Natural Language Processing Research to Drive FinTech: Now and Next

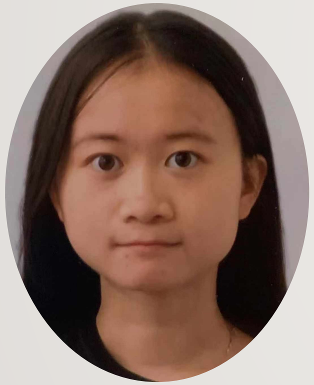
Professor Yulan He

King’s College London, UK

Lin Gui

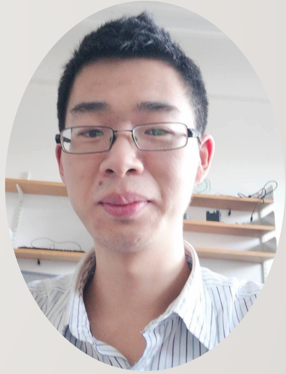
Gabriele Pergola

Xinyu Wang



Runcong Zhao

Yuxiang Zhou



Lixing Zhu

Acknowledgements

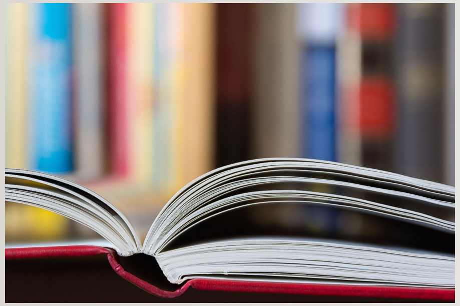
Outline

- Theme Identification
- Financial Information Extraction
- Sentiment Analysis
- End-to-end NLP-Inference approaches

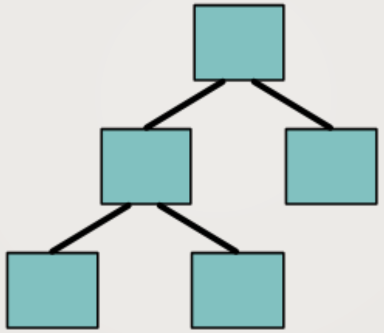
Theme Identification

Making Sense of Text

Aim: We want to gain an insight into topics discussed in a large document collection.



Document Collection

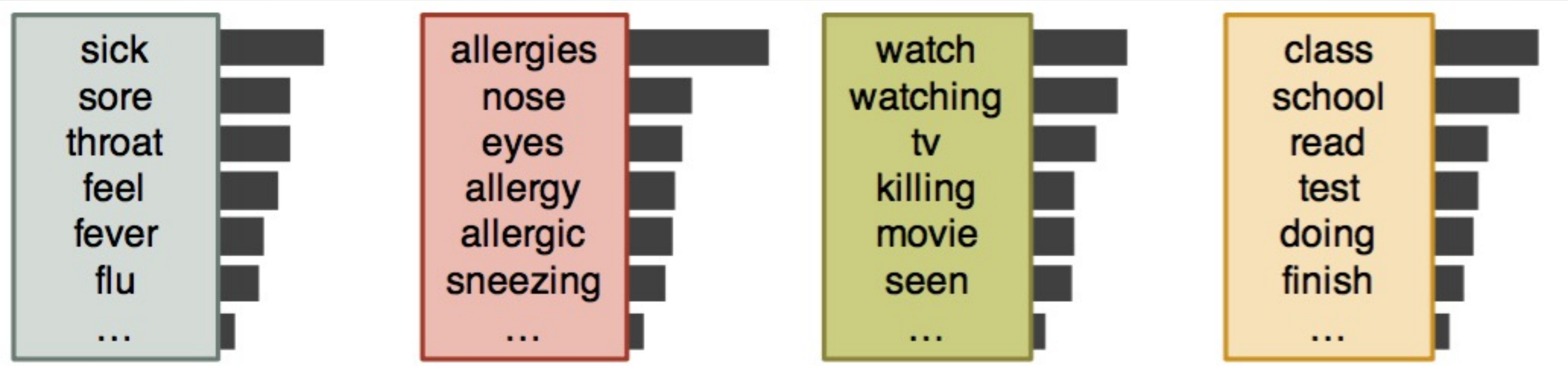


Topic Hierarchy

Insights

Topic Models (e.g., Latent Dirichlet Allocation)

- *Global context of words*



Example Topic Extraction Results by LDA

Topic 1	Topic 2
colonoscopy	pain
bleeding	knee
anemia	fracture
transfusion	leg
red	physical
blood	therapy
chronic	arthroplasty
cells	joint
transfused	surgery
glucose	osteoarthritis

Example Topic Extraction Results

Topic 1	Topic 2
colonoscopy	pain
bleeding	knee
anemia	fracture
transfusion	leg
red	physical
blood	therapy
chronic	arthroplasty
cells	joint
transfused	surgery
glucose	osteoarthritis

red blood cells

Example Topic Extraction Results

Topic 1	Topic 2
colonoscopy	pain
bleeding	knee
anemia	fracture
transfusion	leg
red	physical
blood	therapy
chronic	arthroplasty
cells	joint
transfused	surgery
glucose	osteoarthritis

red blood cells

blood glucose

Problems

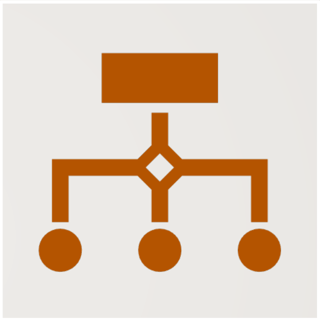
- Documents may contain phrases which are semantically non-decomposable
  - E.g., “red blood cells”, “physical therapy”, …
- Unigram topic models may be ambiguous and generate spurious topics
  - E.g., “red blood cells” are mixed with “blood glucose”

Example Topical Phrases – Latent Dirichlet Allocation

Topic 1	Topic 2	Topic 3
	LDA	
swallowing	infection	surgery
transferred	vancomycin	repair
intubated	antibiotics	wound
intensive_care	culture	signs
speech	fever	removed
wean	intervertebral	nasogastric_tube
extubated	culture_blood	diets
tube_feeds	fluid	female
arrest	levaquin	hospital_course
aspiration	gentamicin	diverticulitis

Unsupervised Topical Phrase Extraction – Topical Phrase Model

Extract phrases first using an off-the-shelf concept extraction tool



Learn a topic model which can model the generation of phrases

Takes a hierarchy of Pitman-Yor Processes as prior

Capture *n*-grams of arbitrary length naturally by considering word orders within phrases

Y. He. Extracting Topical Phrases from Clinical Documents. The 30th AAAI Conference on Artificial Intelligence (AAAI), 2016.

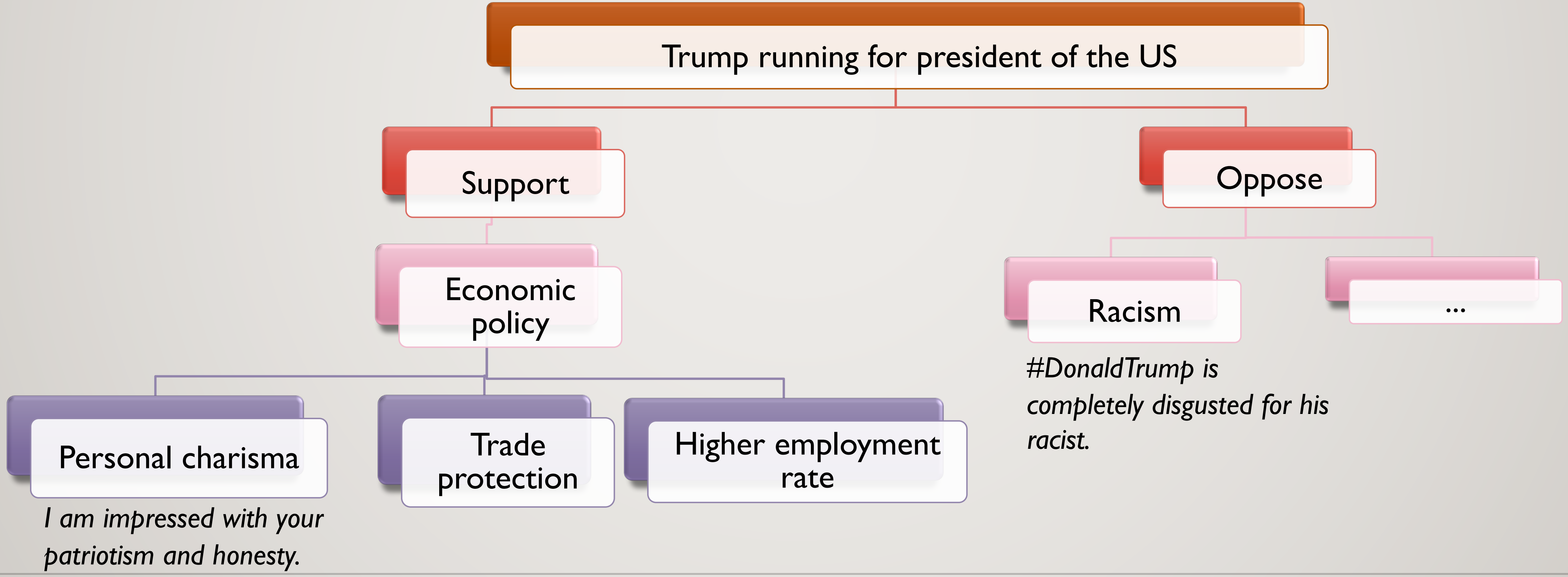
Example Topical Phrases Extracted by Topical Phrase Model

Topic 1	Topic 2	Topic 3
	TPM	
restrictive lung disease	right coronary artery	ciprofloxacin 500 mg
pleuritic chest pain	left upper extremity	levofloxacin 500 mg
ferrous sulfate 325 mg	systemic vascular resistance	levofloxacin 250 mg
dyspnea on exertion	systolic ejection murmur	ciprofloxacin 250 mg
vq scan	pulmonary vascular resistance	metronidazole 500 mg
breath chest pain	flash pulmonary edema	chronic urinary tract infection
interstitial lung disease	shortness of breath	white blood cell count
morbid obsesity	transesophageal echocardiogram	benign prostatic hyperplasia
arterial blood gas	left internal mammary artery	recurrent urinary tract
pulmonary function tests	coronary artery disease	irbesartan 150 mg

Disease name: Symptoms: Diagnosis method:

Hierarchical Viewpoint Discovery

- When users express their stances towards a topic in social media, they might elaborate their viewpoints or reasoning.
- Oftentimes, viewpoints expressed by different users exhibit a hierarchical structure.



Hierarchical Opinion Phrase Model

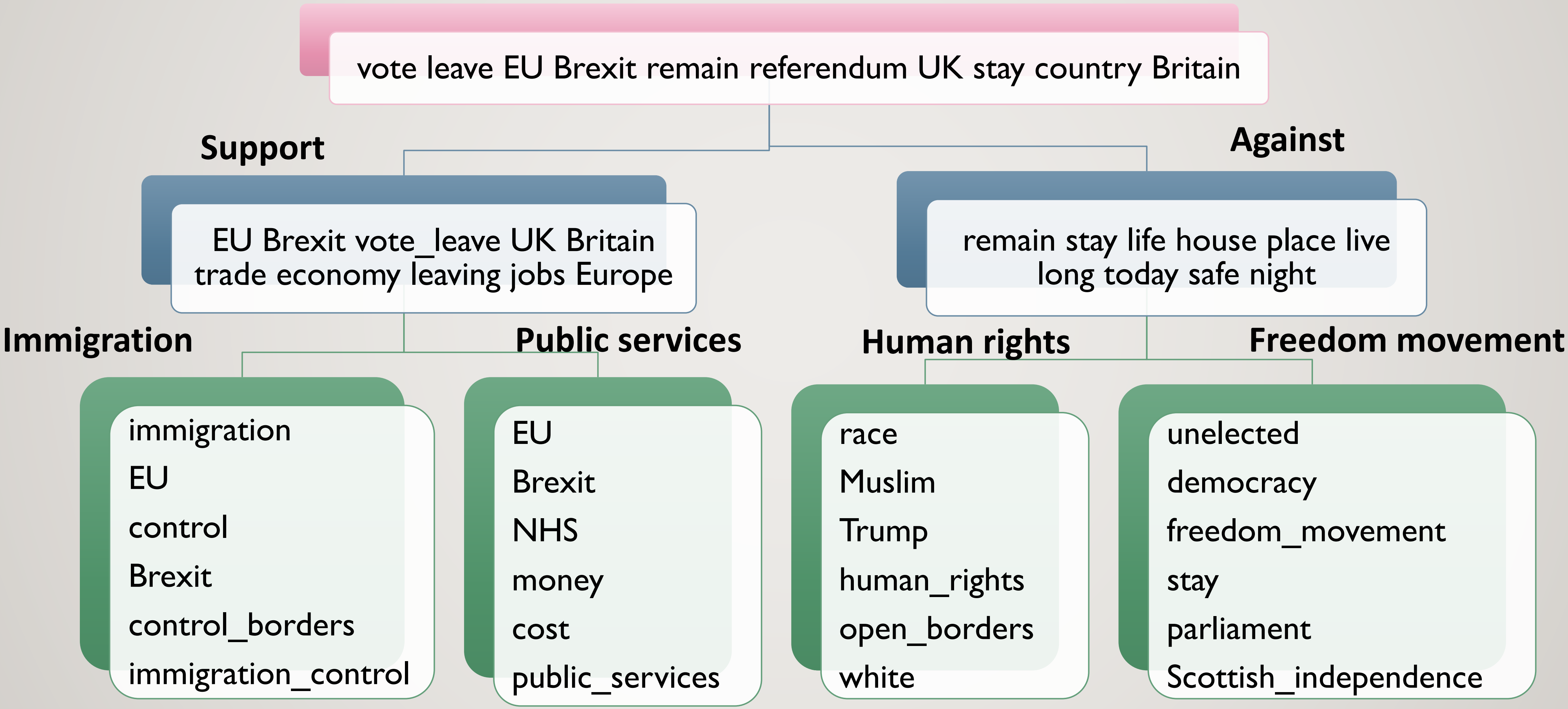
- The Topical Phrase Model can be extended to generate hierarchical viewpoints.
  - The root node (level-1) contains the topic of interest (e.g., 'Trump run for president')
  - The level-2 topics indicate stance (e.g., either "***Support***" or "***Oppose***")
  - Topics in the level-3 and below contain viewpoints under different stances.

