Domain-Adaptive Sentiment Analysis Across Online Social Networks

Jingli Shi (19091064)

Supervisors & Advisor:

Dr. Weihua Li, AUT

Dr. Sira Yongchareon, AUT

A/Prof. Jian Yu, AUT

A/Prof. Quan Bai, UTAS

AUT UNIVERSITY NEW ZEALAND

UNIVERSITY of TASMANIA AUSTRALIA

Outline

- ❑ Introduction
  - Research Background
  - Literature Review
  - Research Limitations & Questions
- ❑ Design of Study
  - Research Methodology
  - Proposed Method
  - Datasets Analysis
  - Evaluation Metrics
- ❑ Current Progress
  - Aspect and Relation Extraction
  - Feature Selection-based Aspect Extraction
- ❑ Timetable for Completion
- ❑ Publications

Research Background (1/3)

- **What is Sentiment Analysis?**

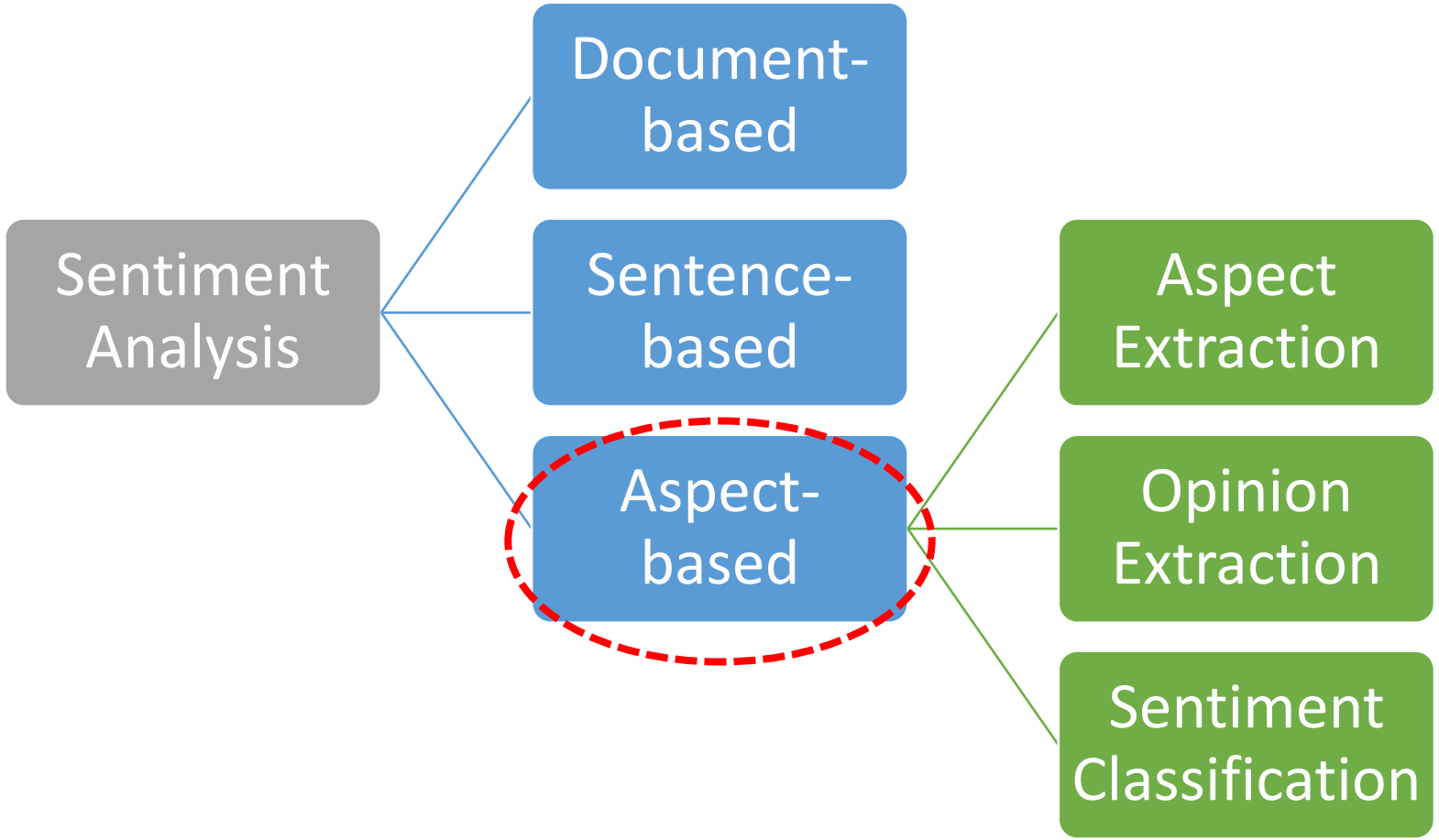
  Sentiment analysis aims to apply natural language processing (NLP), text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

