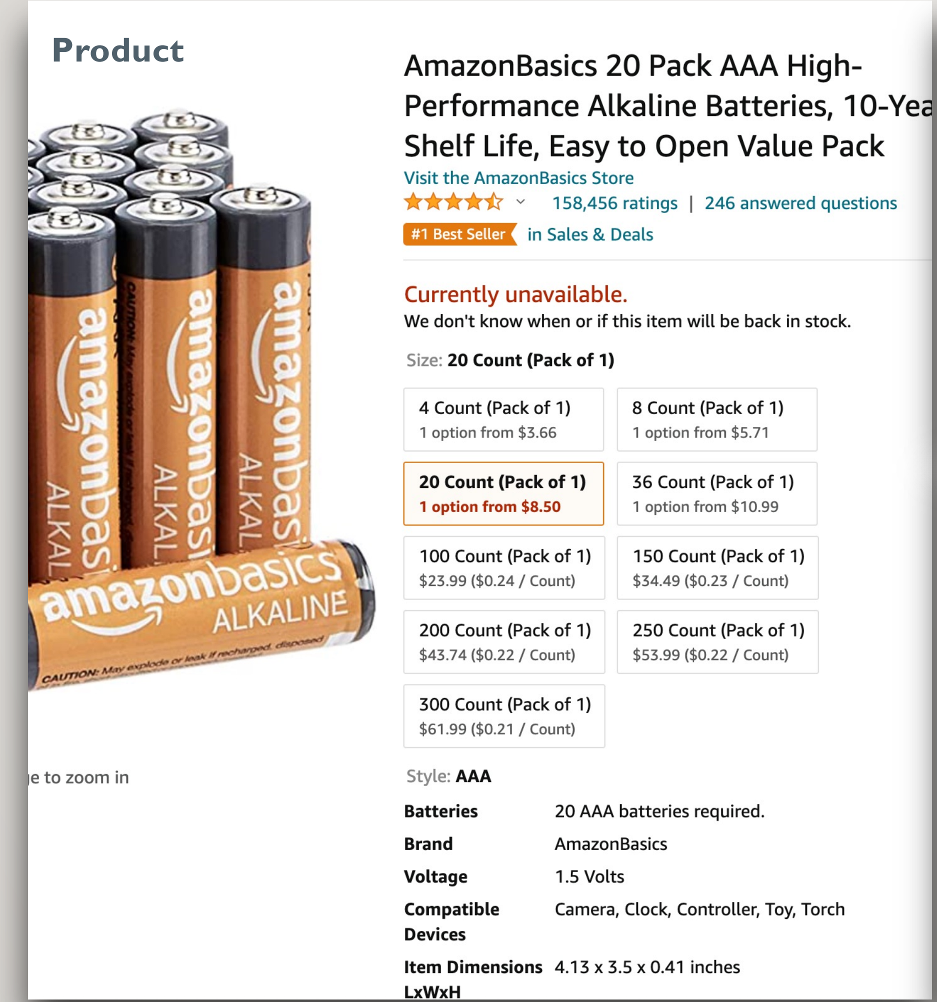
Multimodal emotion detection – also considering images, facial expressions and/or speech

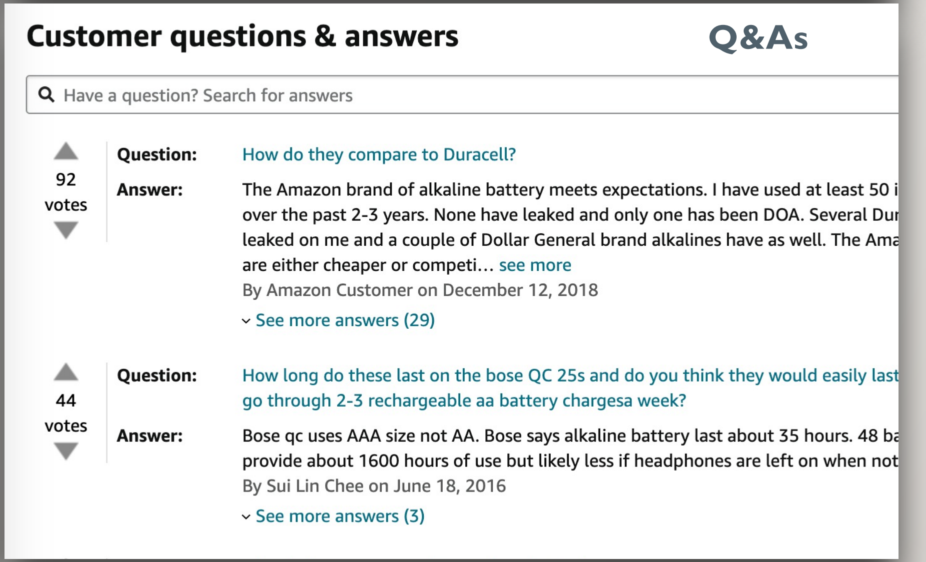
Outline

- Aspect-Based Sentiment Analysis
- Sentiment-Topic Extraction
- Emotion Cause Detection
- Dialogue Emotion Detection
- Review Question-Answering
- Interpretability in Sentiment Analysis

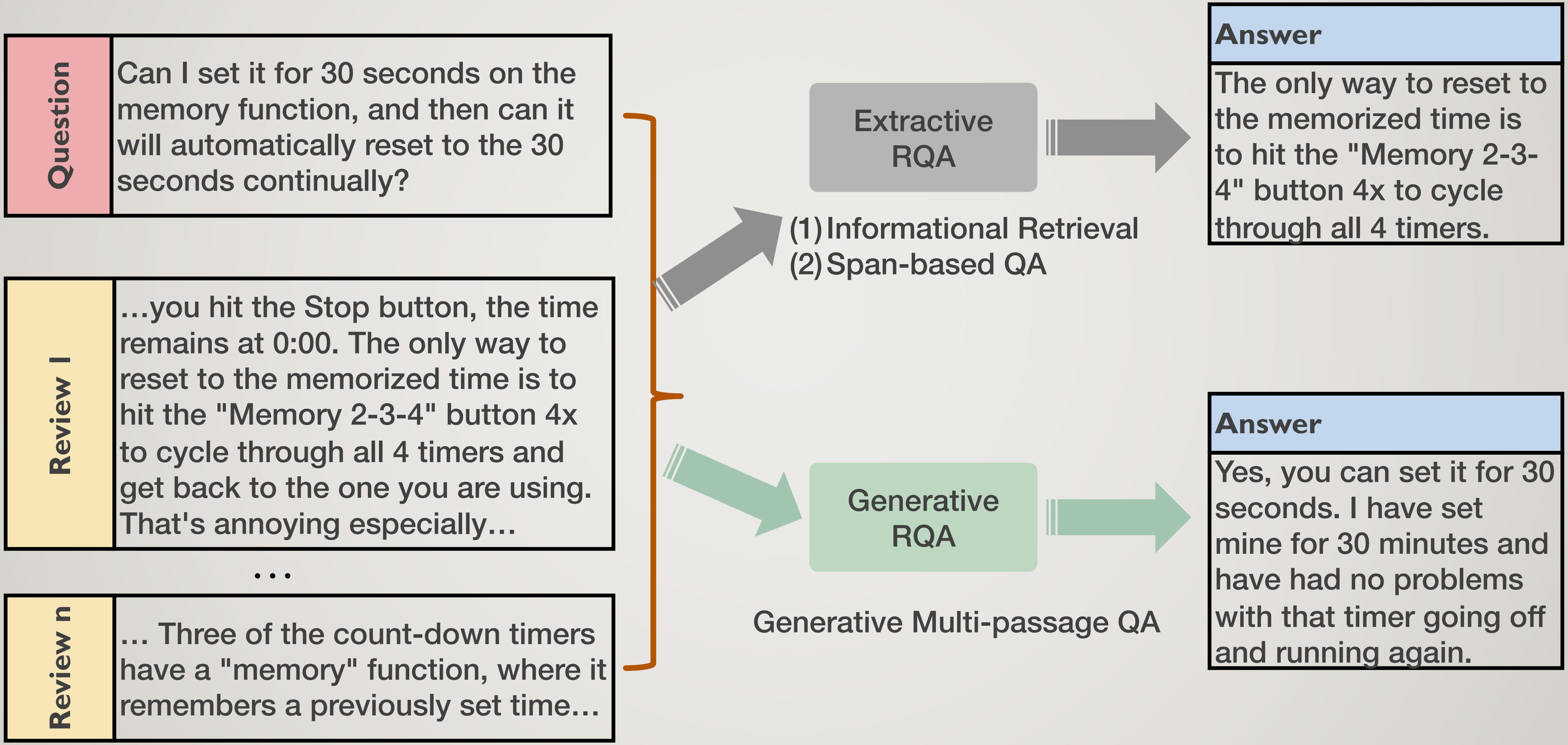
Review Question-Answering (RQA)

- To answer questions based on subjective personal experiences or opinions towards certain products / services