- Generative process:
  - Document generation can be modelled by a nested Chinese Restaurant Process (Blei et al., 2010).
  - The level-2 topic is sampled from a Bernoulli distribution.
  - Generation of phrases under each topic follows the Hierarchical Pitman Yor process (Teh, 2006).

D.M. Blei, T.L. Griffiths, M.I. Jordan. The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies Journal of the ACM, 57(2):1-30, (2010),

Y.W. Teh. A hierarchical Bayesian language model based on Pitman–Yor processes. ACL 2006,

Brexit



L. Zhu, Y. He and D. Zhou. Hierarchical Viewpoint Discovery from Tweets Using Bayesian Modelling. Expert Systems with Applications, 116:430-438, 2019.

US General Election

- Trump Hillary Donald_Trump president America
  - **Vote Trump**: Trump win Hillary vote Clinton
    - **Email scandal**: Hillary_Clinton FBI campaign emails Human indicted spooky_story email FBI_investigation jail
    - **Drain swamp**: Trump America country great president repeal_ObamaCare change world drain_swamp won
      - deport Hunger illegal_immigrans million_immigrants
      - reliable lead_unified_republican wind trusted
  - **Vote Hillary**: Hillary Trump Clinton Obama America
    - **Racist**: Trump women white racist black women_rights black_voters shit build_wall child_rape
    - **Russia**: Trump Putin markets Russian economy crash Russia nuclear_war energy civil_war

L. Zhu, Y. He and D. Zhou. Hierarchical Viewpoint Discovery from Tweets Using Bayesian Modelling. Expert Systems with Applications, 116:430-438, 2019.

Open Challenges

Hierarchical topic extraction on streaming data – dealing with topic drift

Interpretability of topics

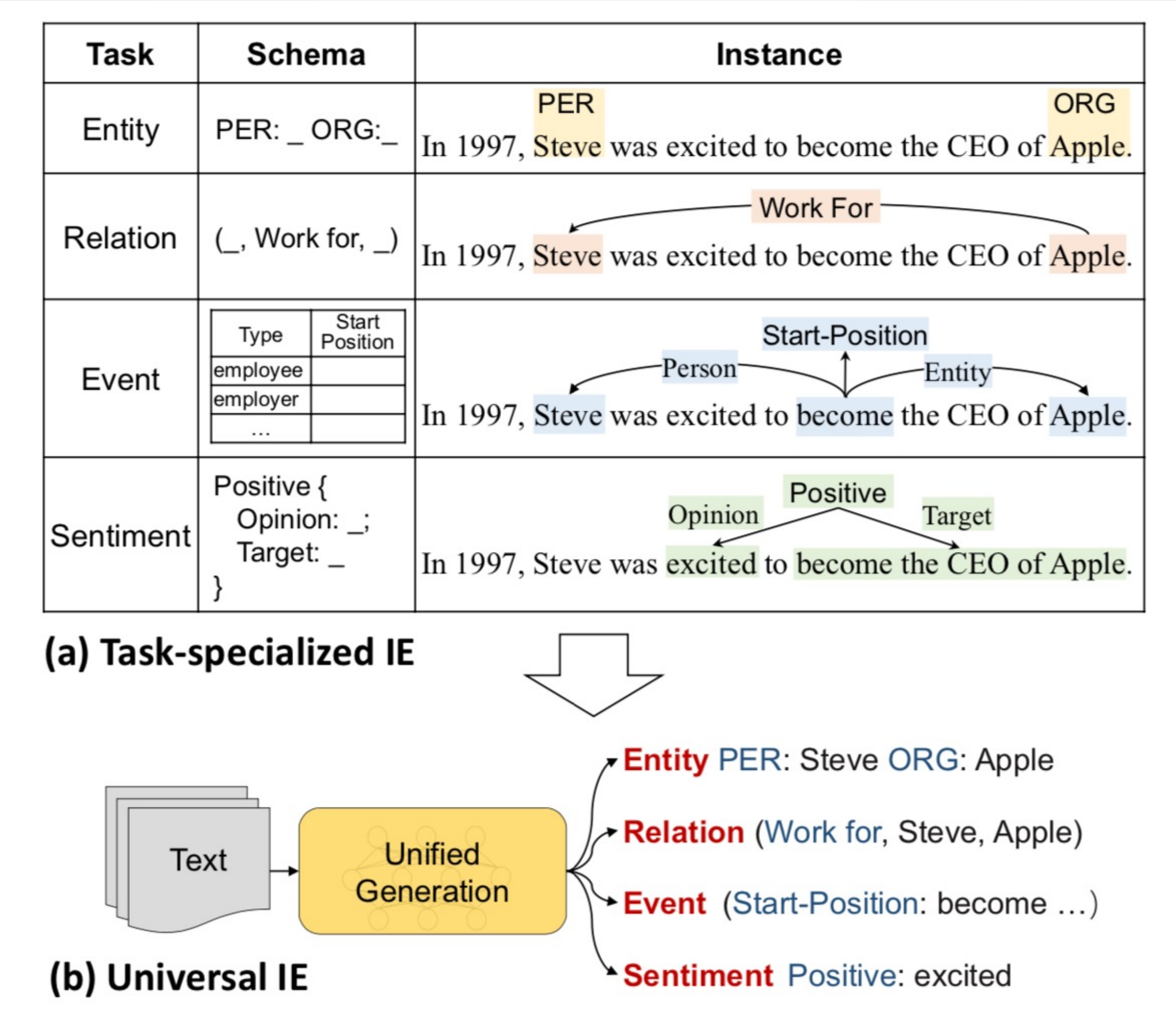
Human-in-the-loop topic extraction

Financial Information Extraction

State-of-the-Art in Named Entity Recognition (NER)

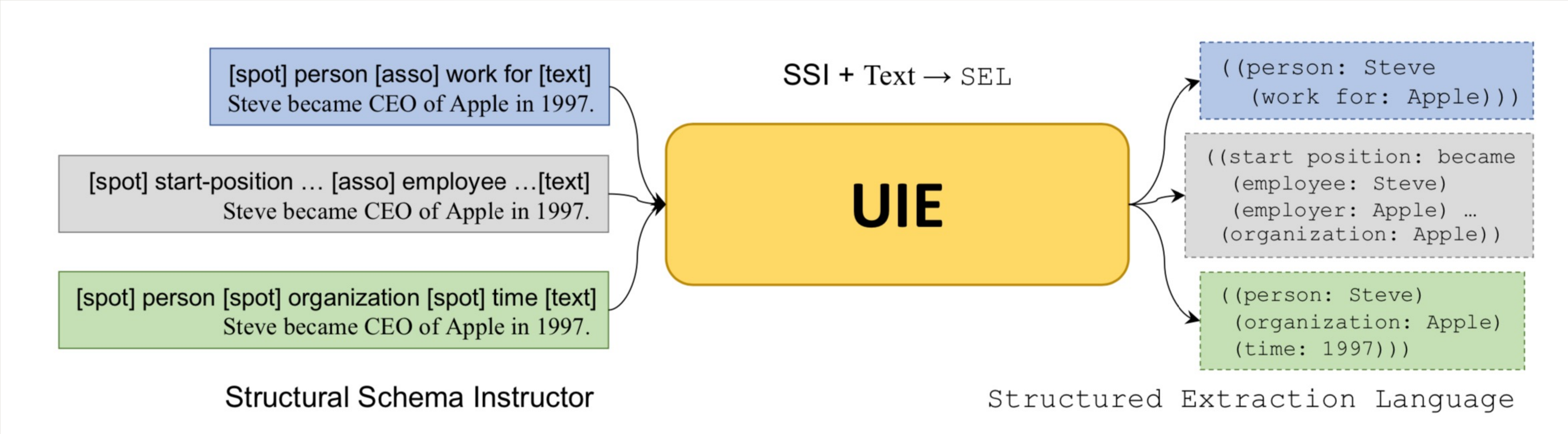
Dataset	Entity Types	Model	F1
CoNLL 2003 NER task Newswire text from the Reuters RCV1 corpus	four entity types (PER, LOC, ORG, MISC)	Automate the process of finding better concatenations of embeddings for structured prediction (Wang et al., 2021)	94.6
Ontonotes corpus v5	18 tags, consisting of 11 types (PERSON, ORGANIZATION, etc) and 7 values (DATE, PERCENT, etc)	BERT+Key-Value Memory Network (Nie et al., 2020)	90.32
Few-NERD consists of 188,238 sentences from Wikipedia, 4,601,160 words and each is annotated as context or a part of a two-level entity type	8 coarse-grained types, 66 fine-grained types	BERT-Tagger (Ding et al., 2021)	68.88

Universal Information Extraction



Lu et al., [Unified Structure Generation for Universal Information Extraction](). ACL 2022

Universal Information Extraction



Lu et al., Unified Structure Generation for Universal Information Extraction. ACL 2022

Few-Shot Learning on CoNLL03

Model	1-Shot	5-Shot	10-Shot	AVE-S	1%	5%	10%	AVE-R
T5-v1.1-base	12.73	30.17	58.89	33.93	75.74	85.71	87.70	83.05
Fine-tuned T5-base	24.93	54.85	65.31	48.36	78.51	87.67	88.91	85.03
UIE-base w/o SSI	43.52	64.76	72.47	60.25	81.91	**88.41**	**89.84**	86.72
UIE-base	**46.43**	**67.09**	**73.90**	**62.47**	**82.84**	88.34	89.63	**86.94**

- Models:
  - T5-v1.1-base – an initial model of UIE- base;
  - Fine-tuned T5-base – fine-tuned with sequence generation tasks such as summarization,;
  - UIE- base w/o SSI – without structural schema instructor (SSI) in the pre-training stage
- Experiments on six different partitions of the original training sets
  - 1/5/10- shot, 1/5/10% ratio
  - Sample 1/5/10 sentences for each entity type in the training set.

Financial Event Extraction

...Shenkai Petrochemical Co., Ltd. received the receipt from the company's shareholder, Yexiang Investment Ltd. on the evening of November 15, 2016. Regarding the notice of the shares being frozen. ... On November 14, 2016, Yexiang Investment received the Notice of Litigation Preservation from the Court of Binjiang District, and granted a total of 47577481 shares held by Yexiang Investment will be frozen, and the freezing period is from October 31, 2016 to October 30, 2019 ... Yexiang Investment is … holding 47577481 shares of the company, accounting for 13.07% of the company's total share capital. ... On February 2, 2016, the 42 million shares held by it are pledged to Haitong Securities Ltd., and the repurchase transaction date was February 1, 2017. ...

Financial Event Extraction

...Shenkai Petrochemical Co., Ltd. received the receipt from the company's shareholder, **Yexiang Investment Ltd.** on the evening of November 15, 2016. Regarding the notice of the shares being frozen. ... On November 14, 2016, Yexiang Investment received the Notice of Litigation Preservation from the **Court of Binjiang District**, and granted a total of **47577481 shares** held by Yexiang Investment will be frozen, and the freezing period is from **October 31, 2016** to **October 30, 2019** ...Yexiang Investment is … holding **47577481 shares** of the company, accounting for **13.07%** of the company's total share capital. ... On **February 2, 2016**, the **42 million shares** held by it are pledged to **Haitong Securities Ltd.**, and the repurchase transaction date was **February 1, 2017**. ...