- **Why Sentiment Analysis is important?**
  - **Business and Organization**: benchmark products and services; place related products based on customer opinion.
  - **Individuals**: make decision to buy products or services;

- **What Sentiment Analysis can do?**

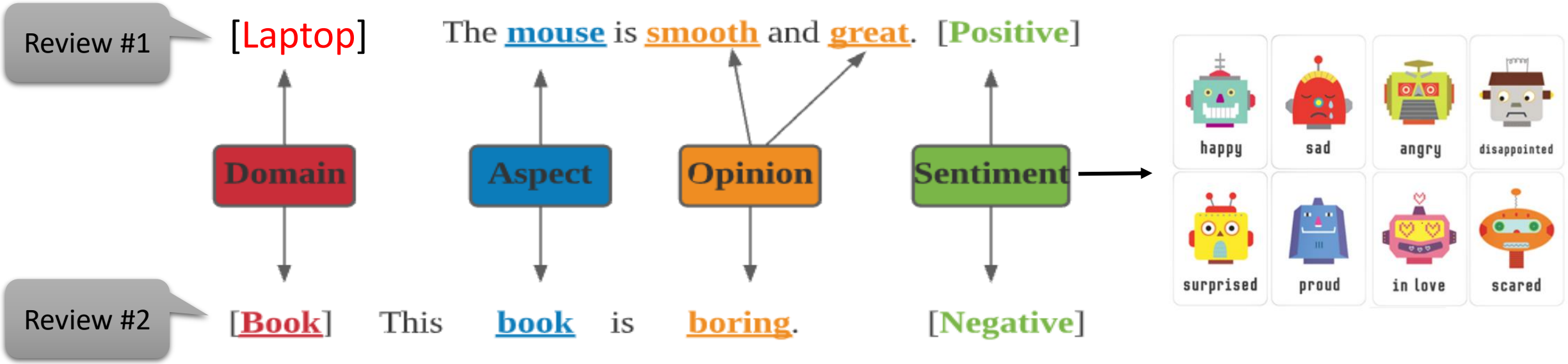
  Social media monitoring, customer support management, customer feedback, brand monitoring and reputation management, voice of customer, voice of employee, product analysis, market research and competitive research, etc.

Research Background (2/3)