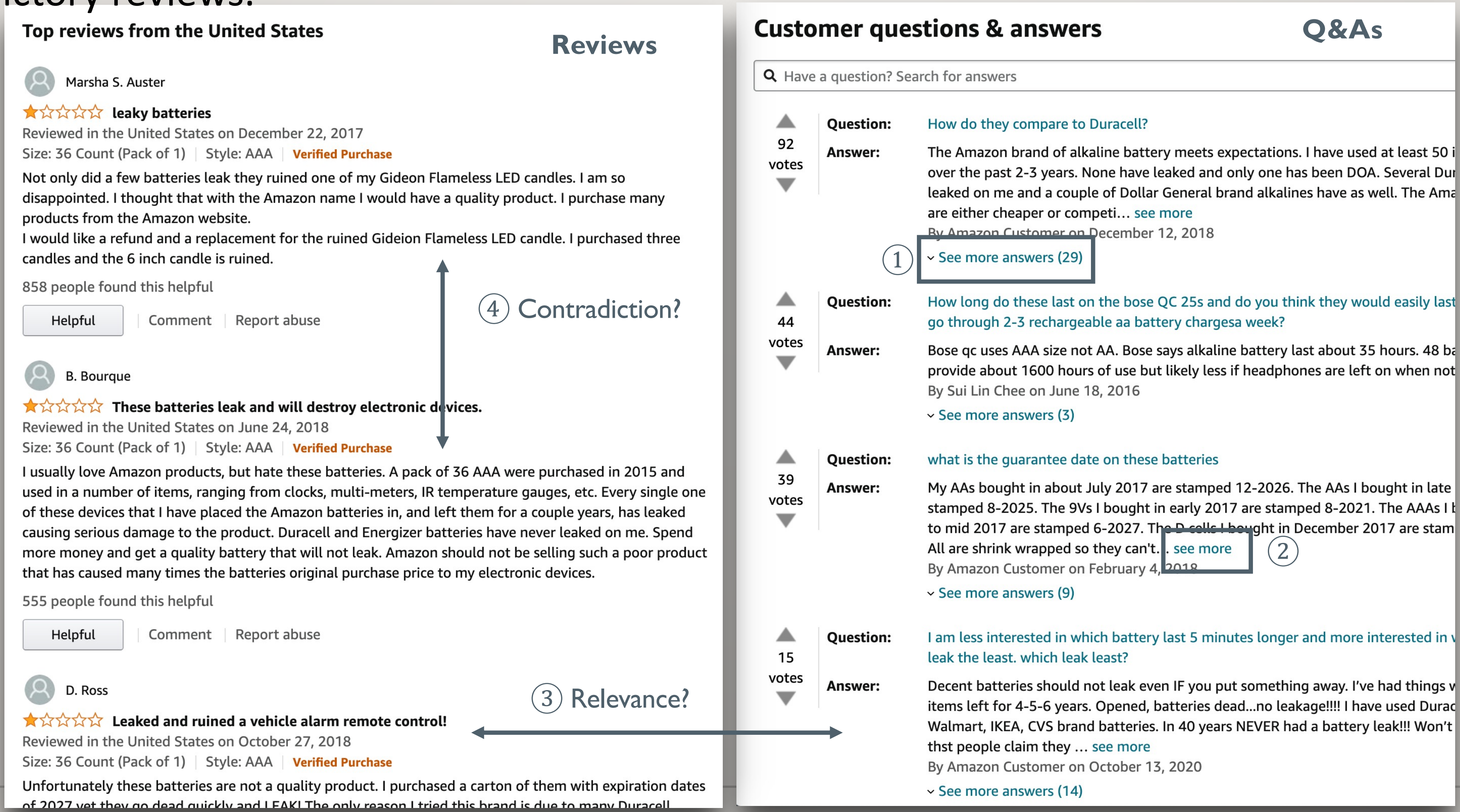


Extractive RQA vs. Generative RQA



Challenges of Review QA

(1) Multiple ground-truths; (2) Long text length; (3) Lack of explicit relevance annotations; (4) Contradictory reviews.



Extractive QA: Fine-tuning BERT

- Label a 2.5k span-based review QA dataset
- Jointly fine-tuning BERT for span-based review QA, aspect extraction and aspect sentiment classification

Questions

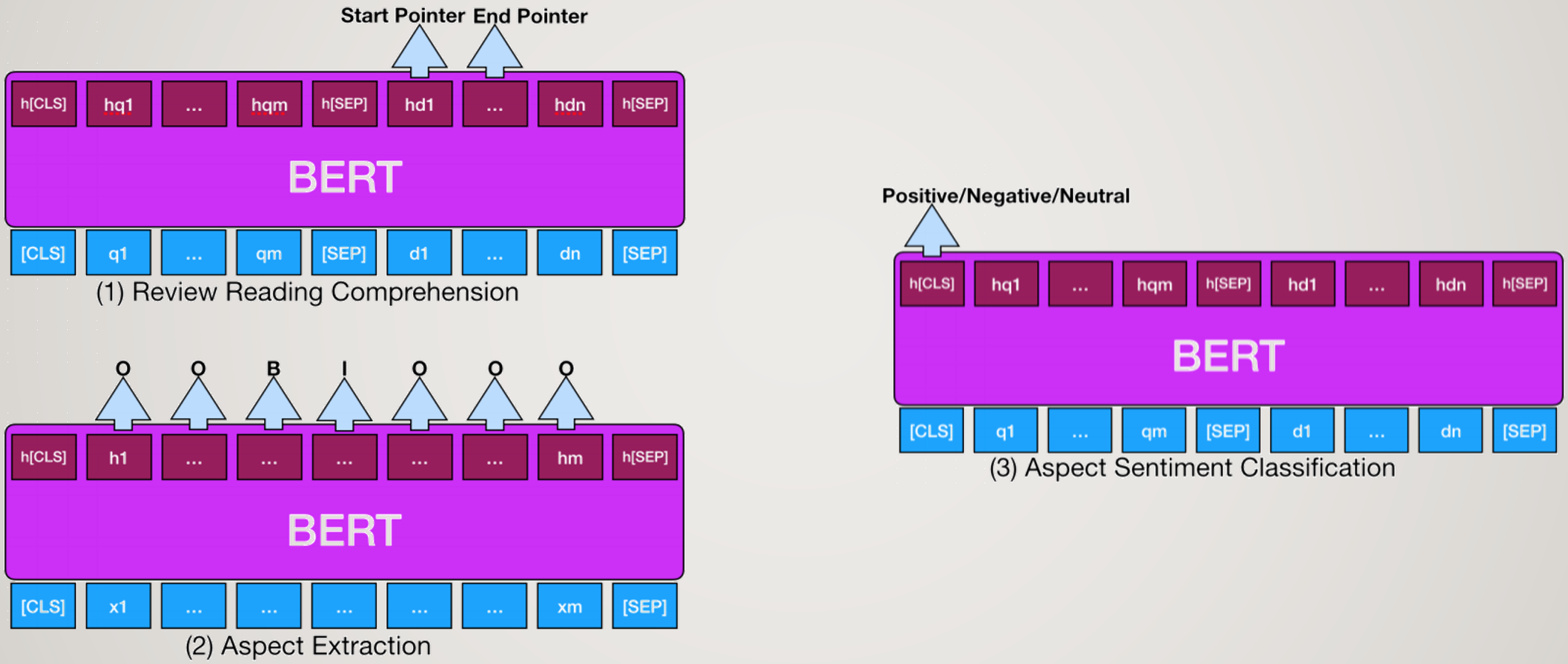
Q1: Does it have an internal hard drive ?
Q2: How large is the internal hard drive ?
Q3: is the capacity of the internal hard drive OK ?

Review

Excellent value and a must buy for someone looking for a Macbook . You ca n't get any better than this price and it **come with**$_{A1}$ an internal disk drive . All the newer MacBooks do not . Plus you get **500GB**$_{A2}$ which is also a **great**$_{A3}$ feature . Also , the resale value on this will keep . I highly recommend you get one before they are gone .

Dataset	Num. of Questions	Num. of Reviews
Laptop Training	1015	443
Laptop Testing	351	79
Restaurant Training	799	347
Restaurant Testing	431	90

Extractive QA: Fine-tuning BERT



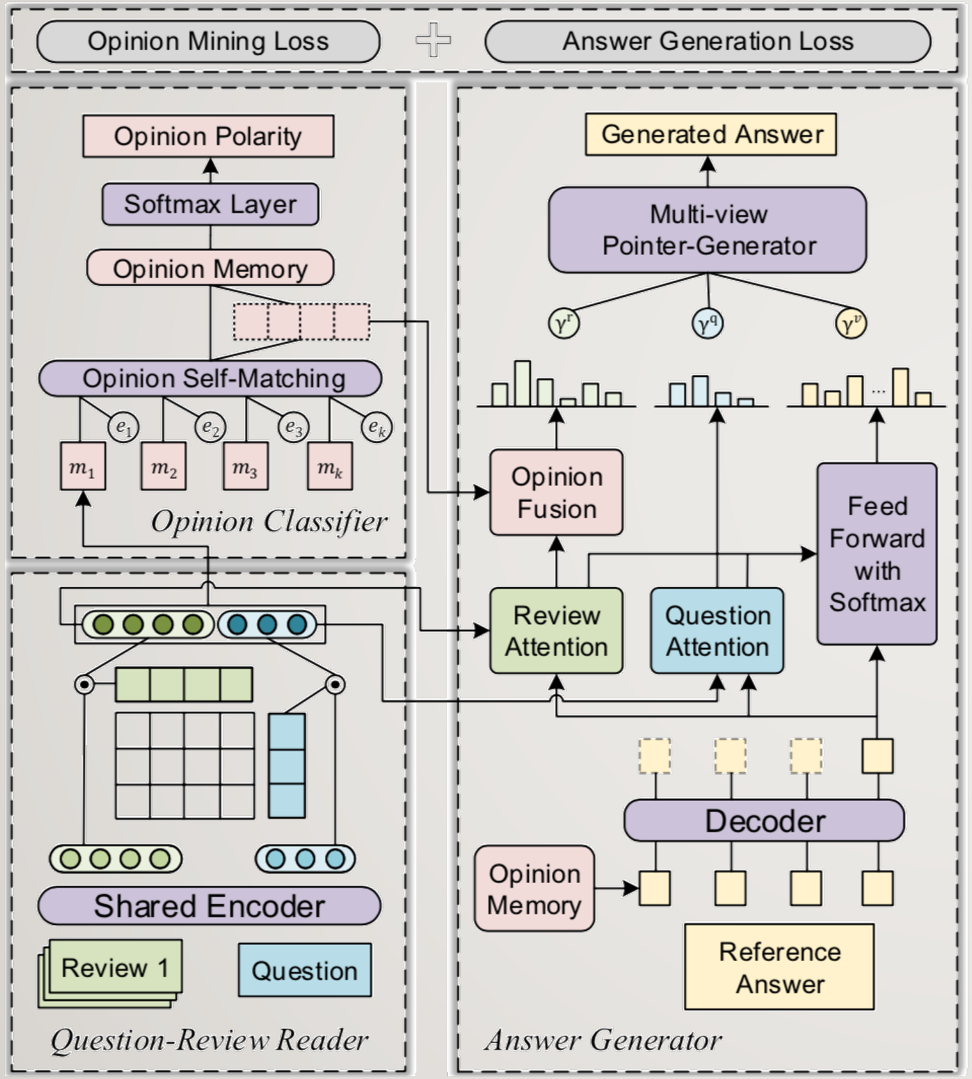
Generative QA: Opinion-Aware Answer Generation (OAAG)

Question	Is this really a good buy for cycling?	Does this device play Blu-ray on PC?
Reference Answer	Yes it is. It's pretty well padded.	Yes. It comes with software to play Blu-ray discs on a PC.
Relevant Review Snippets (Partial)	1. This seat is comfortable and works well. I mentioned to my sister, about wanting to buy a new seat, and she suggested buying a gel cover. 2. This cushion is high quality and very comfy. This one holds up even if it gets wet. 3. Not perfect in my opinion, could have used a bit more padding.	1. No complaints one way or the other, the device works as expected and allowed me to view Blu-ray disks on a PC that didn't have a Blu-ray device. 2. I have only used it on a XBMC and Win 7 PC for movies since I don't own a player for Mac. 3. Again, it does not play discs, so what did I just buy?