Event #1: Equity Pledge

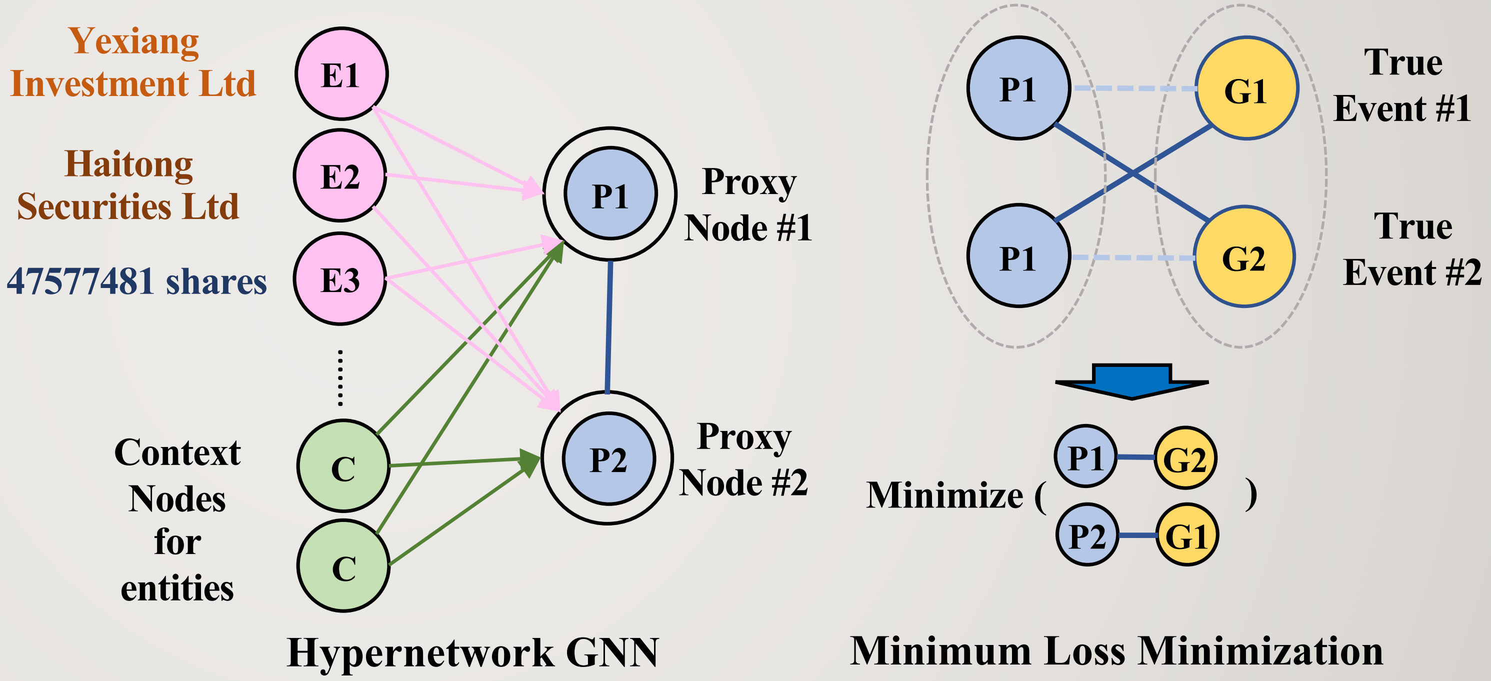
Pledger: Yexiang Investment Ltd.
Pledgee: Haitong Securities Ltd.
HoldingShares: 47577481 shares
HoldingRatio: 13.07%
PledgedShares: 42 million shares
StartDate: February 2, 2016
EndDate: February 1, 2017

Event #2: Equity Freeze

EquityHolder: Yexiang Investment Ltd.
LegalInstitution: Court of Binjiang District
HoldingShares: 47577481 shares
HoldingRatio: 13.07%
FrozeShares: 47577481 shares
StartDate: October 31, 2016
EndDate: October 30, 2019

ProCNet – Proxy Nodes Clustering Network

- **ProCNet:** a novel approach via clustering with **proxy nodes** representing events.
- The process can be viewed as an **iterative metric learning** where
  - proxy nodes are centroids representing events;
  - entities are mapped to an event-oriented metric space by the Hypernetwork GNN.



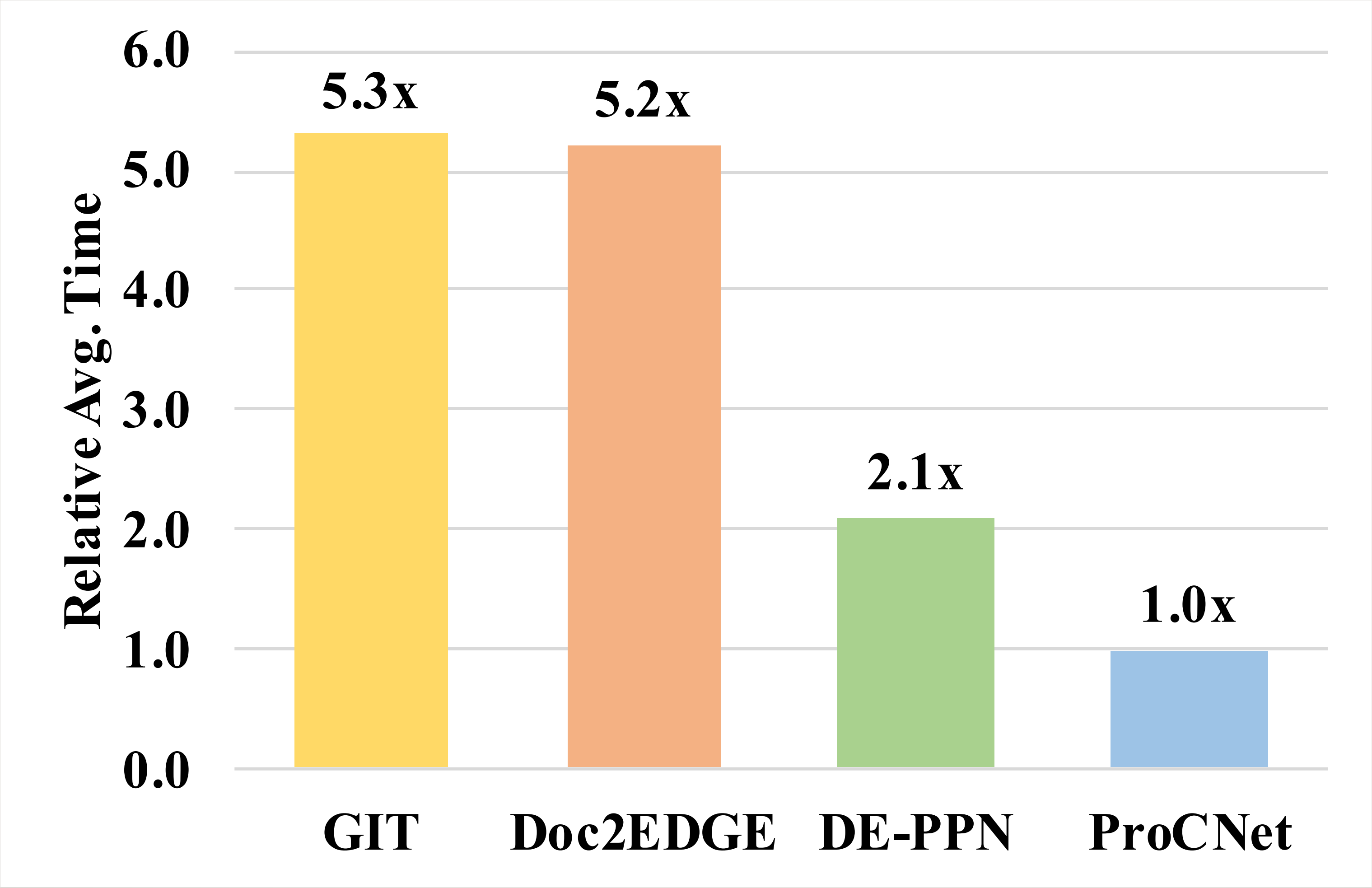
X. Wang, L. Gui and Y. He. Document-Level Multi-Event Extraction with Proxy Nodes and Hausdorff Distance, in submission.

Financial Event Extraction Results

... [4] On October 16, 2018, Jinhui Wine Co., Ltd. (hereinafter referred to as the "Company") received the shareholder **Yingzhijiu Equity Investment Limited Partnership** (hereinafter referred to as "Yingzhijiu") to transfer the company it holds ... [6] **On June 27, 2016**, Yingzhijiu pledged **9,000,000 shares** of the company it held to **Guoyuan Securities Co., Ltd.** (hereinafter referred to as "Guoyuan Securities") for processing pledged repurchase transactions. ... [7] On June 6, 2017, the company implemented the 2016 profit distribution plan, that is, every 10 shares will be increased by 3 shares and a cash dividend of 2.4 yuan will be distributed. After the transfer, the above pledged shares increased from 9,000,000 shares to **11,700,000 shares** ... [8] On **October 15, 2018**, Yingzhijiu released all the above-mentioned **11,700,000 shares** from the pledge, ... [12] On **June 26, 2017**, Yingzhijiu pledged **2,000,000 shares** of the company it held to Guoyuan securities. ... [13] On **October 12, 2018**, Yingzhijiu pledged **1,000,000 shares** of the company (accounting for 0.275% of the total share capital of Jinhui Wine) to Guoyuan Securities as a supplementary pledge for the above-mentioned pledged shares, and handled the related procedures. ... [16] As of the date of this announcement, Yingzhijiu held **20,653,685 shares** of the company, accounting for **5.674%** of the company's total share capital, and pledged **4,993,030 shares** in total, accounting for 24.175% of the company's shares and 1.372% of the company's total share capital. . ...

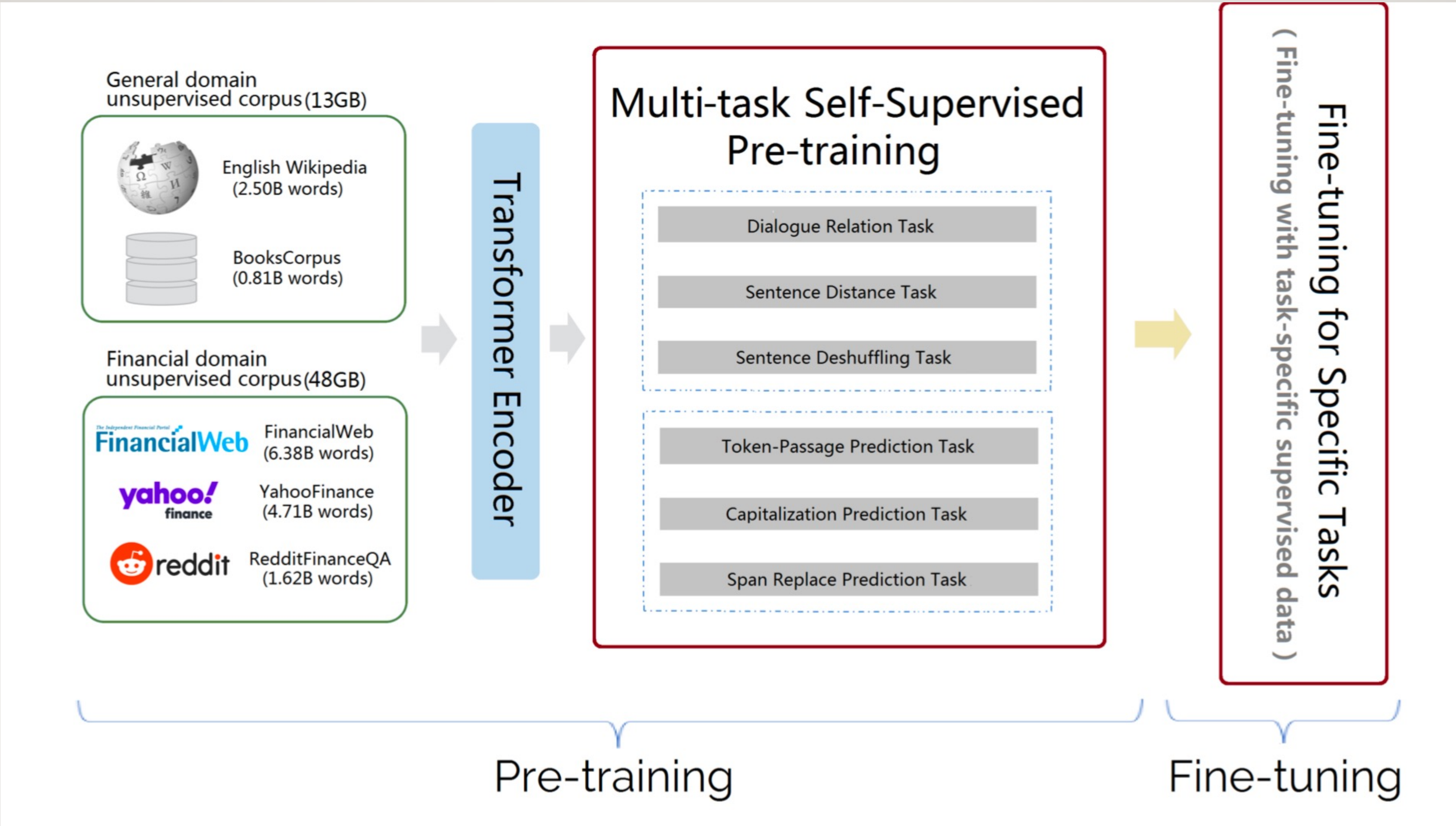
Event #1: Equity Pledge	**Event #2: Equity Pledge**	**Event #3: Equity Pledge**
Pledger: Yingzhijiu Equity Investment Limited Partnership	Pledger: Yingzhijiu Equity Investment Limited Partnership	Pledger: Yingzhijiu Equity Investment Limited Partnership
PledgedShares: 11,700,000 shares	PledgedShares: 1,000,000 shares	PledgedShares: 2,000,000 shares
Pledgee: Guoyuan Securities Co., Ltd.	Pledgee: Guoyuan Securities Co., Ltd.	Pledgee: Guoyuan Securities Co., Ltd.
TotalHoldingShares: 20,653,685 shares	TotalHoldingShares: 20,653,685 shares	TotalHoldingShares: 20,653,685 shares
TotalHoldingRatio: 5.674%	TotalHoldingRatio: 5.674%	TotalHoldingRatio: 5.674%
TotalPledgedShares: 4,993,030 shares	TotalPledgedShares: 4,993,030 shares	TotalPledgedShares: 4,993,030 shares
StartDate: June 26, 2016	StartDate: October 12, 2018	StartDate: June 26, 2017
EndDate: null	EndDate: null	EndDate: null
ReleasedDate: October 15, 2018	ReleasedDate: null	ReleasedDate: null