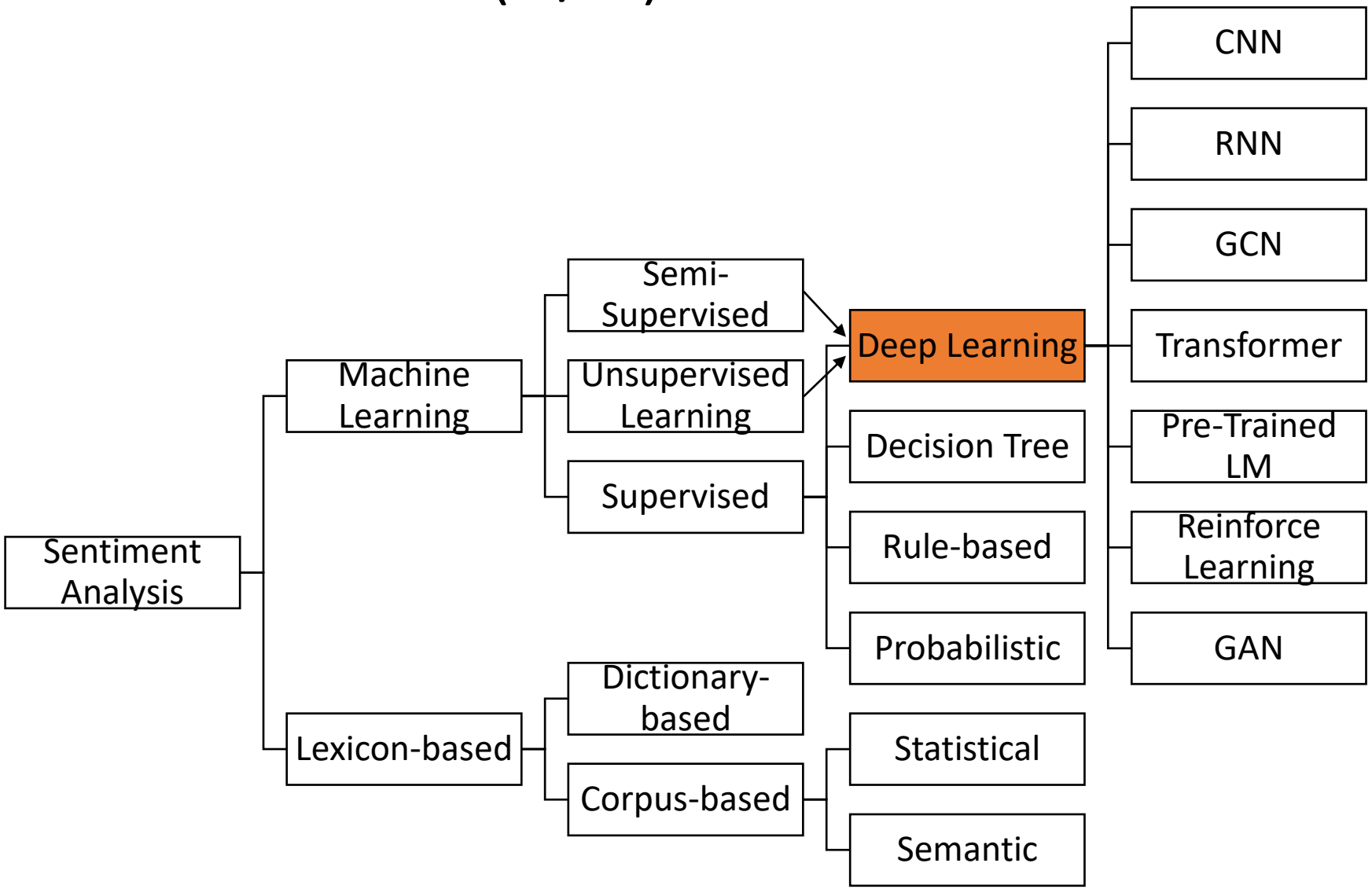


Research Background (3/3)

- Aspect Extraction
- Opinion Extraction
- Sentiment Classification



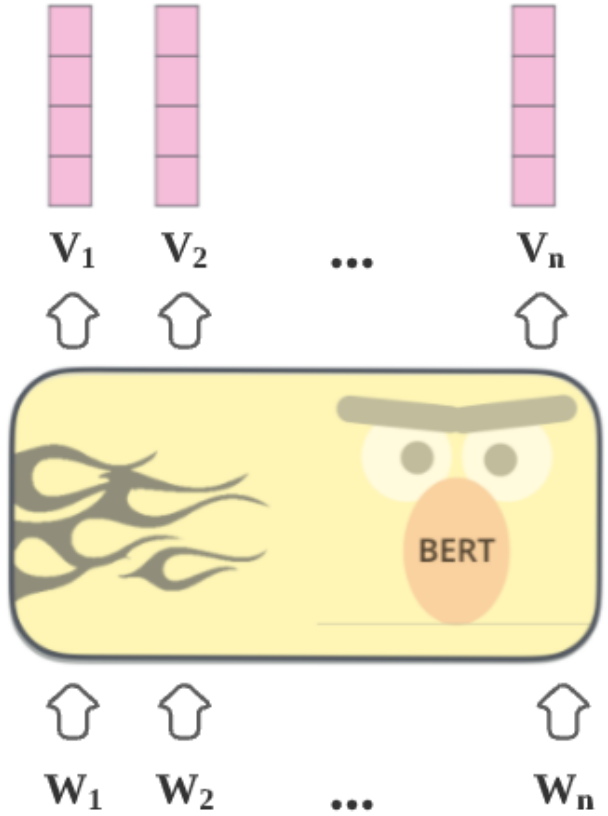
Literature Review (1/7)



Literature Review (2/7)

❑ Aspect Extraction

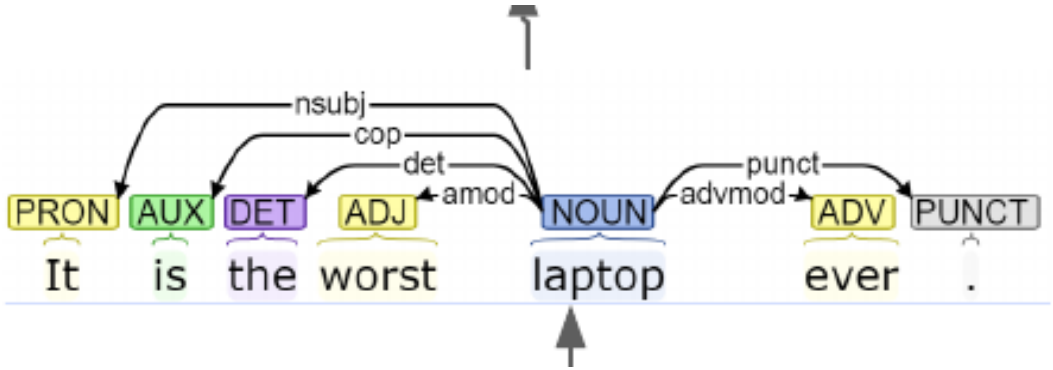
- Pre-Trained LM based Methods
- Graph based Methods

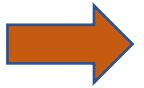


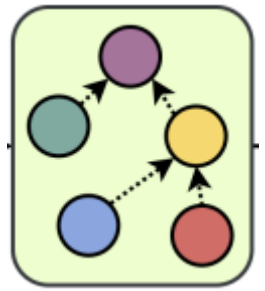
Literature Review (3/7)