Approach	**Answer**	
Opinion-based	Yes	Yes
Retrieval-based	This cushion is high quality and very comfy. This one holds up even if it gets wet.	No complaints one way or the other, the device works as expected and allowed me to view Blu-ray disks on a PC that didn't have a Blu-ray device.
Generation-based	I don't think so. I don't think it would be too big for cycling.	I don't see why it wouldn't work with the Blu-ray player, but it does have an HDMI input.
Opinion-aware Generation	Yes, it is a very good seat. I have been using it for several months now and have not had any problems.	Yes, it does work with the Blu-ray player. I havenâĂŹt had any problems with them at all.

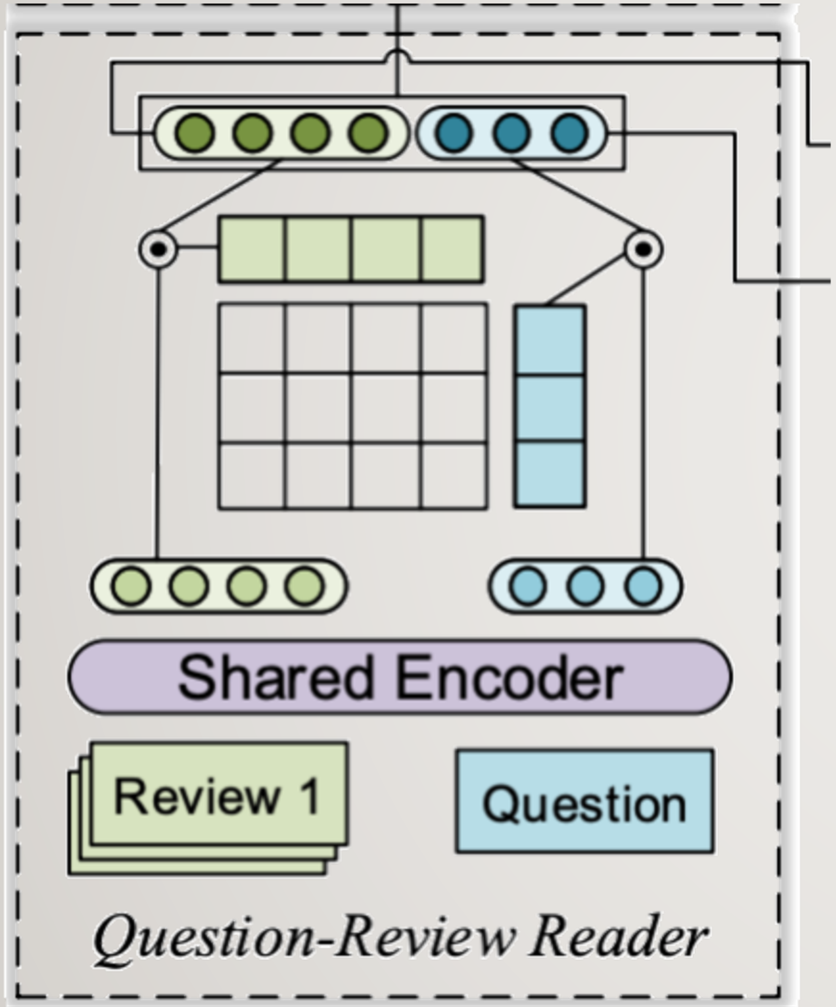
Deng et al. [Opinion-aware Answer Generation for Review-driven Question Answering in E-Commerce](). CIKM 2020.

Generative QA: Opinion-Aware Answer Generation (OAAG)



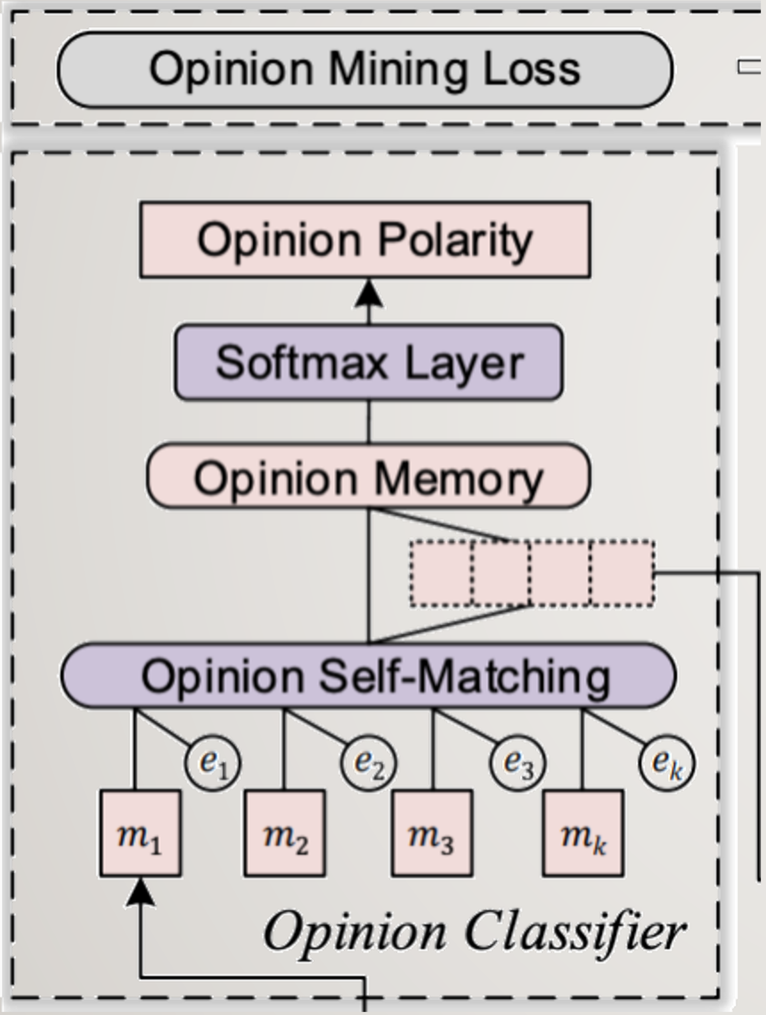
- Question-Review Reader
- Opinion Classifier
- Answer Generator

OAAG – Question-Review Reader



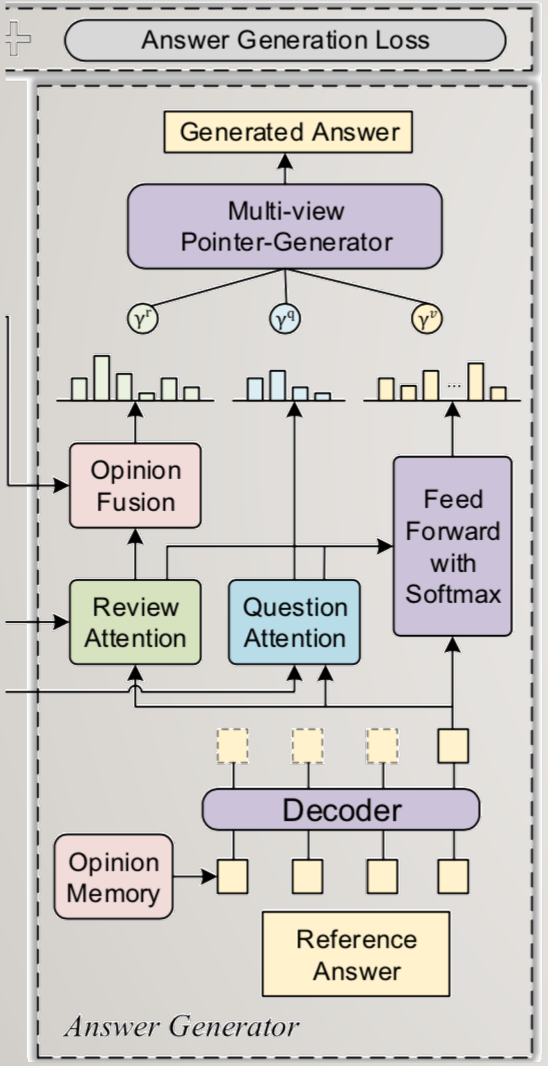
- Use Bi-LSTM to encode each review and question separately
- Compute the co-attention between the question and review representations to get review-aware question vector and question-aware review vector
- Concatenate final vectors to get question-review matching representation

OAAG – Opinion Classifier



- Concatenate *question-review matching representation* m_k with one-hot corresponding *review rating vector* e_k (e.g. 1 to 5 stars) to generate matching vectors
- Apply a vanilla attention over the resulting vectors to get review-level opinion attention weights
  - To determine the common opinion among all the reviews
- Derive the *opinion memory representation* by the dot product of the matching vectors and the opinion attention weights
- Add a Softmax layer to classify the common opinion polarity of the reviews given the question

OAAG – Answer Generator



- *Opinion memory representation* is used as the initialized state of the Unidirectional LSTM decoder

- A multi-view pointer-generator network are applied on the top of decoder to generate final answer. The views includes:
  - Probability distributions over the fixed vocabulary, using review attention and question attention from the question-review encoder;
  - Probability distributions over *extra question tokens*, using only question attention;
  - Probability distribution over *extra review tokens*, using fused opinion-aware review attention

Deng et al. [Opinion-aware Answer Generation for Review-driven Question Answering in E-Commerce](). CIKM 2020.

OAAG: Example Outputs

Question: Are the leg height adjustments easy to manipulate and change? (*Electronics*)
Reference Answer: Yes they are. Very easy and they stay where you set them. (*positive*)

OAAG: Yes, they are very easy to set up. I have used them for several years and haven't had any issues with them.

Question: Can these be used in the microwave? (*Home&Kitchen*)
Reference Answer: Yes, they can. I'm not much of a microwave user, but have used them there on occasion. (*positive*)

OAAG: Yes, they can be used on any kind of microwave, but I don't think they would withstand high temperatures. I haven't had any problems with them at all.

Deng et al. [Opinion-aware Answer Generation for Review-driven Question Answering in E-Commerce](). CIKM 2020.