Run-time Comparison



Sentiment Analysis

FinBERT: A pre-trained financial language representation model for financial text mining



Liu et al. FinBERT: A pre-trained financial language representation model for financial text mining. IJCAI 2020.

Datasets for Financial Sentiment Analysis

- Financial PhraseBank (Malo et al., 2014)
  - 4845 English sentences randomly selected from **financial news** in the LexisNexis database.
  - Annotated as positive/neutral/negative

- FiQA Sentiment Analysis dataset
  - Data from financial news headlines and financial microblogs, with manually annotated target entities, sentiment scores and aspects.
  - The **financial news headlines** dataset contains a total 529 annotated headlines samples (436 samples for the training set and 93 samples for the test set)
  - The **financial microblogs** contains a total 774 annotated posts samples (675 samples for the training set and 99 samples for the test set)

FinBERT Results on Sentiment Analysis

Model	Accuracy	F1
LPS [Malo *et al.*, 2014]	0.71	0.71
HSC [Krishnamoorthy, 2018]	0.71	0.76
ULMFit [Raaci, 2019]	0.83	0.79
FB-SA [Raaci, 2019]	0.86	0.84
$\text{FinBERT}_{\text{BASE}}$ (ours)	0.91	0.89
$\text{FinBERT}_{\text{LARGE}}$ (ours)	**0.94**	**0.93**

Financial PhraseBank

	headline		*post*	
Model	**MSE**	$\mathbf{R}^2$	**MSE**	$\mathbf{R}^2$
CUKG [FiQ, 2018]	0.13	0.46	0.10	0.09
IIT-Dehi ♮	0.20	0.18	0.10	0.08
Inf-UFG ♮	0.21	0.17	0.10	0.16
NLP301 [FiQ, 2018]	-	-	0.31	-1.67
SC-V [Yang *et al.*, 2018]	0.08	0.40	-	-
RCNN [Piao *et al.*, 2018]	0.09	0.41	-	-
FB-SA [Raaci, 2019]	0.07	0.55	-	-
$\text{FinBERT}_{\text{BASE}}$ (ours)	0.29	0.67	0.28	0.26
$\text{FinBERT}_{\text{LARGE}}$ (ours)	**0.38**	**0.77**	**0.37**	**0.36**

FiQA Sentiment Dataset

Liu et al. FinBERT: A pre-trained financial language representation model for financial text mining. IJCAI 2020.

Aspect-Based Sentiment Analysis

- **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

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*”)

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

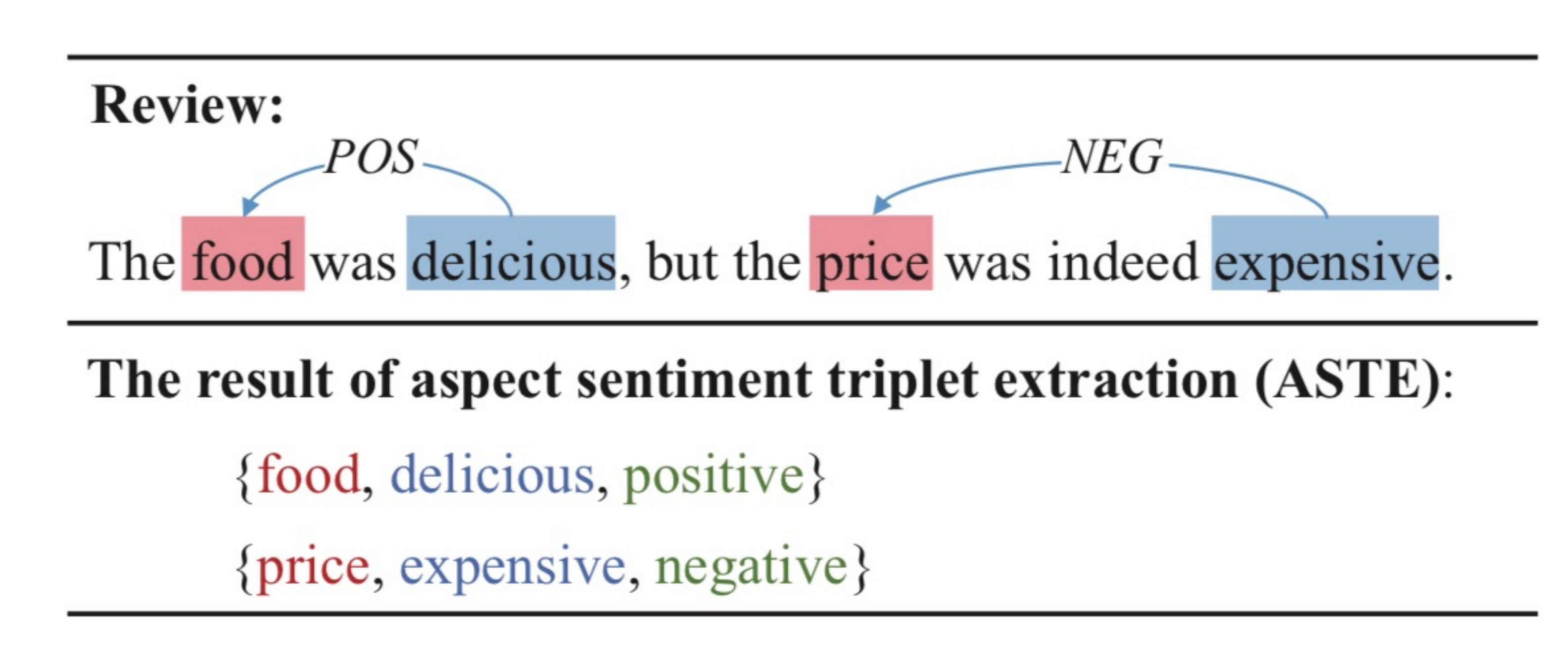
Example Output

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.

Machine Reading Comprehension for Aspect Sentiment Triplet Extraction

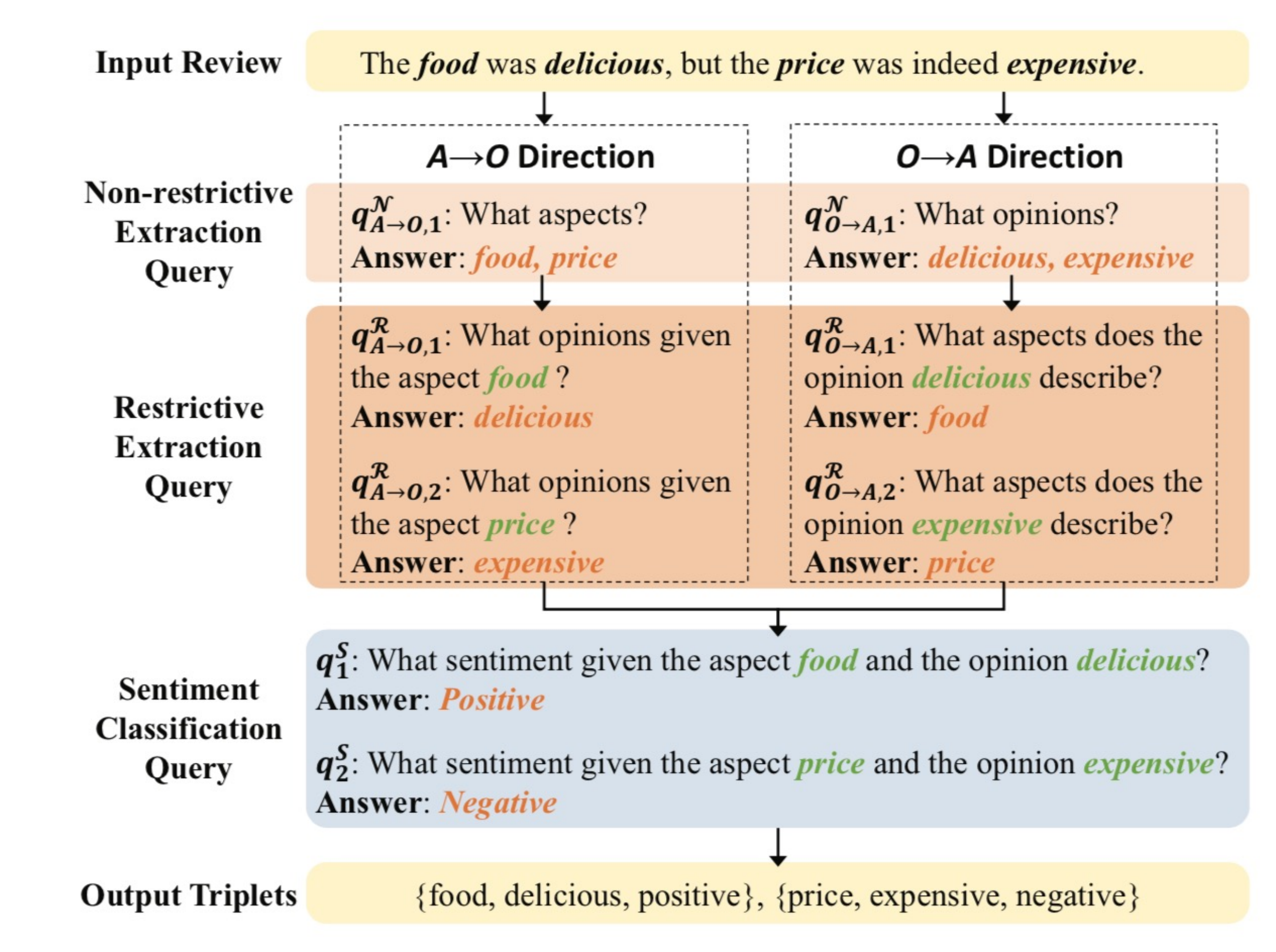
- Extract (aspect, opinion expression, sentiment) triplet from text



Chen et al. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. AAAI 2021.