❑ Aspect Extraction

- Pre-Trained LM based Methods
- Graph based Methods



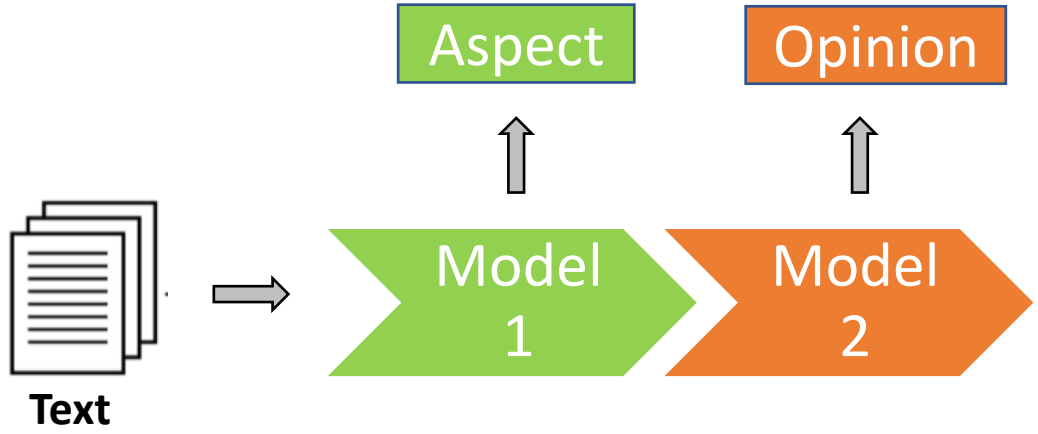




Literature Review (4/7)

❑ Aspect-Opinion Extraction

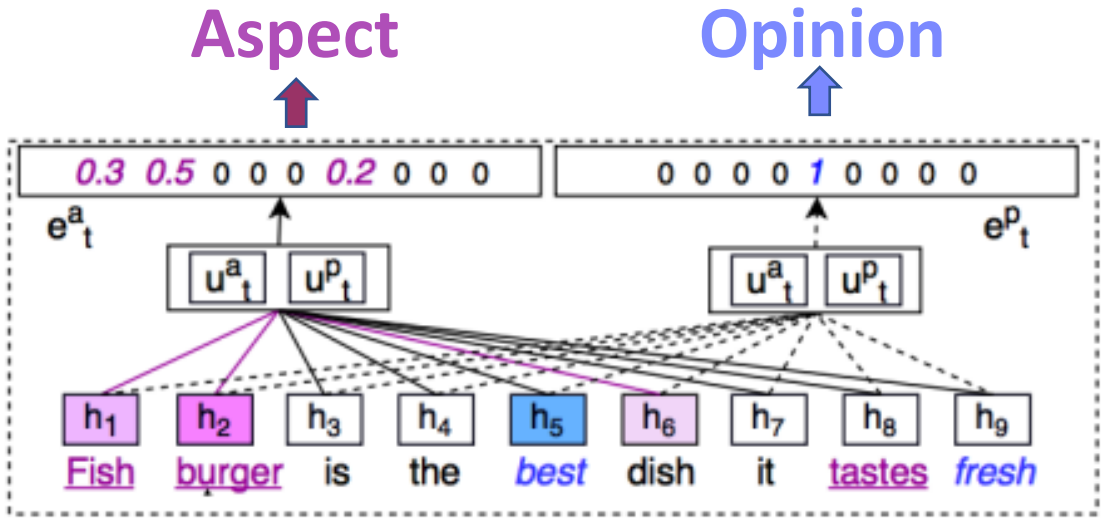
- Pipeline based Methods
- Co-extraction based Methods



Literature Review (5/7)

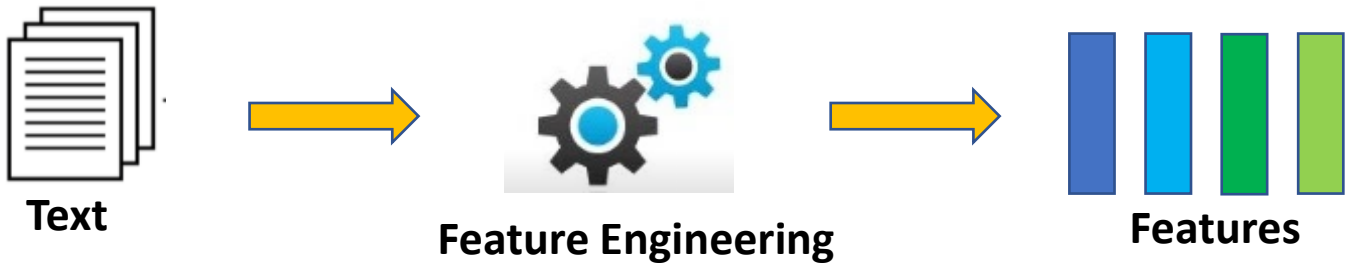
❑ Aspect-Opinion Extraction

- Pipeline based Methods
- Co-extraction based Methods



Literature Review (6/7)