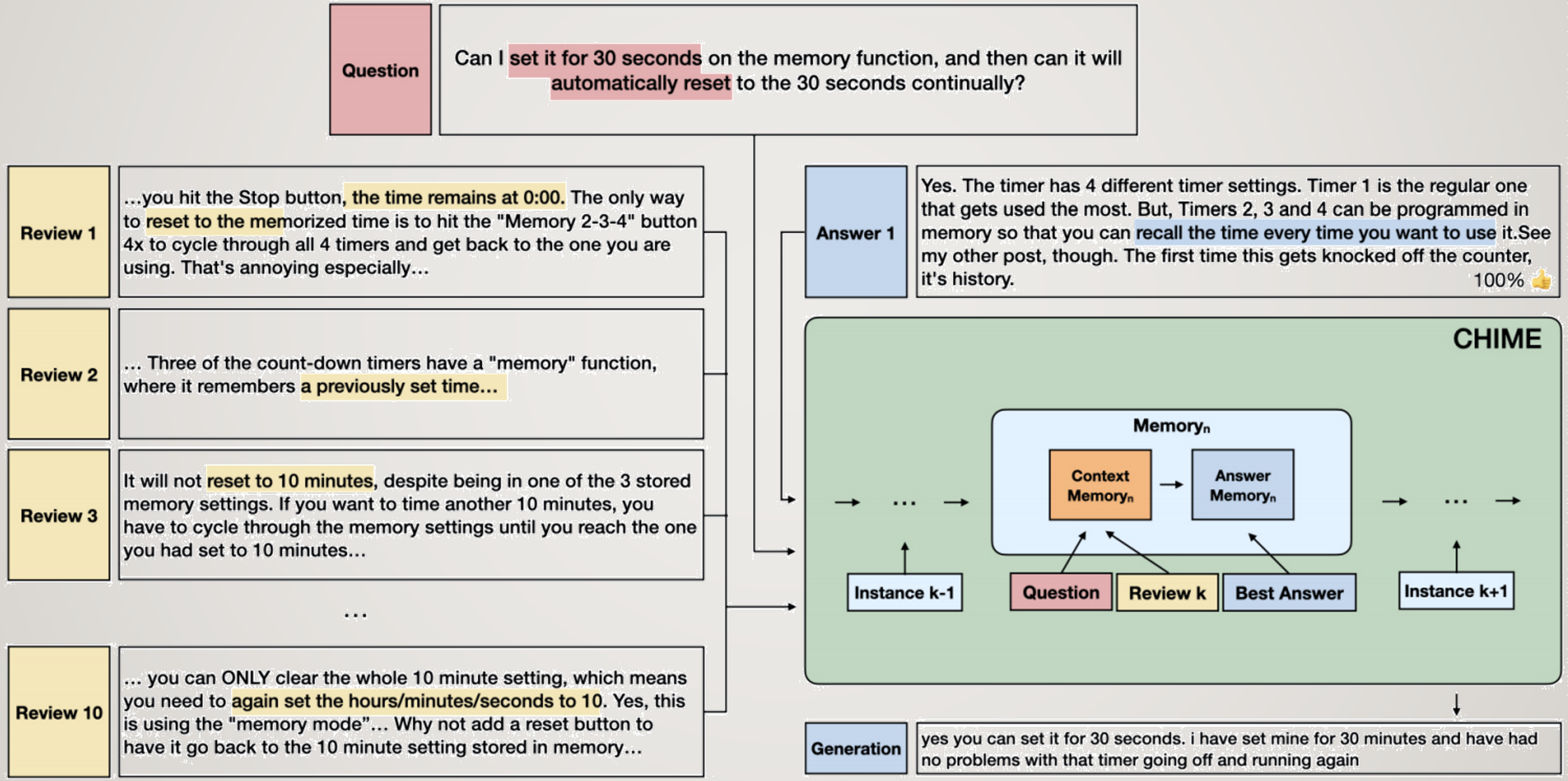
OAAG Limitations

- OAAG used Unidirectional LSTM to encode question and reviews.
  - Each review text is chunked into snippets of length 50, or to the end of a sentence boundary.
  - For a given question, BM25 is used to rank all the review snippets of the corresponding product and top 10 relevant review snippets are used as the model input.

- To use a more effective transformer architecture, it is impossible to concatenate all reviews.
  - E.g., BERT can only take input with a maximum of 512 tokens (or 1024 for BERT-large).

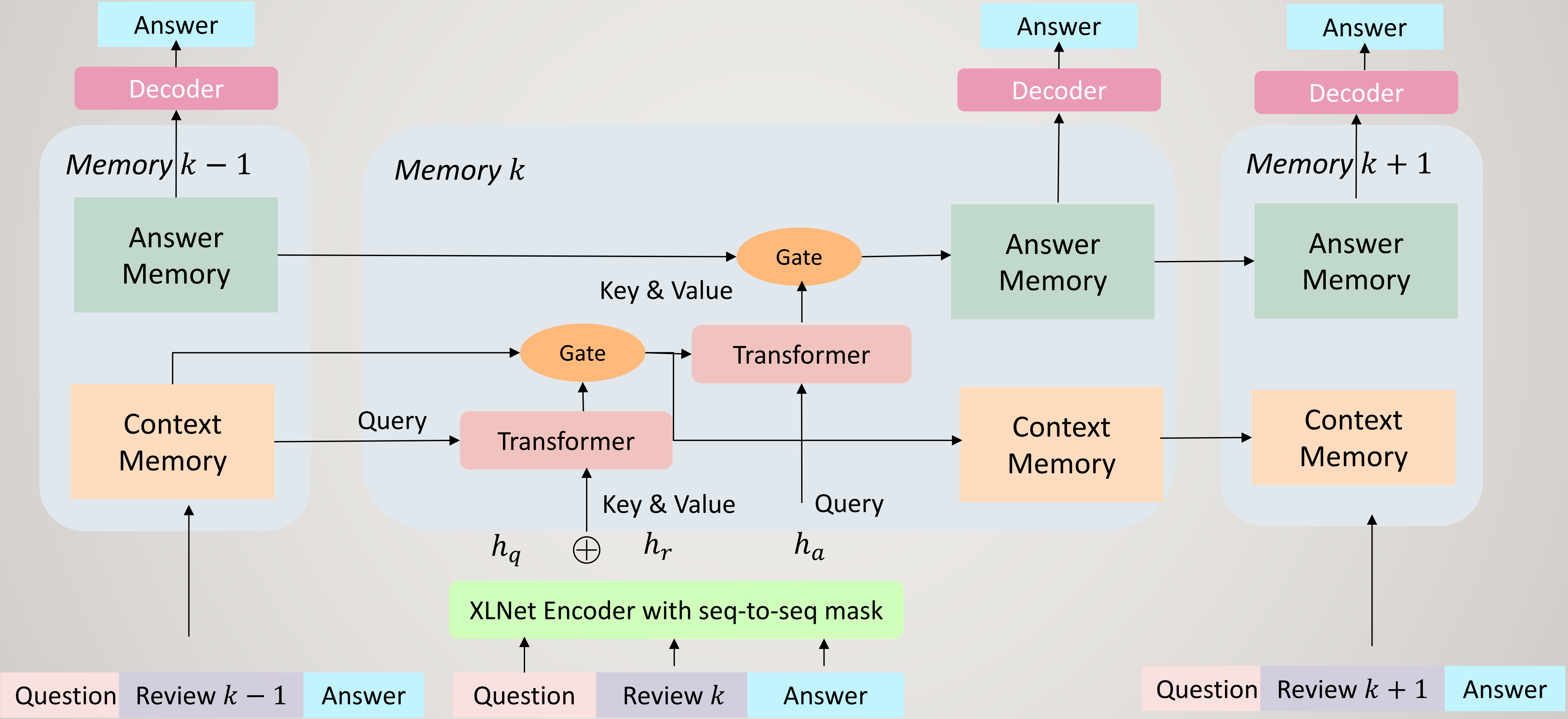
- An alterative approach: Cross-passage HIerarchical Memory nEtwork (CHIME).
  - It extends XLNet introducing an auxiliary memory module consisting of two components:
    - **Context memory** collects cross-passage evidence;
    - **Answer memory** working as a buffer continually refining the generated answers.

Generative QA: Cross-passage HIerarchical Memory nEtwork (CHIME)



Lu et al. CHIME: Cross-passage Hierarchical Memory Network for Generative Review Question Answering. COLLING 2020.

Generative QA: CHIME

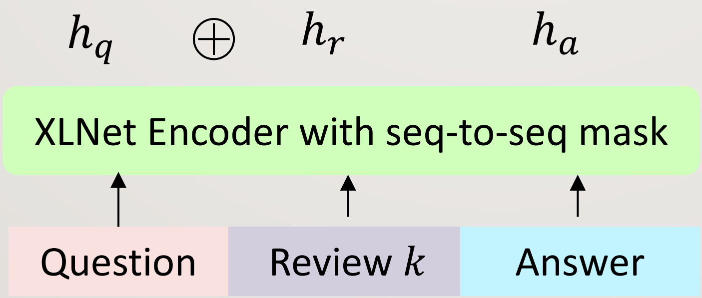


Lu et al. CHIME: Cross-passage Hierarchical Memory Network for Generative Review Question Answering. COLLING 2020.

Generative QA: CHIME

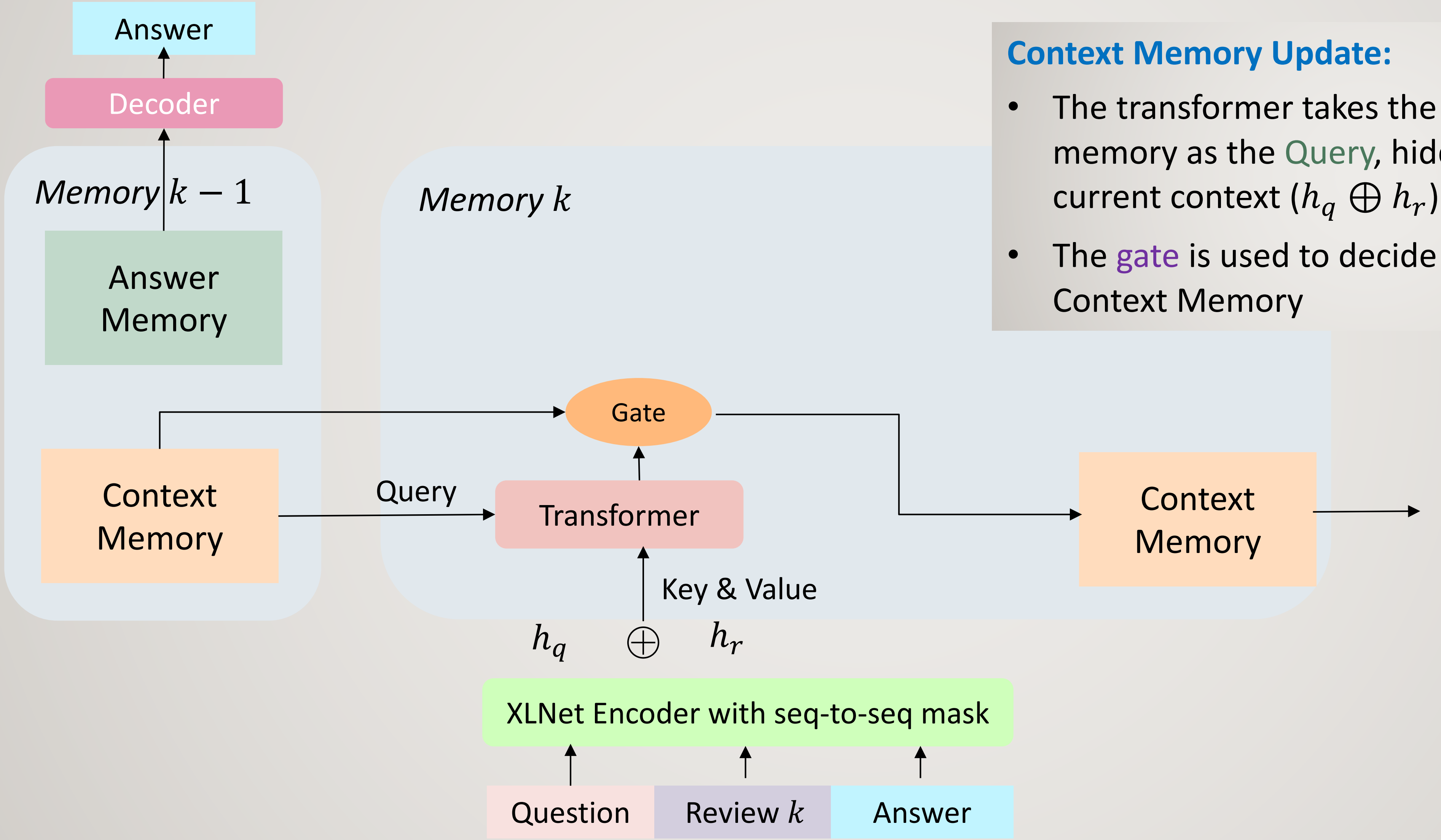
Encoder:

- A training instance contains (question, kth review, Answer)
- XLNet is fine-tuned on the training set for answer generation (with seq-to-seq mask)
- The resulting XLNet is used as the input encoder



Lu et al. CHIME: Cross-passage Hierarchical Memory Network for Generative Review Question Answering. COLLING 2020.