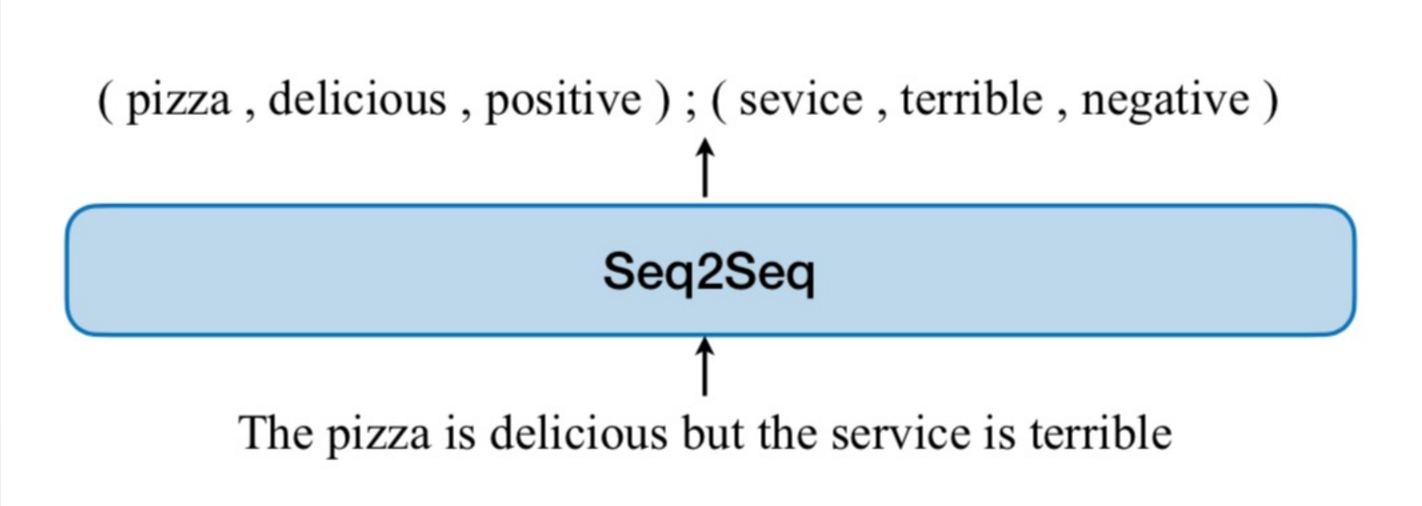
Machine Reading Comprehension for Aspect Sentiment Triplet Extraction

- Extract (aspect, opinion expression, sentiment) triplet from text



Generative Aspect-Based Sentiment Analysis

- Train T5 model for seq2seq generation of aspect-opinion-sentiment triplets.



Zhang et al. Towards generative aspect-based sentiment analysis. ACL 2021.

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

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

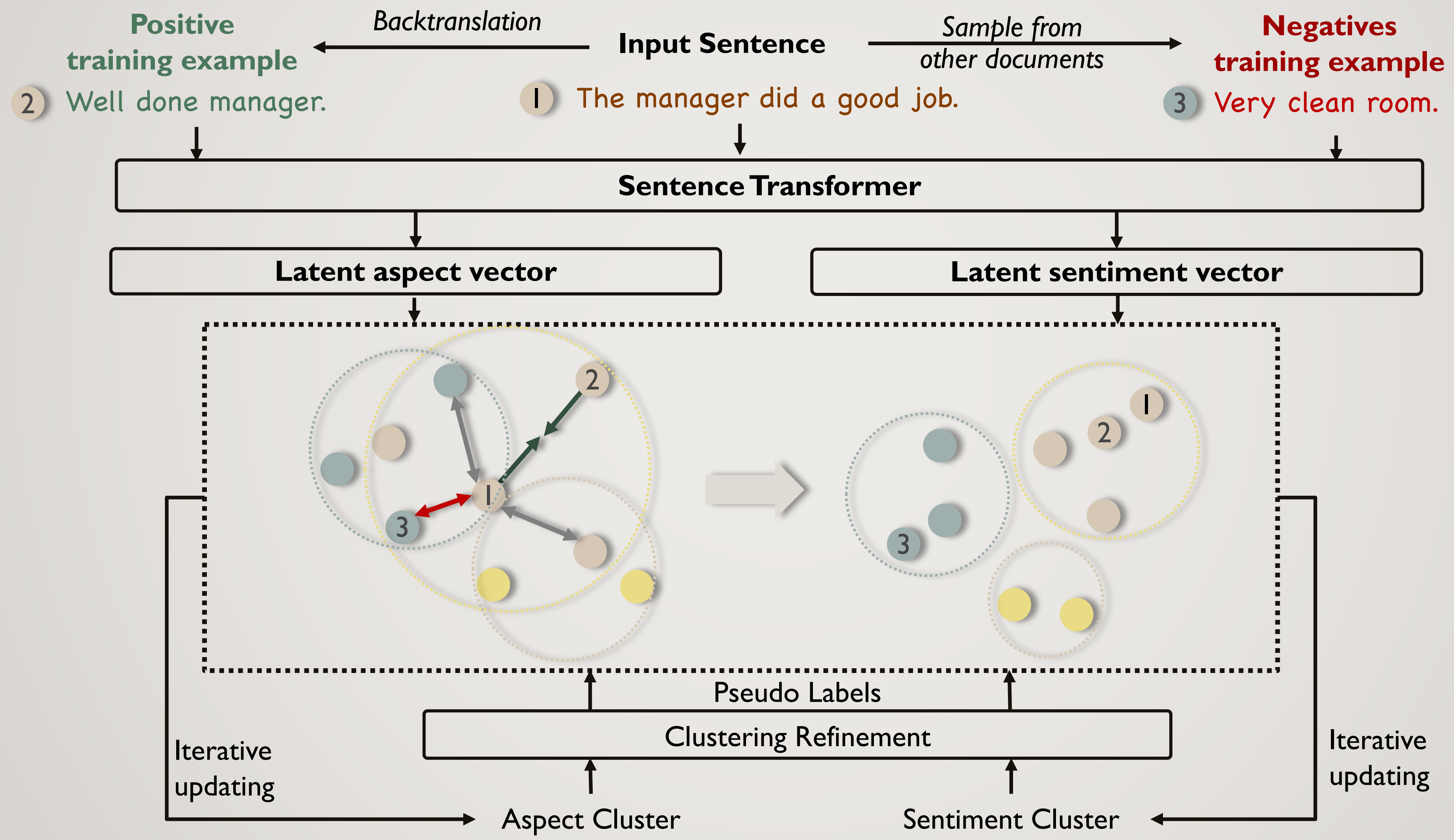
ABSA in domains beyond product reviews, e.g., patient reviews

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.

Unsupervised Contrastive Opinion Extraction

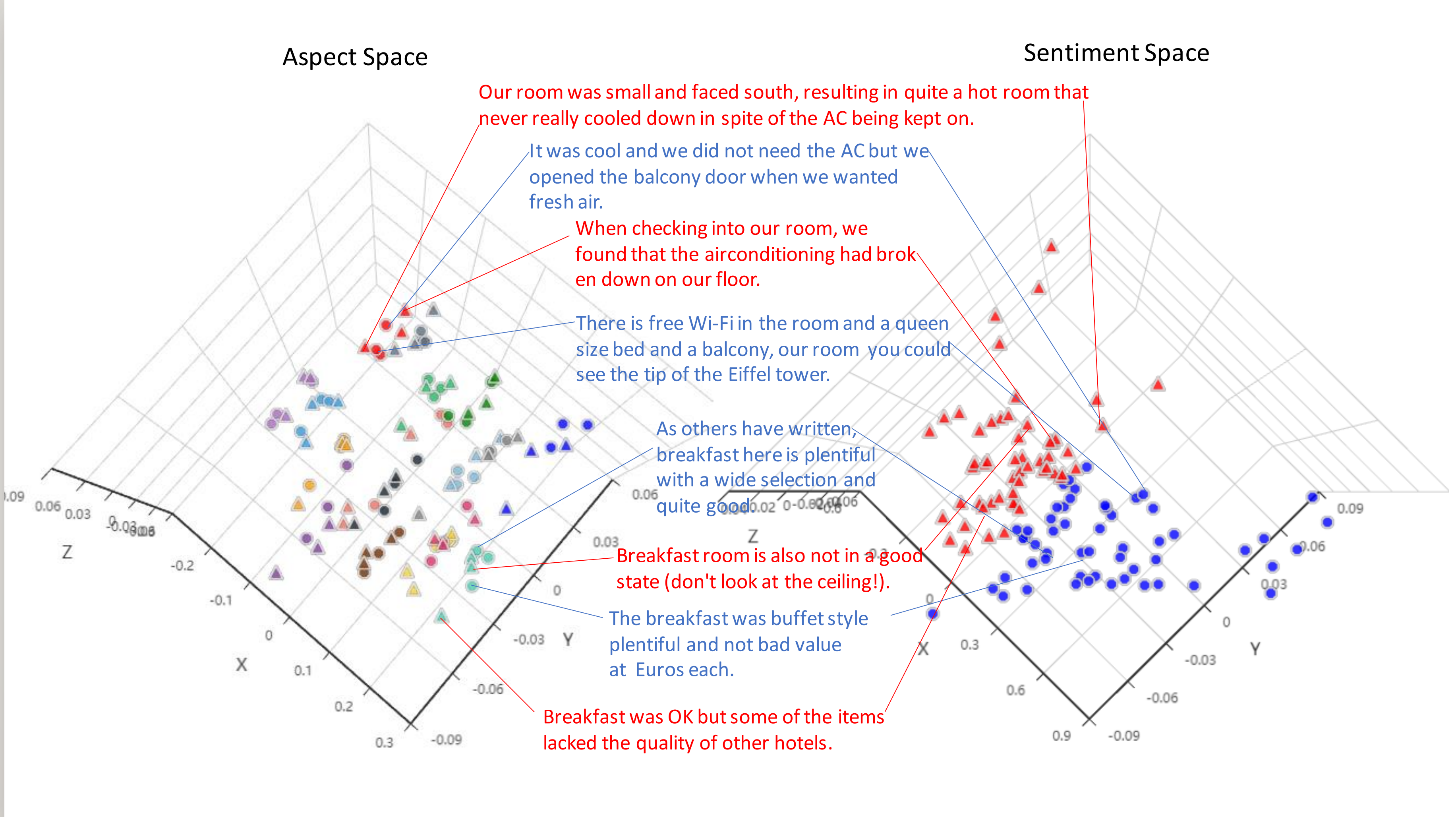
Aspect	Negative	Negative (%)	Positive	Positive (%)
Rooms	• Our room was small and faced south, resulting in quite a hot room that never really cooled down in spite of the AC being kept on. • When checking into our room, we found that the air conditioning had broken down on our floor.	16.6	• There is free Wi-Fi in the room and a queen size bed and a balcony, our room you could see the tip of the Eiffel tower. • It was cool and we did not need the AC but we opened the balcony door when we wanted fresh air.	83.4%
Services	• The hotel was completely full when we went and got very very busy for breakfast. • We reported my fall to the duty manager; he couldn't have cared less and other than having a non-slip mat put in the bath on the following morning ignored my plight totally.	21.5	• Great customer service and good restaurant service is what made this experience so wonderful for my family. • Good job to the manager Massimo (who knows how to hire great people)!	78.5

CONE: Contrastive OpinioN Extraction



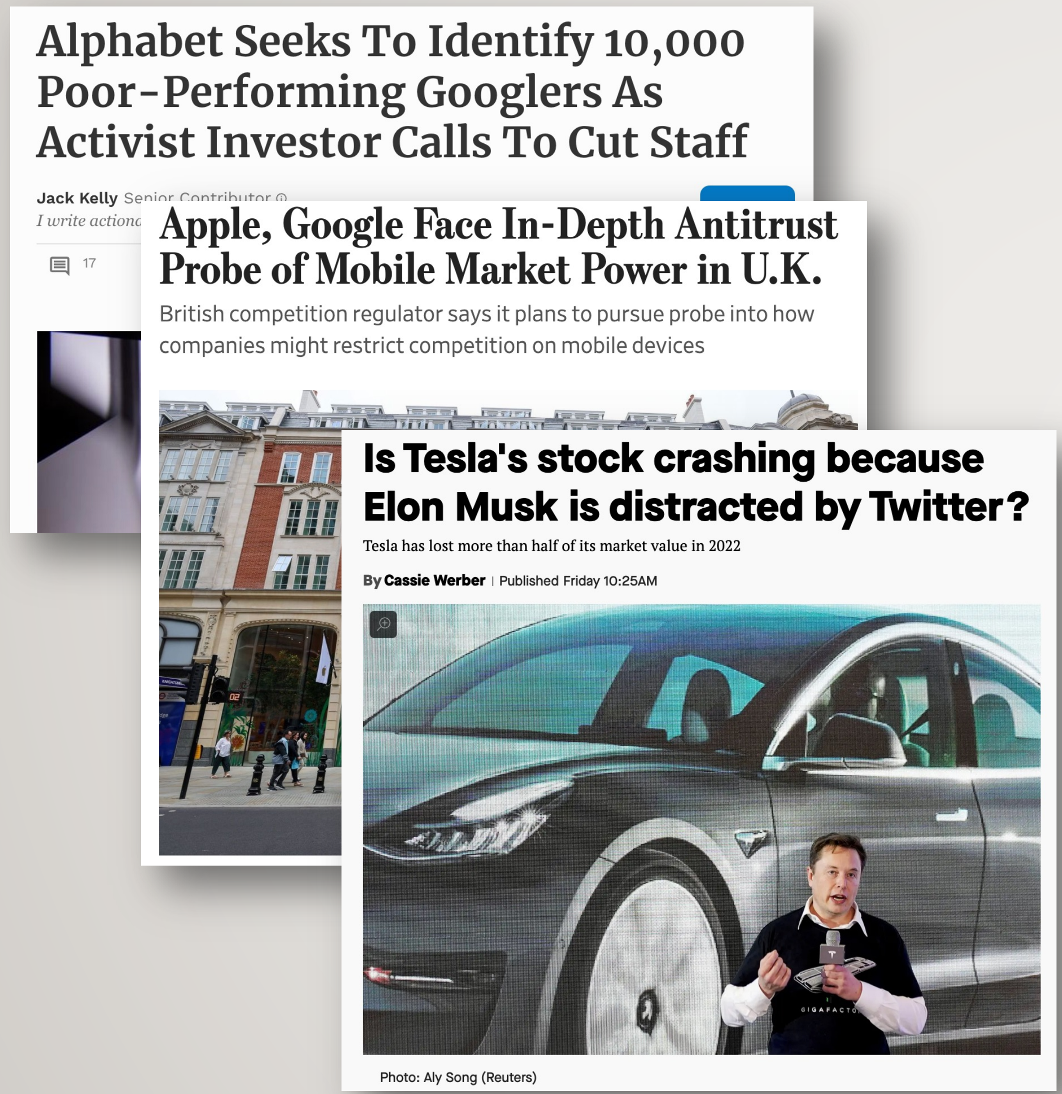
R. Zhao, L. Gui and Y. He. CONE: Unsupervised Contrastive Opinion Extraction, in submission.