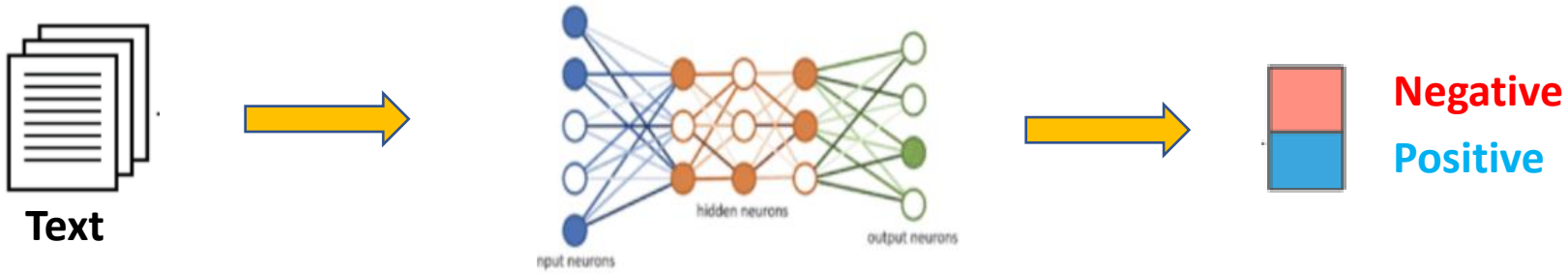
- Cross-Domain Sentiment Classification
  - Feature Engineering based Methods
  - Deep Learning based Methods



Literature Review (7/7)

❑ Cross-Domain Sentiment Classification

- Feature Engineering Methods
- Deep Learning-based Methods



Deep Learning Model

Research Limitations & Questions (1/3)

❑ Aspect Extraction

- Ignore Important Features
  - Linguistic features (i.e., lemma, tag, dep, shape, etc.)
  - Inherent structure (i.e., relation between aspects)
- Manually labelled Dataset

❑ RQ1

How to effectively extract contextual and linguistic features for aspect extraction in sentiment analysis?

Research Limitations & Questions (2/3)

❑ Aspect-Opinion Extraction

- Extract aspect and opinion individually
- Manual Annotation of Syntactic Info

❑ RQ2

How to automatically generate syntax structure information and learn useful syntactic representations for aspect-opinion extraction?

Research Limitations & Questions (3/3)

❑ Cross-Domain Sentiment Classification

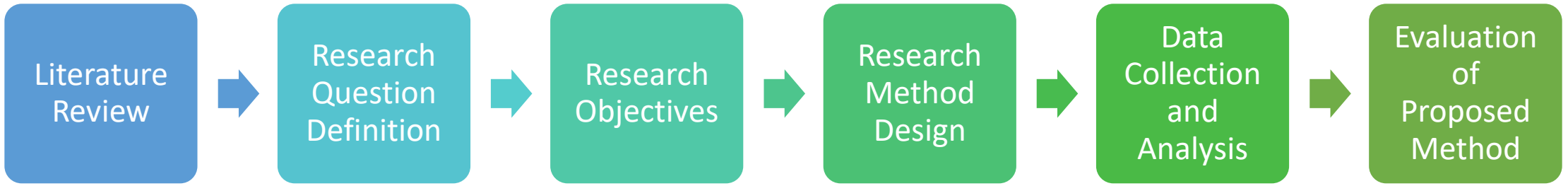
- Ignore Important Info in Target Data
- Ignore Sentiment-Specific Features

❑ RQ3

How to effectively extract domain-specific and domain-invariant features for cross-domain sentiment classification?

[Domain: Restaurant]	“fast” ➔ “service”	(**Positive**)
[Domain: Laptop]	“fast” ➔ “CPU”	(**Positive**)
	“fast” ➔ “power consumption”.	(**Negative**)