Generative QA: CHIME

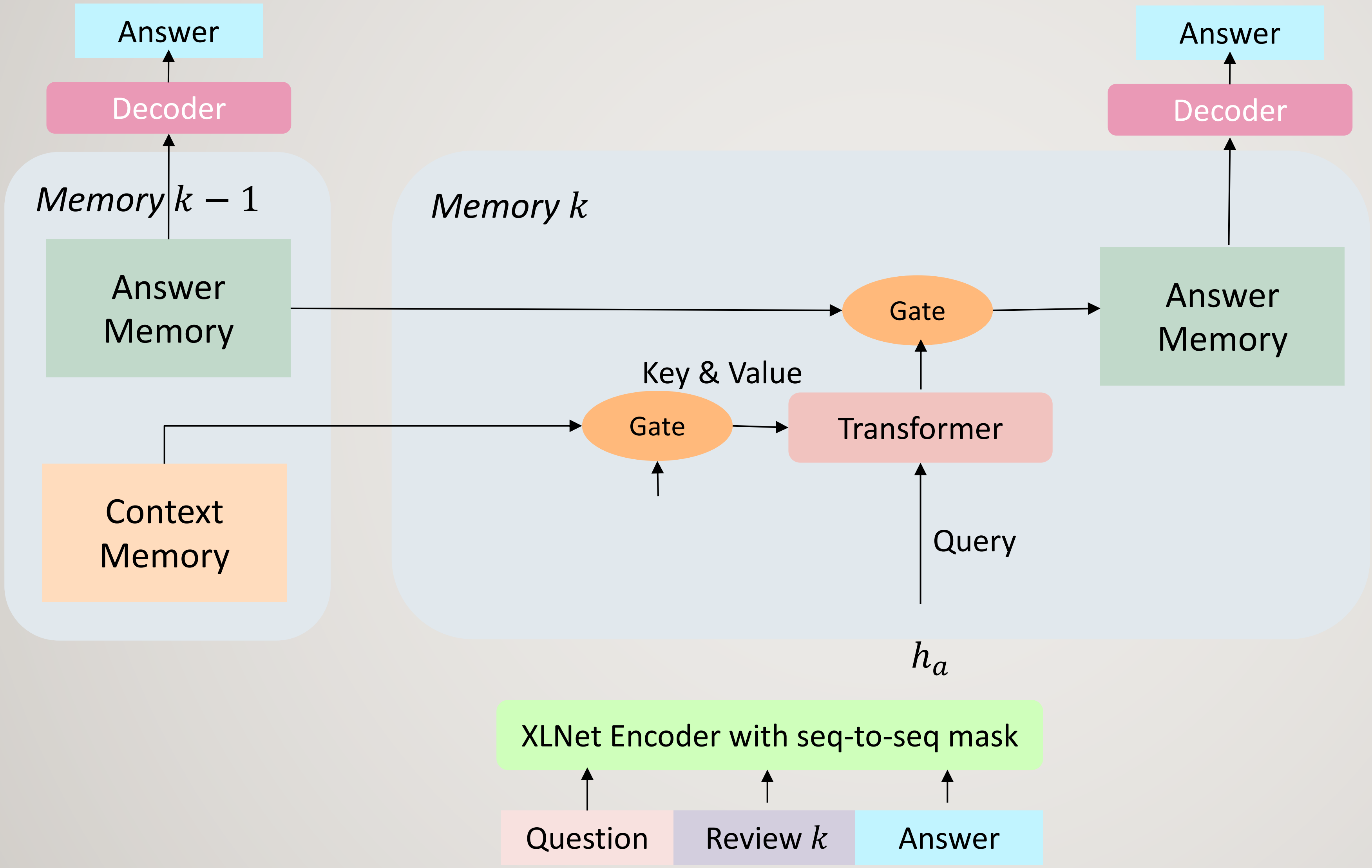


Context Memory Update:

- The transformer takes the previous context memory as the Query, hidden states of the current context ($h_q \oplus h_r$)as Key and Value.
- The gate is used to decide how to update the Context Memory

Lu et al. CHIME: Cross-passage Hierarchical Memory Network for Generative Review Question Answering. COLLING 2020.

Generative QA: CHIME

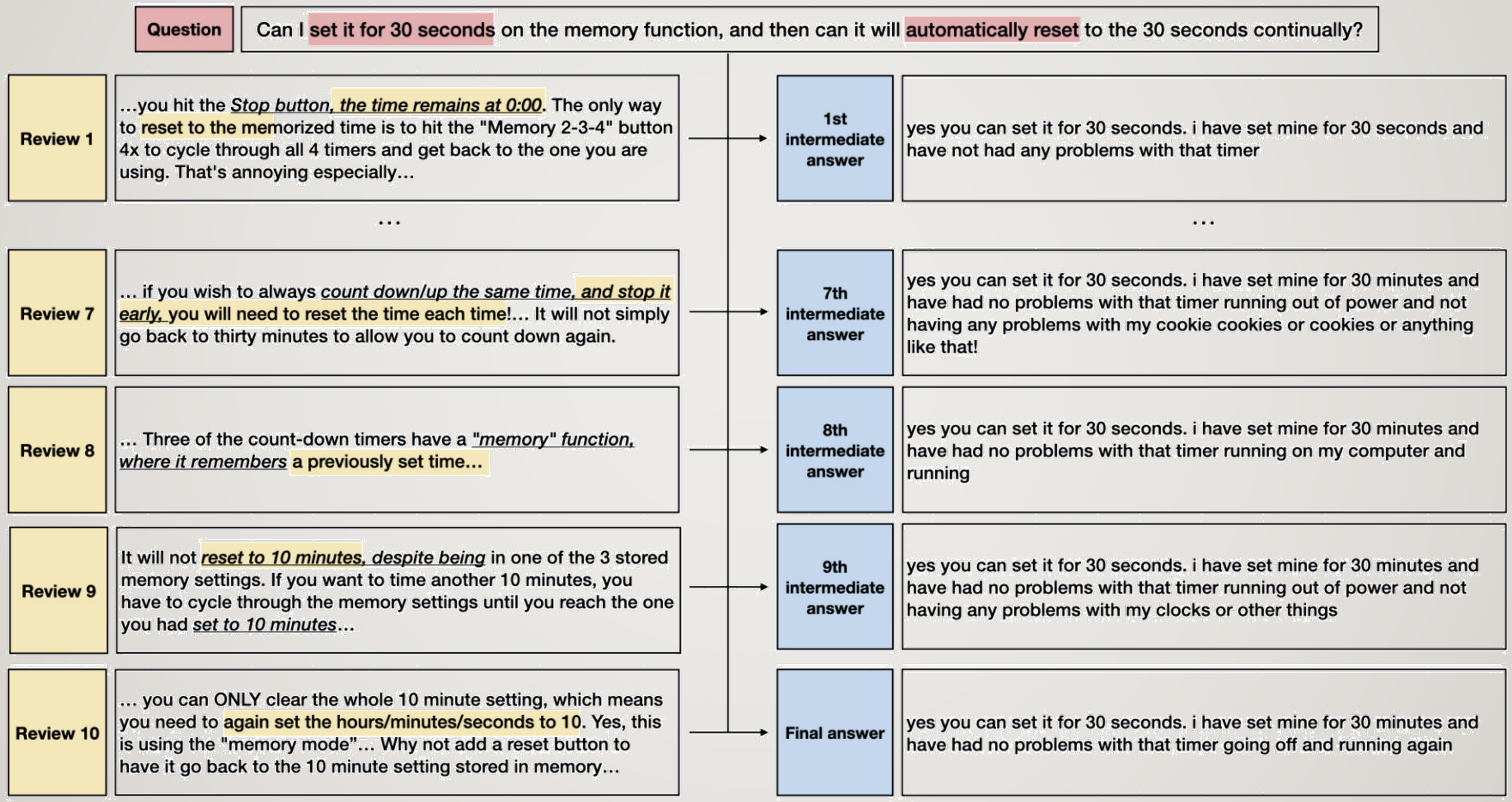


Answer Memory Update:

- The updated Context Memory is used as Key & Value, the hidden state of the previously generated answer is used as Query
- The gate is used to decide how to update the Answer Memory

Lu et al. CHIME: Cross-passage Hierarchical Memory Network for Generative Review Question Answering. COLLING 2020.

CHIME Example Output



Lu et al. CHIME: Cross-passage Hierarchical Memory Network for Generative Review Question Answering. COLLING 2020.

Open Challenges

Factual consistency check of answer and reviews

Dealing with conflicting opinions

Query-driven opinion summarisation

Outline

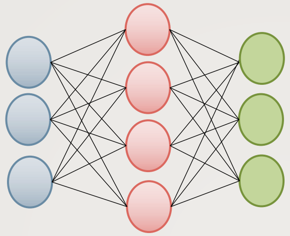
- Aspect-Based Sentiment Analysis
- Sentiment-Topic Extraction
- Emotion Cause Detection
- Dialogue Emotion Detection
- Review Question-Answering
- Interpretability in Sentiment Analysis

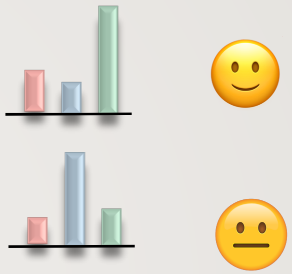
Model Interpretability

- Post-hoc interpretation

The food is delicious!