CONE – Results on Hotel Reviews



End-to-end NLP Inference Approaches

Stock Prediction using News Events



- Dataset
  - **Financial news** from Reuters and Bloomberg (October 2006 – November 2013)
  - **S&P's 500 stock index** and its individual stocks, prices obtained from Yahoo Finance
- Frame stock prediction as a **binary classification** task
  - Use the news information in day $t-1$ to predict price movements of stock market in day t
  - Classification output:
    - $+1$ – the stock closing price will *increase* compared with the opening price in day t;
    - -1 – the stock closing price will *decrease* compared with the opening price in day t

Ding et al. 2019. Event representation learning enhanced with external commonsense knowledge. arXiv preprint arXiv:1909.05190.

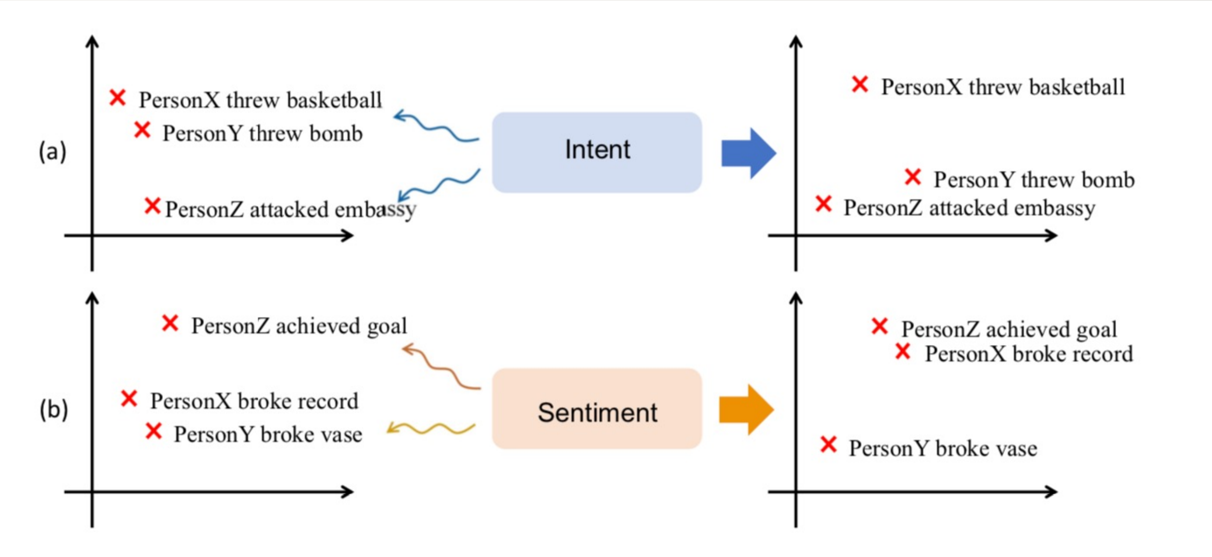
Stock Prediction using News Events

- Event extraction
  - Events are extracted from news using Open Information Extraction and dependency parsing.

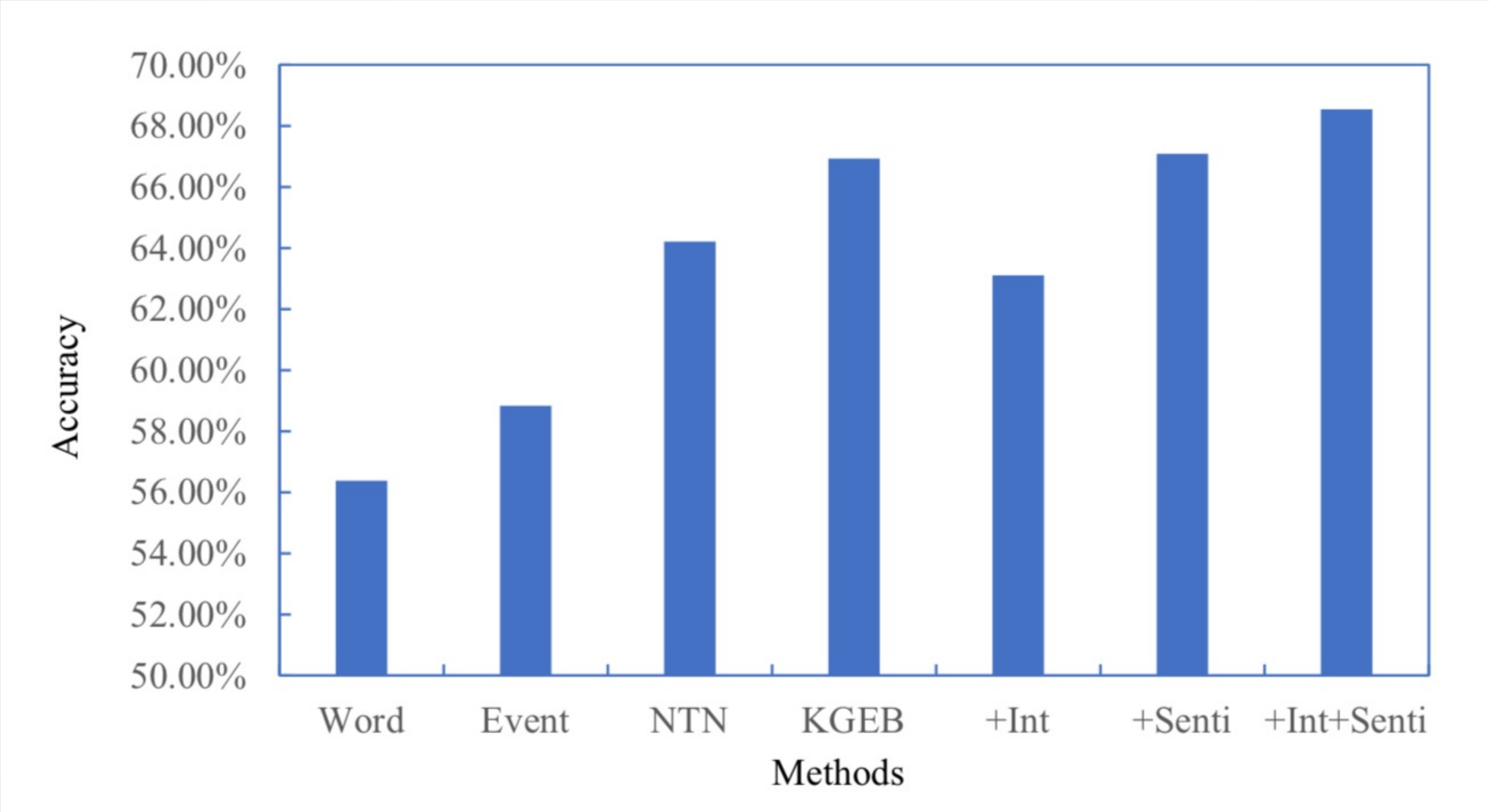
Sep 3, 2013 - Microsoft agrees to buy Nokia's mobile phone business for $7.2 billion

Actor = Microsoft
Action = buy
Object = Nokia's mobile phone business
Time = Sep 3, 2013

- Event representation learning
  - Incorporate information of sentiment and intent of an event from an external commonsense knowledge



S&P 500 Index Prediction Results



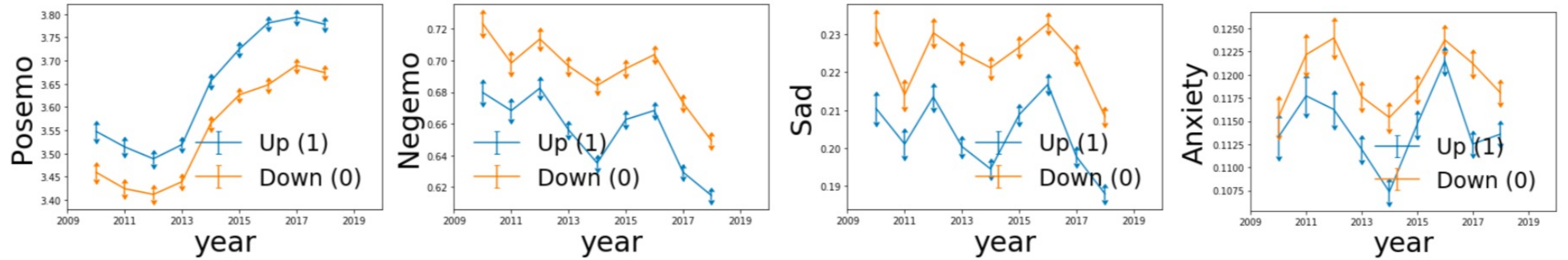
Ding et al. 2019. Event representation learning enhanced with external commonsense knowledge. arXiv preprint arXiv:1909.05190.

Stock Price Movement from Earnings Calls

- Data
  - 100,000 transcripts of earnings calls from 6,300 public companies (January 2010 – December 2019).
- Define the stock price movement as three different classification problem
  - **Input:** a company's earnings call transcript on day d
  - **Output:** the upward/downward movement of the company's stock price
    - Value-based label function, $y = \begin{cases} 1, & \text{if } s_{d+1} > s_{d-1} \\ 0, & \text{Otherwise} \end{cases}$

Sentiments and Stock Price Movements

- Sentiment analysis using the Linguistic Inquiry and Word Count (LIWC) dictionary
  - **Sentiment:** Positive / Negative
  - **Emotions:** Anxiety, Anger, Sad scores
  - **Personality features:** Certainty, Cognitive, Insight, Causation, Discrepancy scores



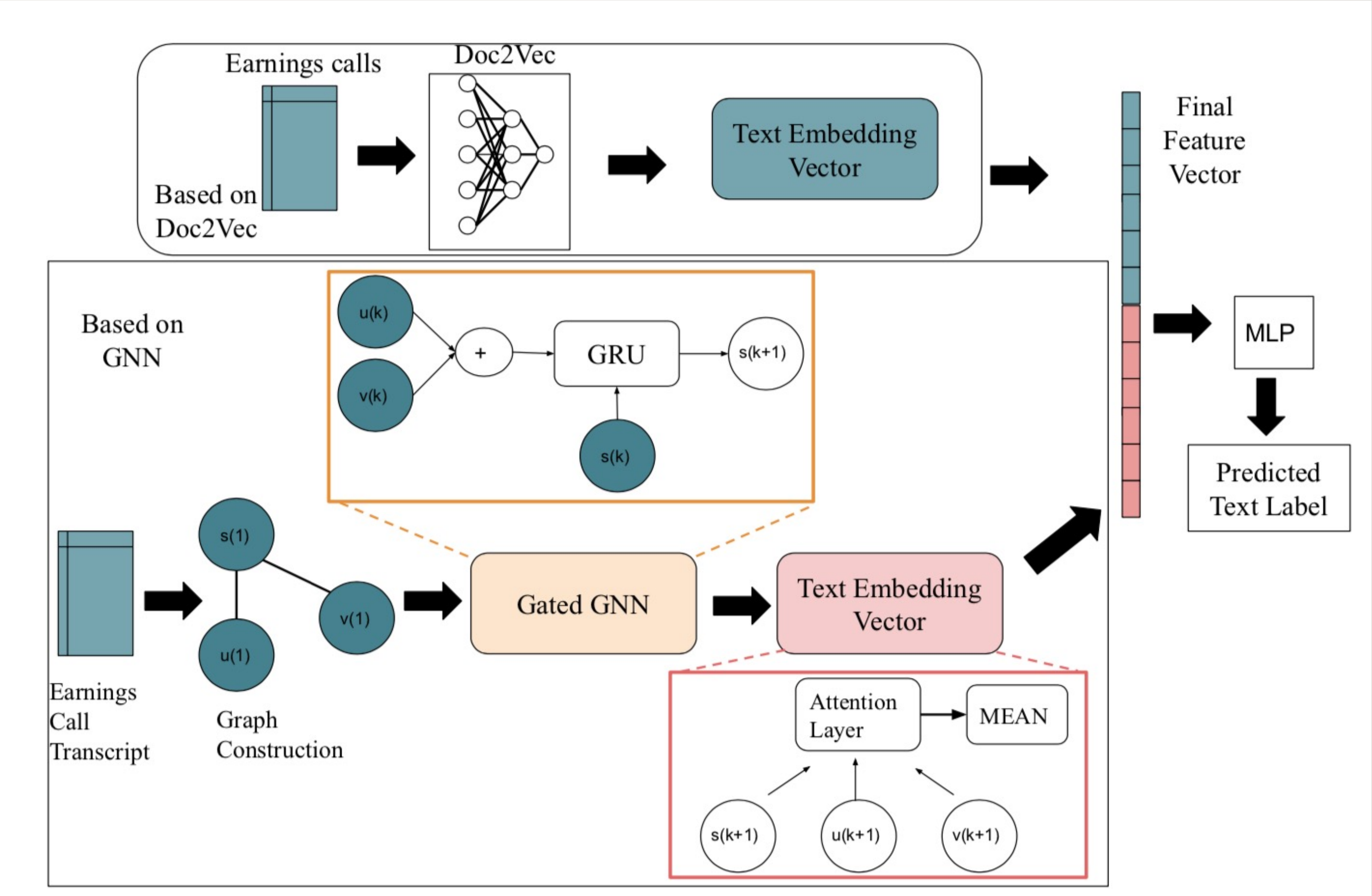
(a) Positive (b) Negative (c) Sad (d) Anxiety

Medya et al. An Exploratory Study of Stock Price Movements from Earnings Calls. WWW 2022.