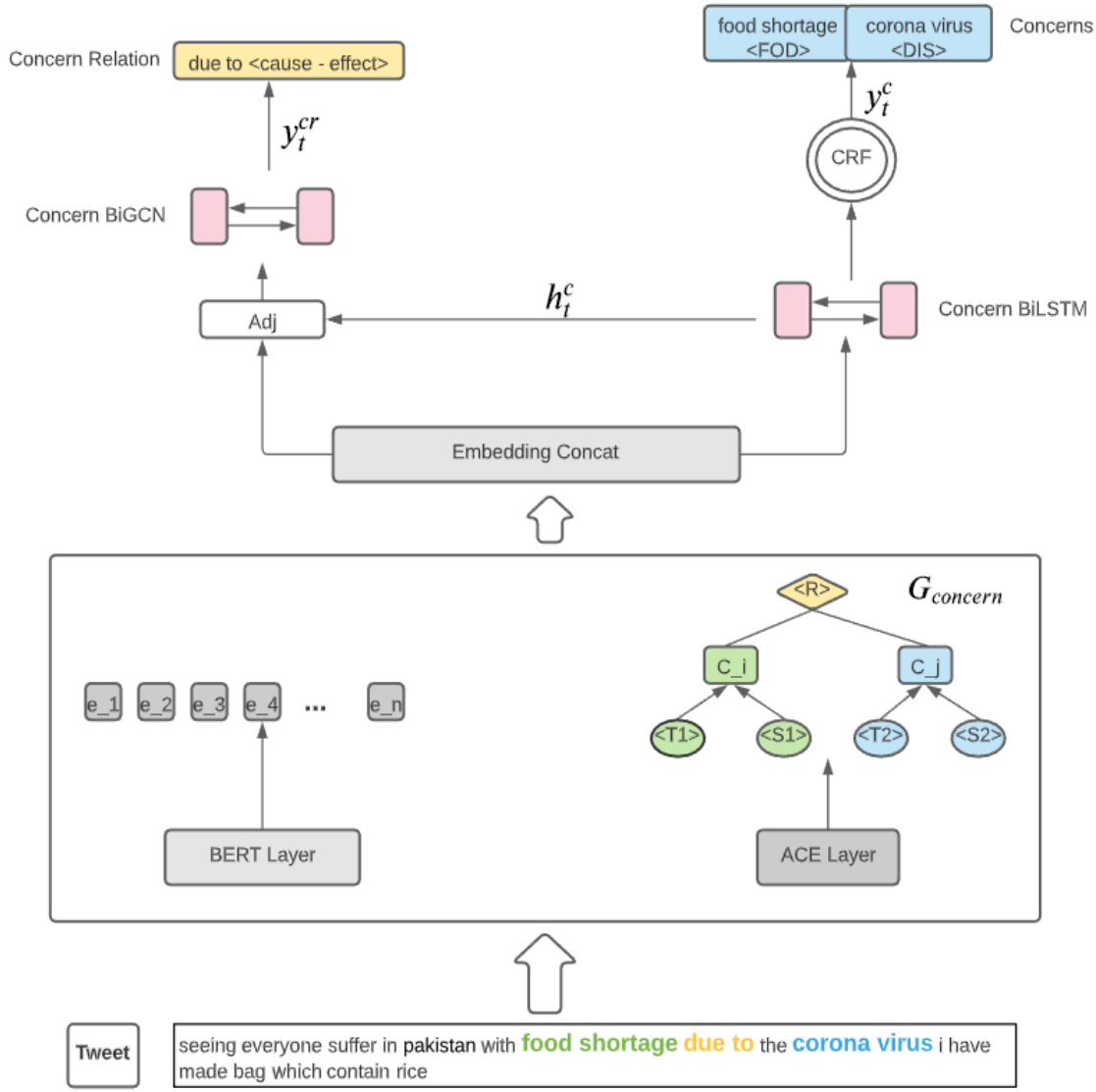
Research Methodology



RQ1:Graph-based Aspect and Relation Extraction

❑Limitations in RQ1:

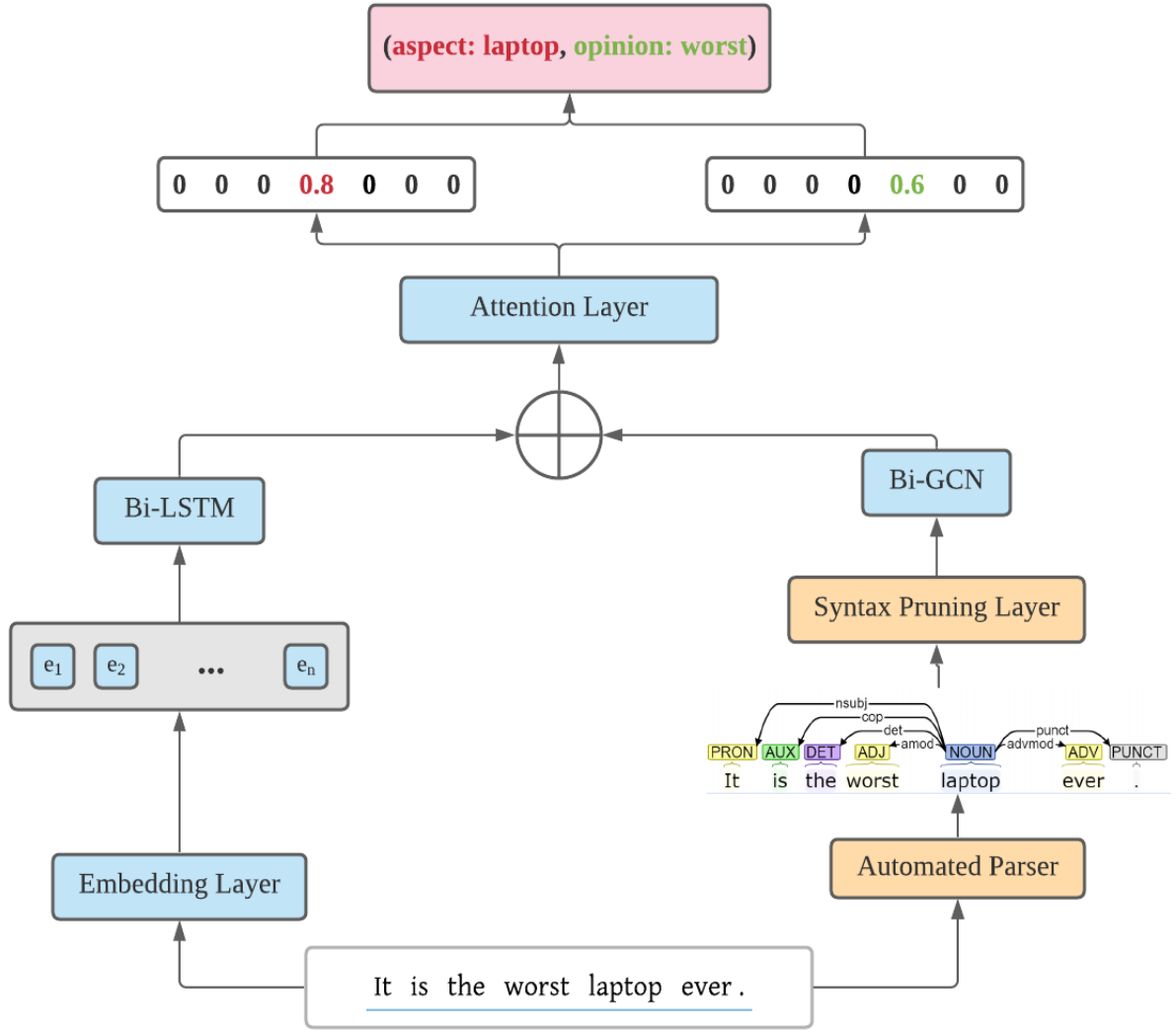
- Ignore Important Features
  - Linguistic features
  - Inherent structure **[CG[1]]**
- Manually labelled Dataset **[ACE[2]]**



1. **Jingli Shi**, Weihua Li, Sira Yongchareon, Yi Yang, Quan Bai. Graph-based Joint Pandemic Concern and Relation Extraction on Twitter. Expert Systems with Applications (2021) (submitted)
2. **Shi, J., Li**, W., Yang, Y., Yao, N., Bai, Q., Yongchareon, S., & Yu, J. (2021). Automated Concern Exploration in Pandemic Situations-COVID-19 as a Use Case. 17th Pacific Rim Knowledge Acquisition Workshop, PKAW 2020, Yokohama, Japan, January 7–8, 2021.

RQ2: Aspect-Opinion Pair Extraction

- Limitations in RQ2:
  - Extract aspect and opinion individually **[AO Pair]**
  - Manual Annotation of Syntactic Info **[Automated Parser]**