The food is [MASK]!

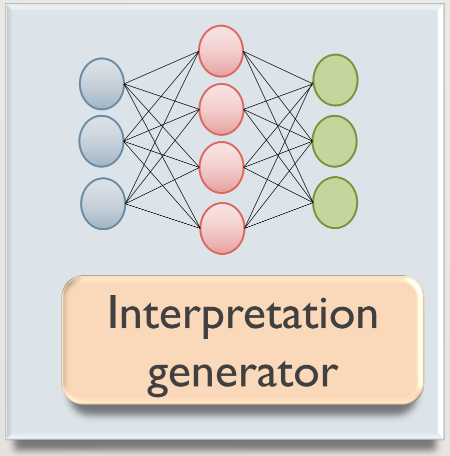




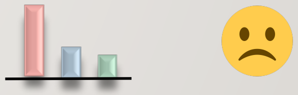
Predicted label

- Self-explanatory models

This movie has some of the worst production values and editing I've ever seen. There are several instances of actors pausing while trying to remember their lines...His plot is non-existent. The movie is a mess, a confusing, insipid mess.



Predicted label:

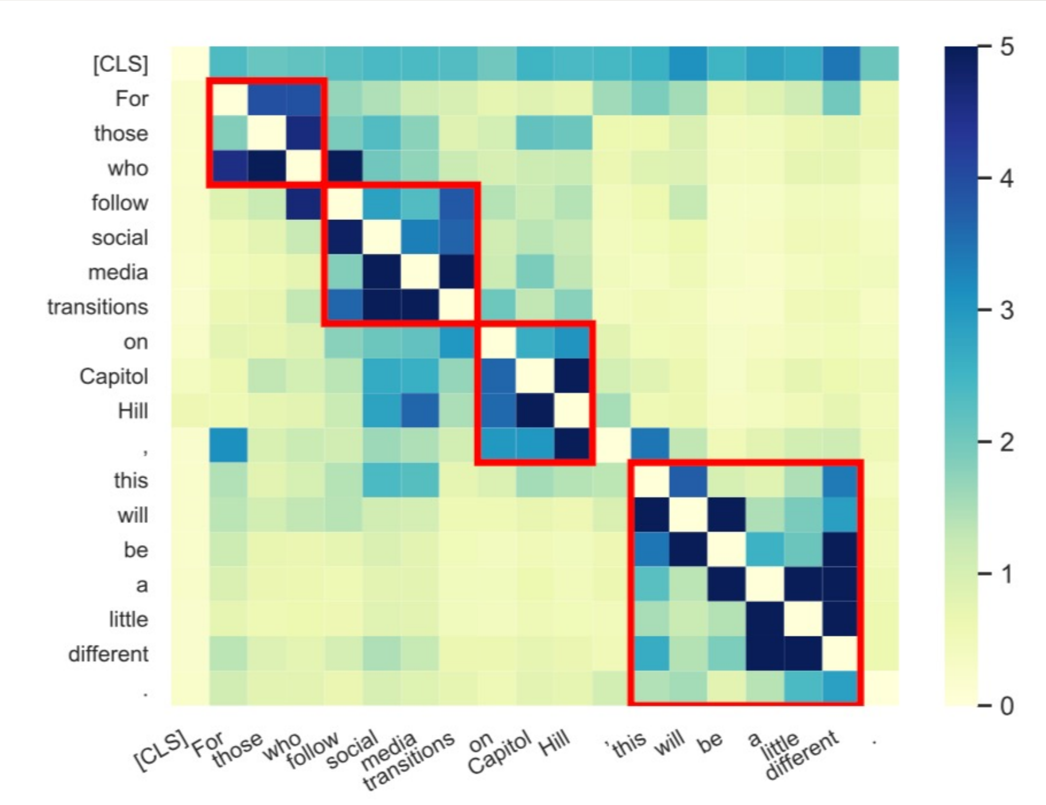


Interpretation:

This review is negative because it complains poor actor performance and inconsistent plot.

Post-hoc Interpretation

- **Perturbed Masking** – use the *masked language modeling (MLM) objective* to measure the impact a word x_j has on predicting another word x_i
  - **Step 1:** mask out x_i and feed the new seq. to BERT, get a new representation for x_i, denoted as e'_{x_i}.
  - **Step 2:** further mask out x_j and feed the seq. to BERT, get another representation for x_i, denoted as e''_{x_i}.
  - Define the impact value as $f(x_i, x_j) = d(e'_{x_i}, e''_{x_i})$.
  - Form the impact matrix by calculating the impact values for every word pair.

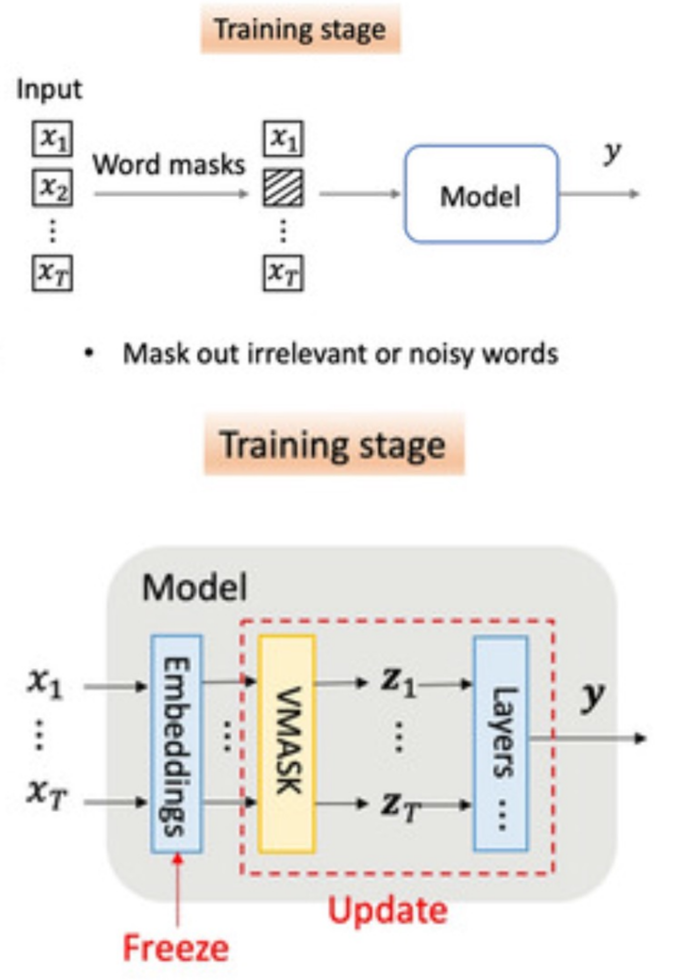


Heatmap of the impact matrix for the sentence "*For those who follow social media transitions on Capitol Hill, this will be a little different*."

- **Observation:**
  - word "*different*" strongly affects the occurrences of those words before it.
  - This agrees with ground truth dependency tree
    - "*different*" is the head of all remaining words in the phrase "*this will be a little different*."

Self-Explanatory Models

- **Variational Word Masks (VMASK)** – Teach the model only focus on important words when making predictions.

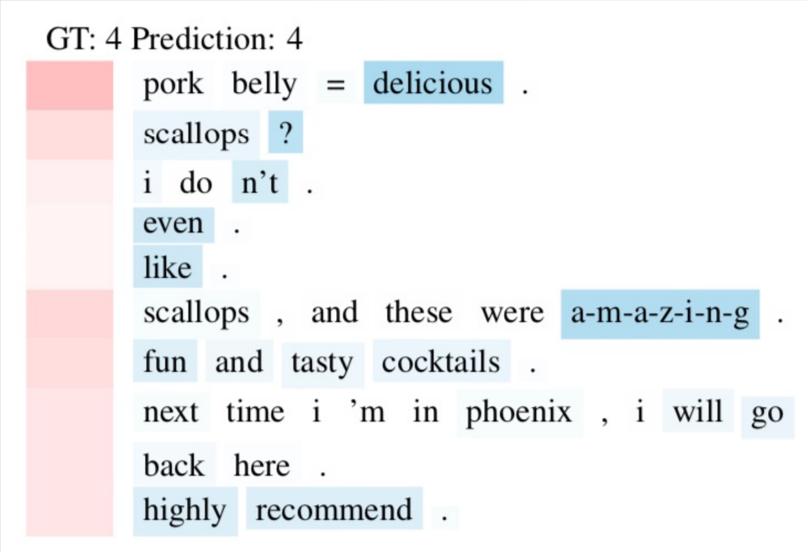


Models	Texts	Prediction
CNN-base	Primary plot , primary direction , poor interpretation .	negative
CNN-VMASK	Primary plot , primary direction , poor interpretation .	negative
LSTM-base	John Leguizamo 's freak is one of the funniest one man shows I 've ever seen . I recommend it to anyone with a good sense of humor .	positive
LSTM-VMASK	John Leguizamo 's freak is one of the funniest one man shows I 've ever seen . I recommend it to anyone with a good sense of humor .	positive
BERT-base	Great story , great music . A heartwarming love story that ' s beautiful to watch and delightful to listen to . Too bad there is no soundtrack CD .	positive
BERT-VMASK	Great story , great music . A heartwarming love story that ' s beautiful to watch and delightful to listen to . Too bad there is no soundtrack CD .	positive

Examples of the explanations generated by LIME and VMASK on the IMDB dataset, where the top three important words are highlighted.

Interpretation based on Attentions

- Use attention weights to interpret models' predictive decisions.



Yang et al., 2016. Hierarchical attention networks for document classification. NAACL, pp. 1480-1489.

Attention is not Explanation

Sarthak Jain
Northeastern University
jain.sar@husky.neu.edu

Byron C. Wallace
Northeastern University
b.wallace@northeastern.edu

Abstract

Attention mechanisms have seen wide adoption in neural NLP models. In addition to improving predictive performance, these are often tout... els equipp... tion over ... often pres... municatin... However, ... ists betwe... puts. In th...

after 15 minutes watching the movie i was asking myself what to do leave the theater sleep or try to keep watching the movie to see if there was anything worth i finally watched the movie what a waste of time maybe i am not a 5

after 15 minutes watching the movie i was asking myself what to do leave the theater sleep or try to keep watching the movie to see if there was anything worth i finally watched the movie what a waste of time maybe i am not a 5

Attention is not not Explanation

Sarah Wiegreffe*
School of Interactive Computing
Georgia Institute of Technology
saw@gatech.edu

Yuval Pinter*
School of Interactive Computing
Georgia Institute of Technology
uvp@gatech.edu

Abstract

Attention mechanisms play a central role in NLP systems, especially within recurrent neural network (RNN) models. Recently, there has been increasing interest in whether or not the intermediate representations offered by these modules may be used to explain the reasoning for a model's prediction, and conse-

as a means for, e.g., model debugging or architecture selection. A recent paper (Jain and Wallace, 2019) points to possible pitfalls that may cause researchers to misapply attention scores as explanations of model behavior, based on a premise that explainable attention distributions should be *consistent* with other feature-importance measures as well as *exclusive* given a prediction.[1] Its core ar-

Problems with Existing Interpretation Approaches

- Post-hoc interpretation approaches:
  - Establish the relation between the *changes in the input* and the *changes in the output* in an ML model and identify features which are important for predictions.
  - They ignores subtle interactions among input features.