Earnings Calls and Stock Price Movement Prediction

StockGNN

- Graph generation – each unique word in a document is a node and the words appear in its neighbourhood (or context) become its neighbours.
- Gated GNN to learn node embeddings.
- Combines GNN based embeddings with traditional context based Doc2Vec embeddings.



Results

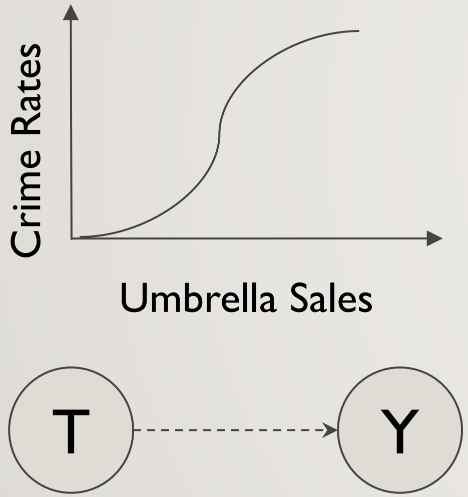
Measures	Accuracy					Avg. Precision					Avg. Recall				
Methods	Fin	Health	Mat	Service	Tech	Fin	Health	Mat	Service	Tech	Fin	Health	Mat	Service	Tech
DEsvm	.544	.582	.554	.567	.597	.54	.577	.555	.568	.597	.54	.579	.555	.567	.598
DElogreg	.55	.584	.556	.565	**.598**	.55	.58	.56	.57	.59	.54	.58	**.56**	.57	**.60**
DEmlp	.547	.552	.55	**.574**	.549	.54	.56	.55	**.58**	.55	.541	.55	.55	**.574**	.55
StockGNN	**.638**	**.606**	**.563**	.56	.55	**.62**	**.609**	**.562**	.56	**.603**	**.544**	**.608**	**.56**	.545	.562

Value-based label (y_v) results in five major sectors

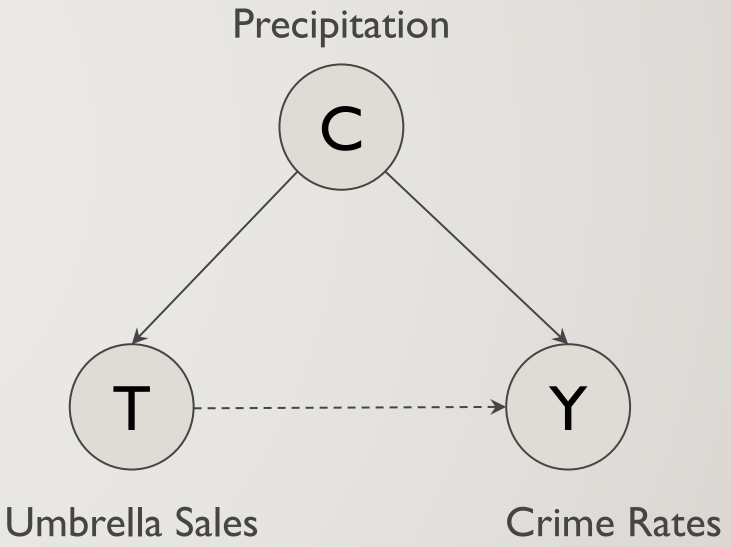
Medya et al. An Exploratory Study of Stock Price Movements from Earnings Calls. WWW 2022.

Motivation: Correlation $\neq$ Causation

- Increase in umbrella sales is correlated with increase in crime



Common cause: Precipitation



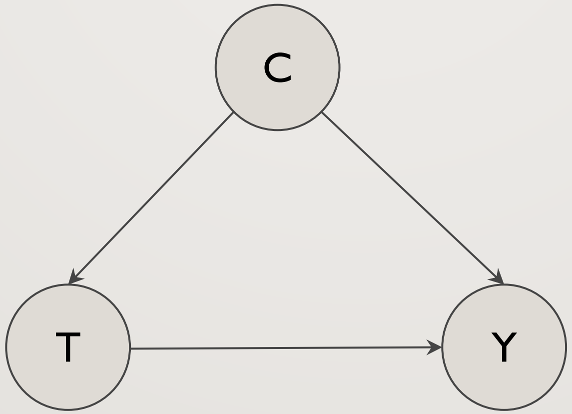
T:	Umbrella Sales	Y:	Crime Rates	C:	Precipitation

Estimating the Causal Effect between Political Risk and Stock Volatility

- Directly measuring the political risk faced by a company is difficult
  - The political risk can be quantified based on the discussions centred on risks associated with politics-related topics in the earnings call (Hassan et al., 2019).
- The firm-level stock volatility is calculated as the variance of the 7-day stock price after the earnings call.

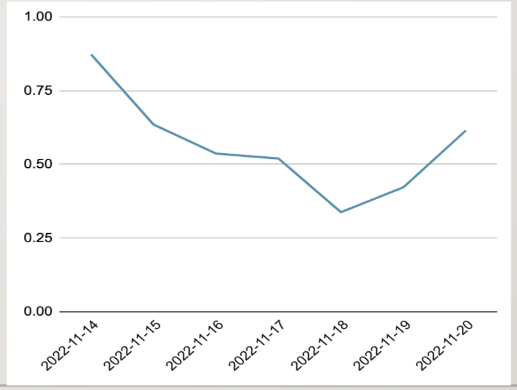
Earnings Call

Delay in defense budget and deferral of several other government programs, unpaid government bills.......



High political Risk: T=1
Low political Risk: T=0

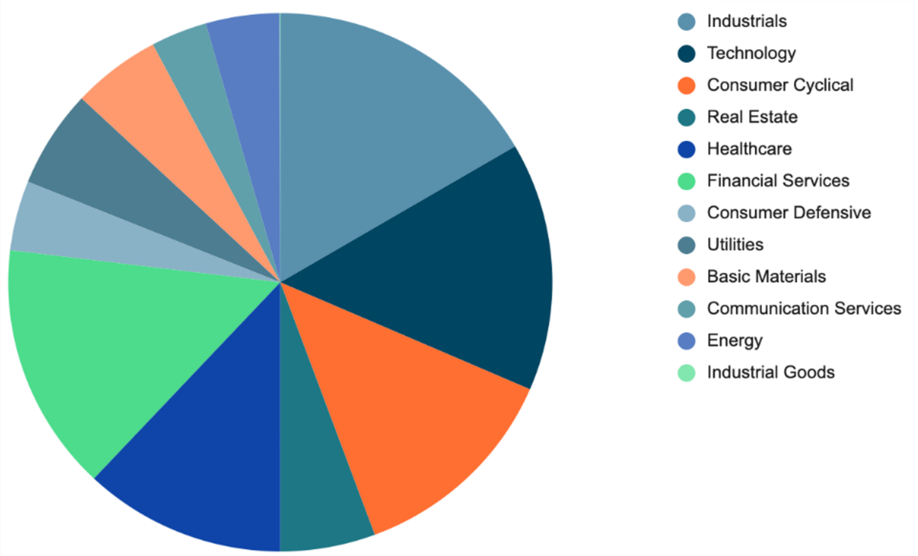
7 day Stock volatility



Datasets

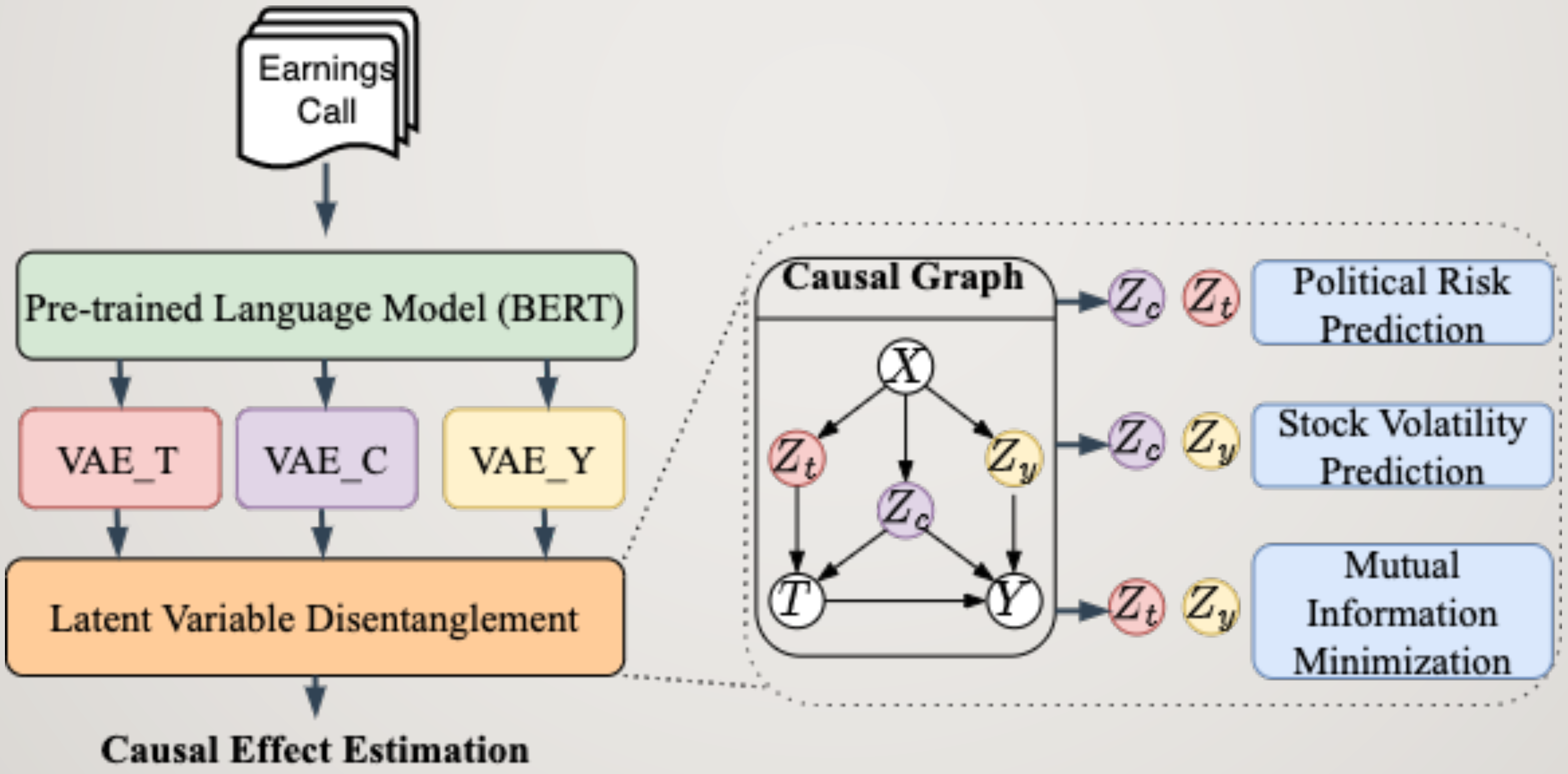
- Downloaded 115,880 transcripts of earnings calls happened between May 2001 to Oct 2019.
  - 1438 companies in 12 different sectors.
- Infer the political risk score from each transcript and select the 10,000 transcripts with the highest and lowest scores as high political risk $(T = 1)$ and low political risk $(T = 0)$ transcripts, respectively.
- Since ground-truth causal effects are unavailable, we generate semi-synthetic data by simulating the outcome.