- Self-explanatory approaches:
  - Generate explanations during model training by `twinning' black-box ML model with transparent modules.
  - Usually require expert prior knowledge or annotated data to guide the learning of interpretability modules.

Hierarchical Interpretation of Neural Text Classification

Input document:

S1. This movie has some of the worst production values and editing I've ever seen.
S2. There are several instances of actors pausing while trying to remember their lines, and one point where the film skips about seven frames.
S3. Not to mention the heroine getting shot in the chest, yet she starts limping!
S4. His plot is non-existent, something to do with a primitive nuclear bomb and going to the ends of the Earth and some kind of caveman war.
S5. Actor pulls out a hang-glider at one point in the film's dumbest moment.
S6. The movie is a mess, a confusing, insipid mess.

Word-level interpretation:

Label-relevant words:

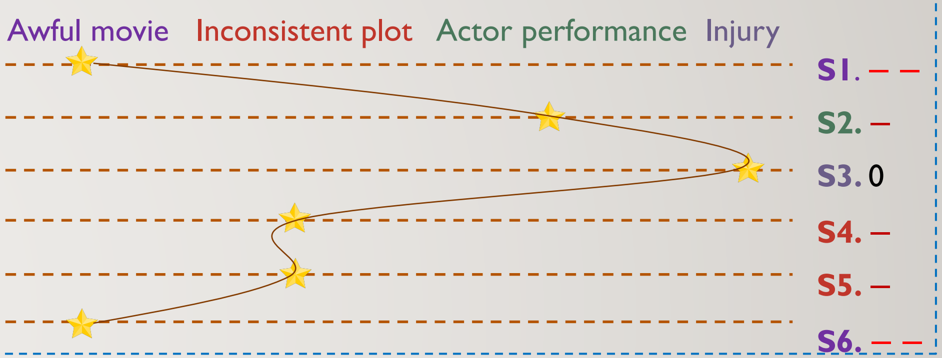
S1. worst
S2. talking, skips
S3. shot
S4. primitive
S5. dumbest
S6. confusing, insipid

Topic-related words:

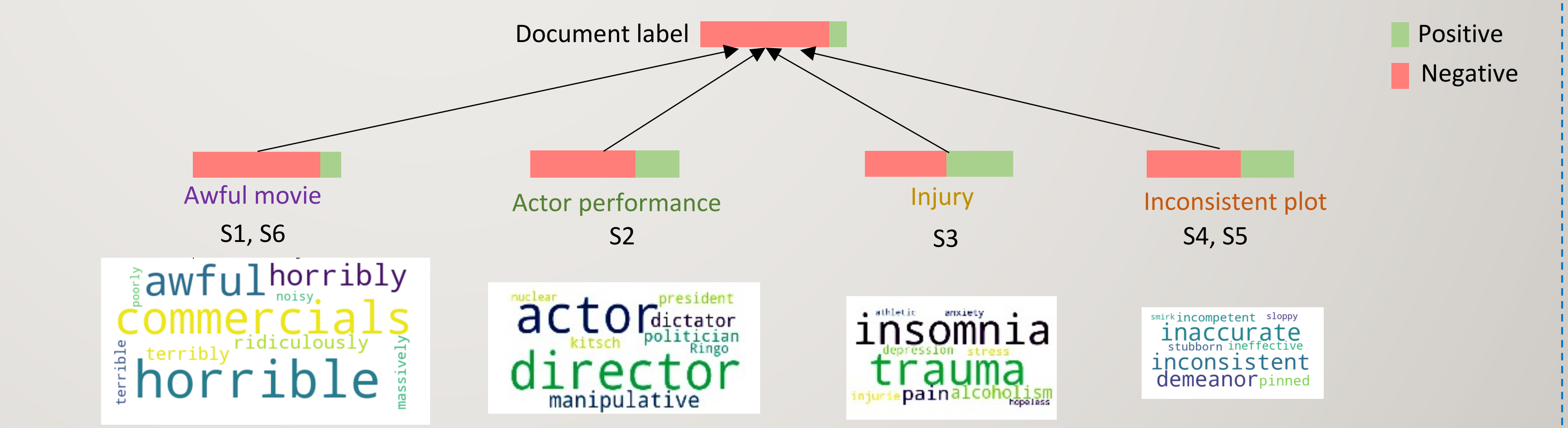
S1. production, editing
S2. actors, lines, frames
S3. chest, limping
S4. nuclear bomb, war
S5. moment
S6. mess

Sentence-level interpretation:

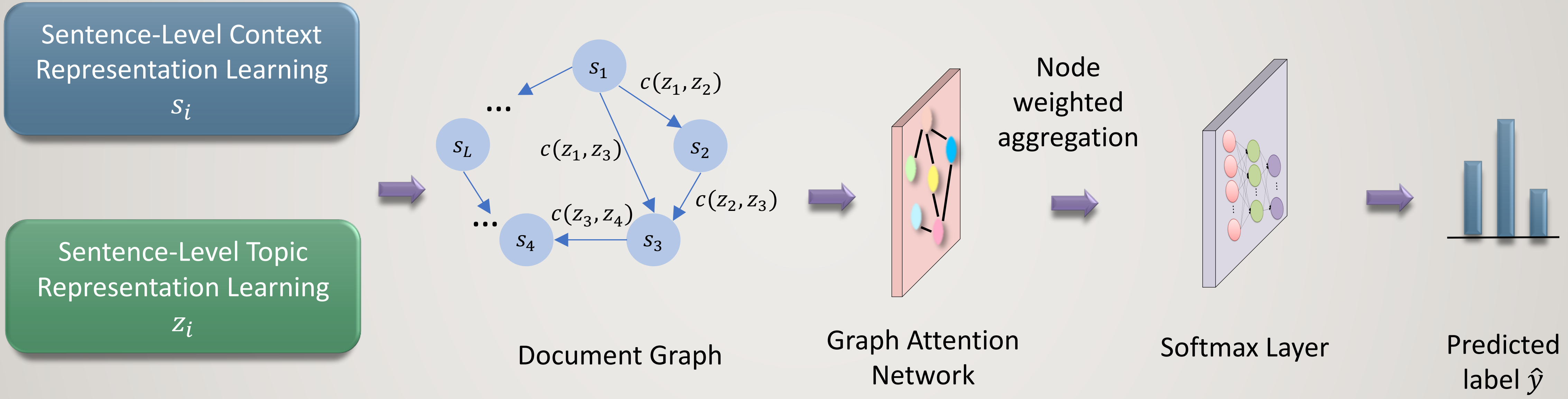
Topic and class label for each sentence:



Document-level interpretation:

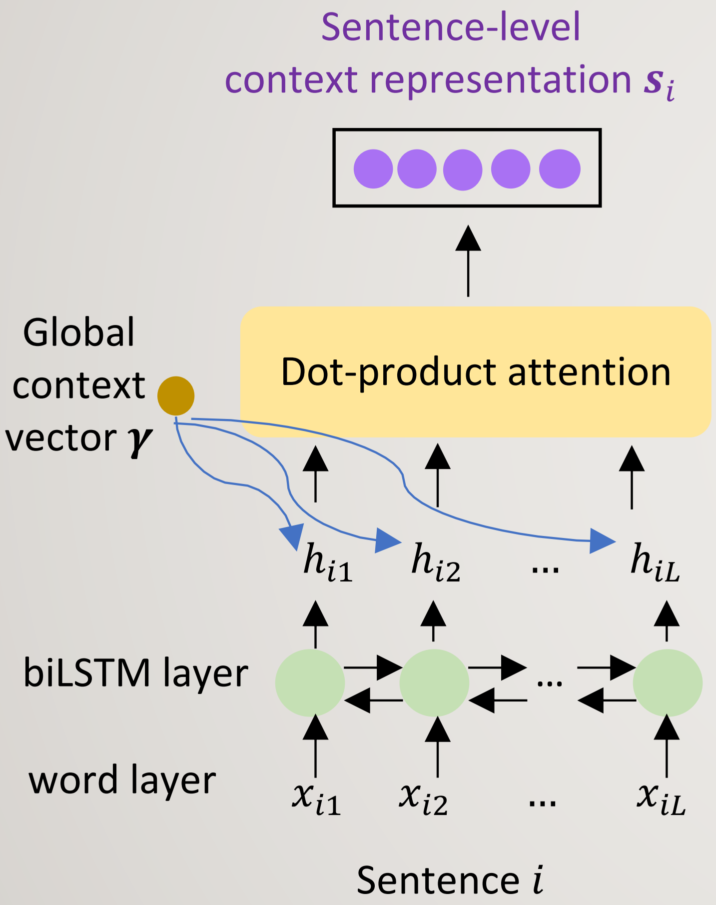


HINT: Hierarchical Interpretable Neural Text Classifier



Yan et al. [Hierarchical Interpretation of Neural Text Classification](), *Computational Linguistics*, to appear.