Statistics	**EarningsCall**
High political Risk (T=1)	10,000
High political Risk (T=0)	10,000
Average Length	1692.71
Training set	12,000
Validation Set	2,000
Test Set	6,000

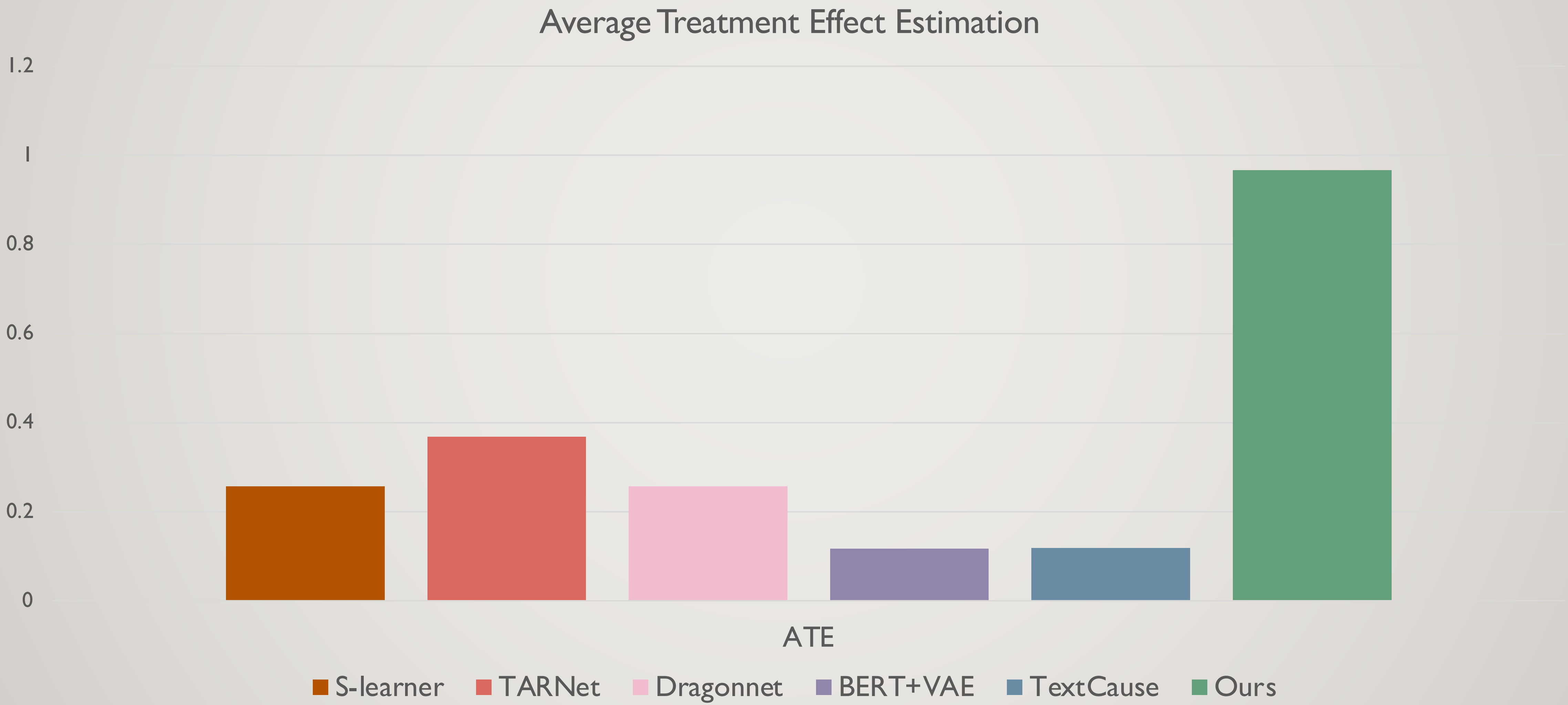


Model: Estimating Causal Effect with Disentangled Variables

- Multi-task learning to disentangle different latent variables that affect treatment, outcome, and both treatment and outcome simultaneously.
- Minimize the mutual information between the latent instrumental variable z_t and latent outcome variable z_y to make them as independent as possible.



Results



Average Treatment Effect Estimation
1.2
1
0.8
0.6
0.4
0.2
0
ATE
S-learner
TARNet
Dragonnet
BERT+VAE
TextCause
Ours

What is Next?

Future Directions

- NLP benchmarking in Finance
- New evaluation frameworks
- Causality and NLP
- Model interpretability

NLP Benchmarking in Finance

GLUE

SuperGLUE

GEM BENCHMARK

BIG-bench

NLP Tasks in FinTech

- Event Extraction
  - Commodity news corpus
  - SentiFM dataset
- Sentiment Analysis
  - Financial PhraseBank
  - FiQA
  - StockSen
- Causality Detection
  - FinCausal shared task
- Numeral Understanding
  - FinNum shared task
- Question-Answering
  - FinQA
  - TAT-QA

Limited Annotated Data

Financial Sentiment Analysis Datasets	Training set Sentences	Test set Sentences
Financial PhraseBank	4845	
FiQA Sentiment – Financial news headlines	436	93
FiQA Sentiment – Financial microblogs	675	99

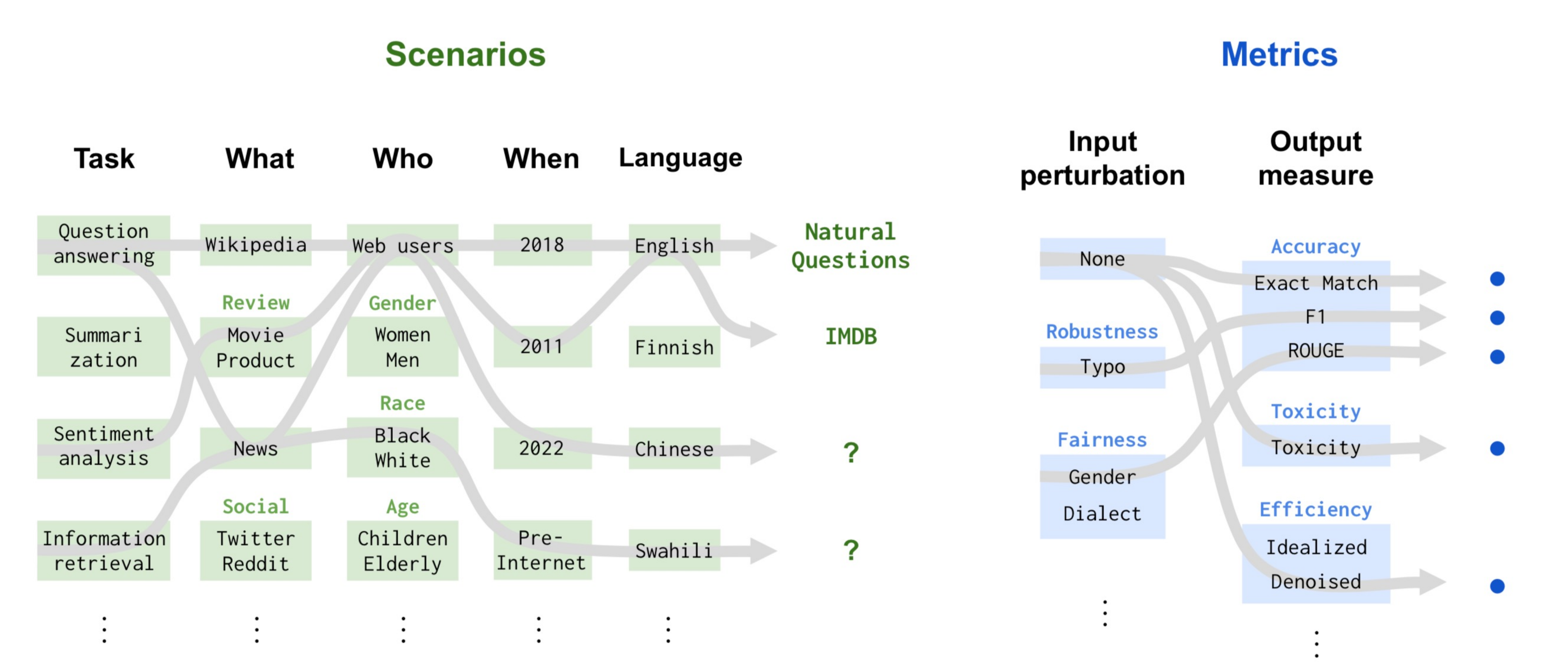
Zero-shot and few-shot learning

Model adaptation and transfer learning

Continual learning

New Evaluation Frameworks

- Holistic Evaluation of Language Models (HELM)



Liang, et al. 2022. Holistic Evaluation of Language Models. arXiv preprint arXiv:2211.09110.

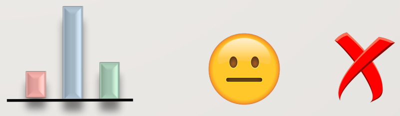
Robustness Evaluation Framework

- Static data quickly becomes obsolete
- Update models with adversarial or synthetic examples
- Dynamic adversarial data collection

DynaBench

Robustness Gym

I wonder if episodes had quality control before broadcasting.



Predicted label

Causality and NLP

Causal information extraction

Detection of cause and effects from text

Detection of causal relations of two events from text

Causal graph construction

Causal question-answering

Counterfactual question-answering

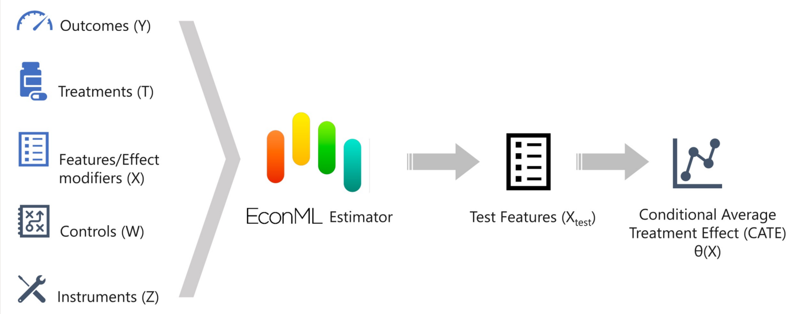
Causal effect estimation with text

Text as confounders

Text as outcome

Text as treatment

https://github.com/uber/causalml



https://www.microsoft.com/en-us/research/project/econml/

Causal Reasoning for NLP

Incorporating causal knowledge between observations and output labels.

Counterfactual example generation

Disentangled learning of domain-invariant and domain-dependent features

Causal analysis for fairness and bias

Causal analysis can be used to identify the fairness properties of an observed distribution of data and predictions.

Counterfactual data augmentation can be used to reduce bias in pre-trained language models.

Counterfactual explanations

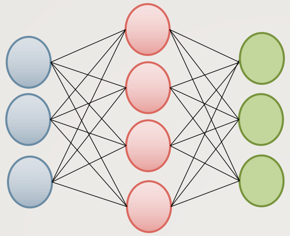
Counterfactual generation can be done at the feature representation level.

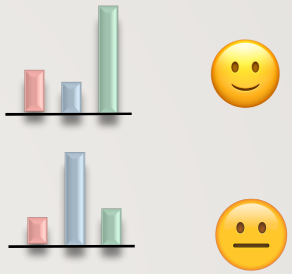
Explanation can be performed by comparing predictions for each input instance and its generated counterfactual.

Model Interpretability

The food is delicious!

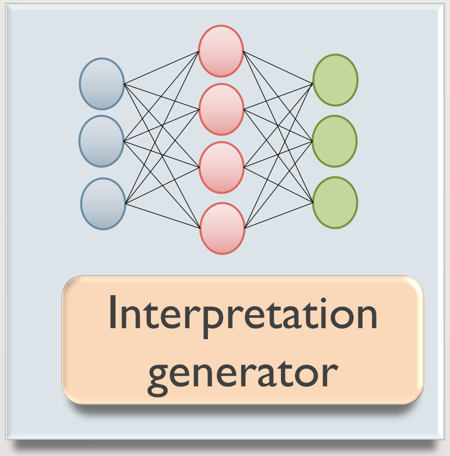
The food is [MASK]!



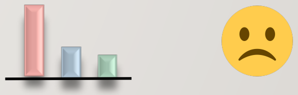


Predicted label

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.

Model Interpretability

Interpretation beyond word- and phrase-level

Uncertainty interpretation

Conversational XAI

QUESTIONS

yulan.he@kcl.ac.uk

Twitter: @yulanhe