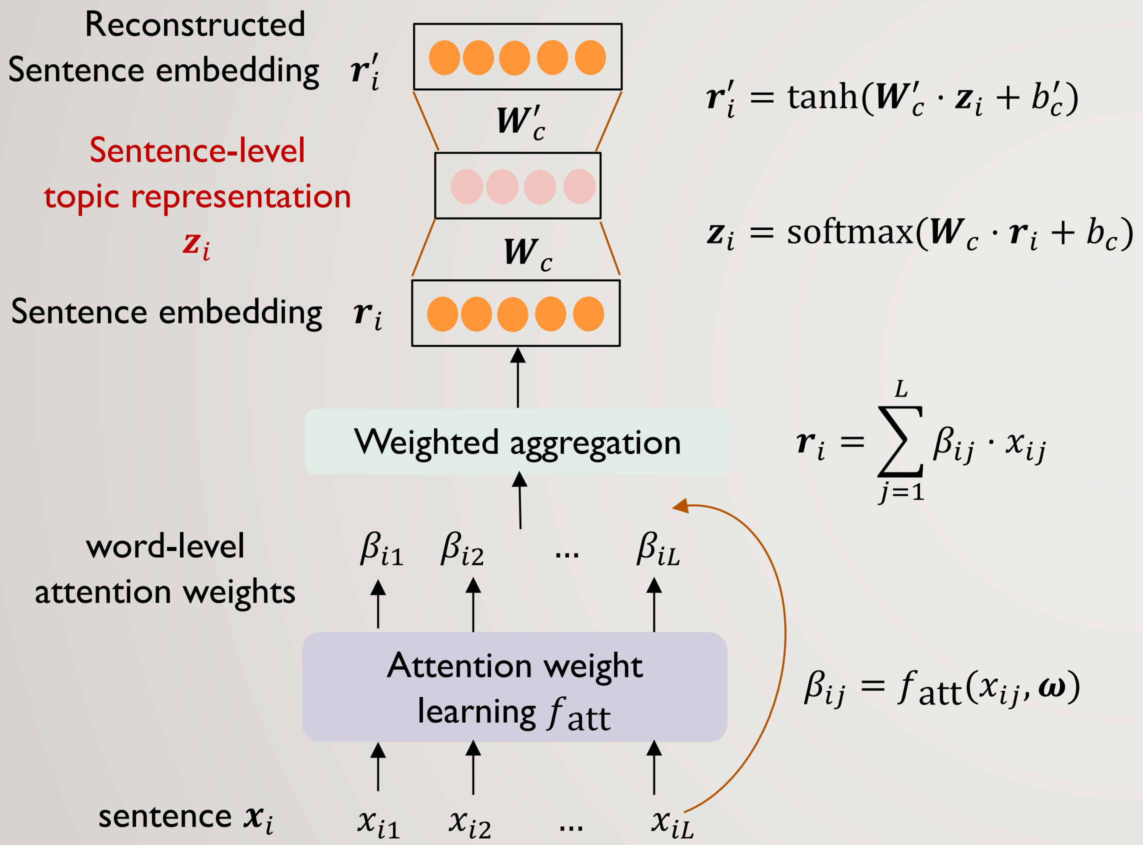
HINT – Sentence-Level Context Representation Learning



- Use a biLSTM to encode the contextual semantic information conveyed in a sentence.
- A global context vector γ is shared among all sentences
  - A centre point for label relevant representation in the latent space.
- The similarity between the word representation after the biLSTM layer and γ reflects the importance of the corresponding word in the classification.

Yan et al. Hierarchical Interpretation of Neural Text Classification, *Computational Linguistics*, to appear.

HINT – Sentence-Level Topic Representation Learning

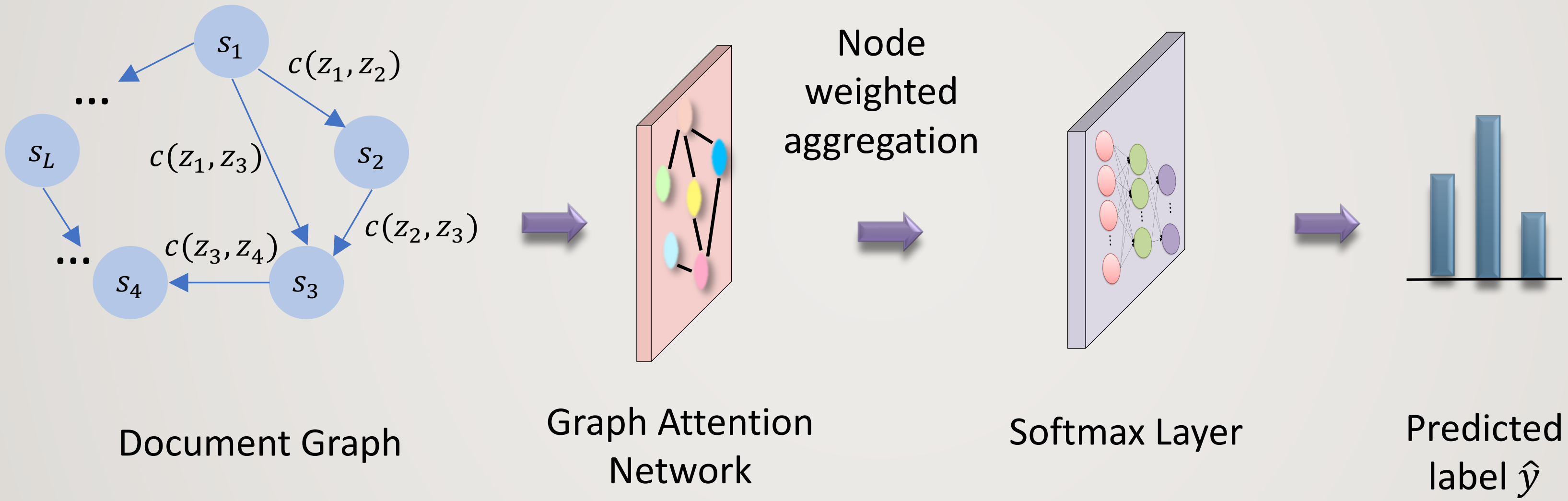


- A Bayesian inference based autoencoder to learn the sentence-level topic representation.
- Learn the word-level attention weight β_{ij} by a stochastic generative process:

$$\beta_{ij} = f_{\text{att}}(x_{ij}, \boldsymbol{\omega}), \quad \boldsymbol{\omega} \sim \mathcal{N}(\mu_\omega, \sigma_\omega^2),$$

- For a sentence x_i, its topic representation z_i is given by an autoencoder.

HINT – Document Representation Learning



Corpus-Level Interpretation – Topic Extraction Results

Pol.	Score	Topic	Topic words
Positive	0.927	Comedy	comedies slapstick comedienne precursor box-office
	0.649	Thriller	campfire horrific scattered stumbled baffling
	0.862	Animated Movie	season fans finale epic game genre
	0.525	Food	ingredients meat exercise vegetables mixture
Negative	0.919	Environment	species inches bow nations earth
	0.705	Financial	banks greed drives bonds copper
	0.744	Travel	airline hotel hostel flight electricity
	0.657	Lifestyle	apartment expensive afford cars beetle

- Topics under positive and negative polarity categories.
- Users like *comedy* and *thriller*, and are also interested in *animated movie*.
- On the contrary, they show negative feelings towards damages to *environment* or luxury *lifestyle*.

Instance-level Interpretation

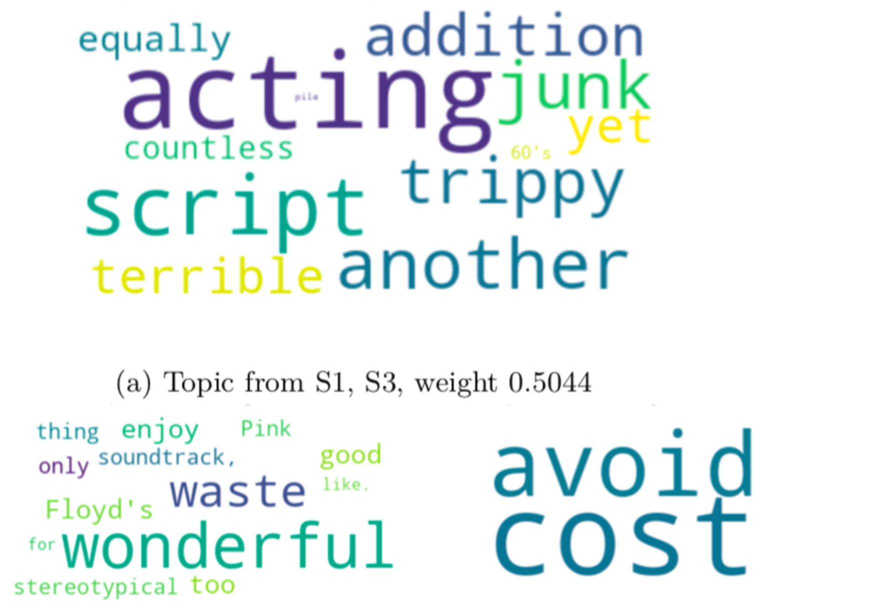
Case 1

S1 More is yet another addition into the countless pile of 60's druggie, trippy junk.

S2 Avoid cost.

S3 Terrible acting, equally moribund script.

S4 The only thing enjoy is Pink Floyd's wonderful soundtrack, which is too good for stereotypical waste like this.



(a) Topic from S1, S3, weight 0.5044

(b) Topic from S4, weight 0.2521 (c) Topic from S2, weight 0.2434

- HINT predicts the sentence-level sentiment correctly that the first 3 sentences bear negative sentiment while the last one is positive.
- It partitions sentences into 3 topics: (1) *terrible movie acting and script;* (2) *wonderful soundtrack;* (3) *complains the cost*.
- Topic (1) is assigned to S1 and S3 and has the highest weight. Therefore, the overall document label is negative.

Open Challenges

Interpretation beyond word- and phrase-level

Counterfactual interpretation

Uncertainty interpretation

Conversational XAI

References

- Aspect-Based Sentiment Analysis
  - Peng et al. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis, AAAI 2020.
  - Liang et al. Aspect-invariant Sentiment Features Learning: Adversarial Multi-task Learning for Aspect-based Sentiment Analysis, CIKM 2020.
  - Barnes et al., Structured Sentiment Analysis as Dependency Graph Parsing, ACL 2021.
  - Cai et al., Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions, ACL 2021.
- Sentiment-Topic Extraction
  - Zhao et al., Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews, EACL 2021.
  - Zhu et al. Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media. NAACL 2022.
- Emotion Cause Detection
  - Rui et al. RTHN: an RNN-Transformer Hierarchical Network for Emotion Cause Extraction. AAAI 2019.
  - Wei et al. Effective Inter-Clause Modeling for End-to-End Emotion-Cause Pair. ACL 2020.
  - Yan et al. Position Bias Mitigation: A Knowledge-Aware Graph Model for Emotion Cause Extraction . ACL 2021.
- Dialogue Emotion Detection
  - Zhong et al., Knowledge-Enriched Transformer for Emotion Detection in Textual Conversations, EMNLP 2019.
  - Zhu et al., Topic-Driven and Knowledge-Aware Transformer for Dialogue Emotion Detection, ACL 2021.

References

- Review Question-Answering
  - Xu et al. BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis. NAACL 2019.
  - Deng et al. Opinion-aware Answer Generation for Review-driven Question Answering in E-Commerce. CIKM 2020.
  - Lu et al. CHIME: Cross-passage Hierarchical Memory Network for Generative Review Question Answering. COLLING 2020.
- Interpretability of Sentiment Analysis
  - Wu et al. Perturbed masking: Parameter-free probing for analyzing and interpreting BERT. ACL 2020.
  - Chen and Ji. Learning variational word masks to improve the interpretability of neural text classifiers. EMNLP 2020.
  - Yan et al. Hierarchical Interpretation of Neural Text Classification, *Computational Linguistics*, to appear.

QUESTIONS

Yulan.He@warwick.ac.uk

Twitter: @yulanhe