RQ3: Cross-Domain Sentiment Classification

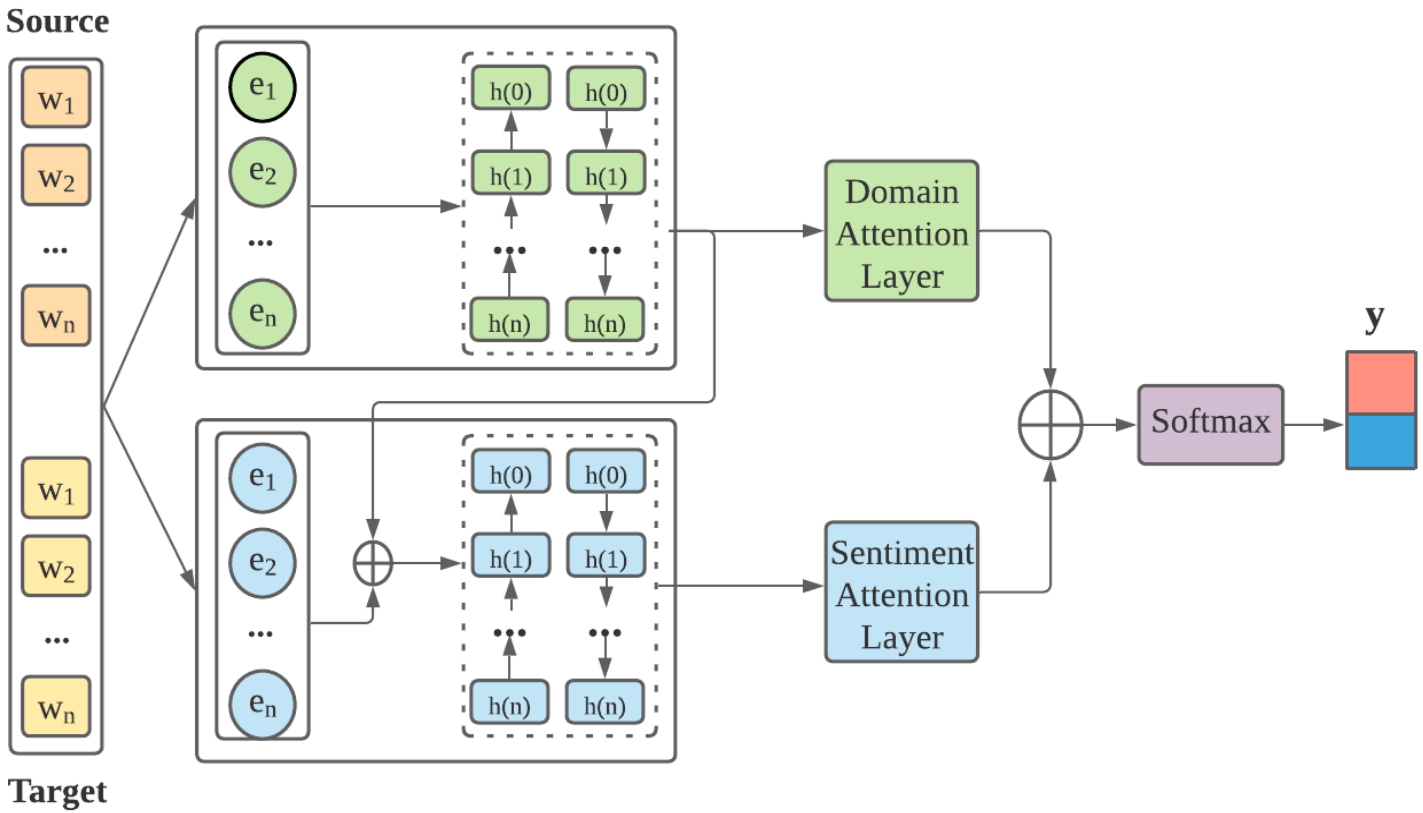
❑ Limitations in RQ3

- Ignore Important Info in Target Data

  [Domain Encoder & Attention]

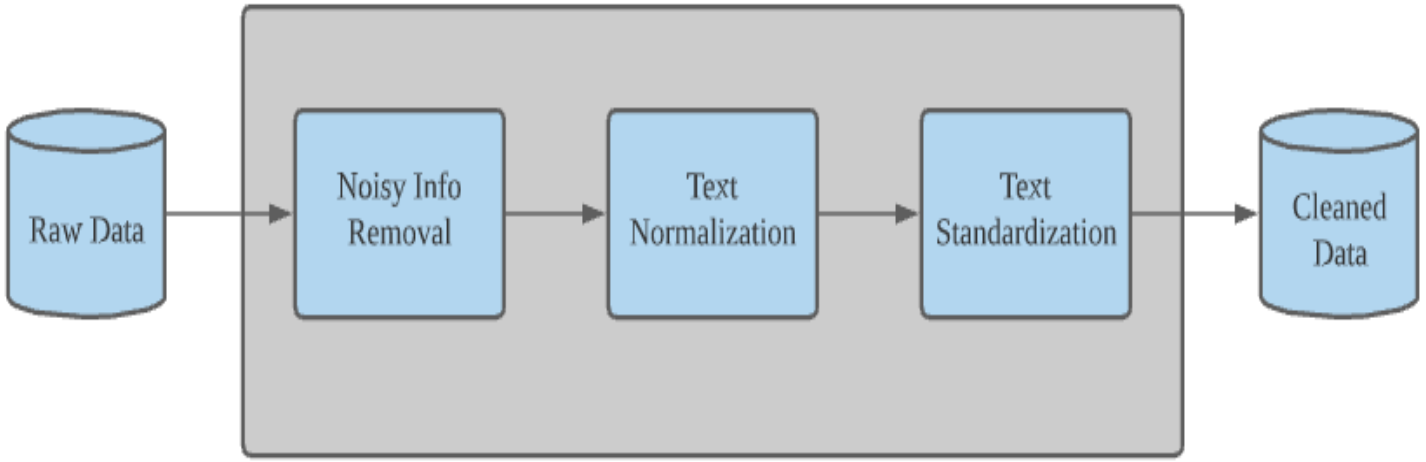
- Ignore Sentiment-Specific Features

[Sentiment Encoder & Attention]

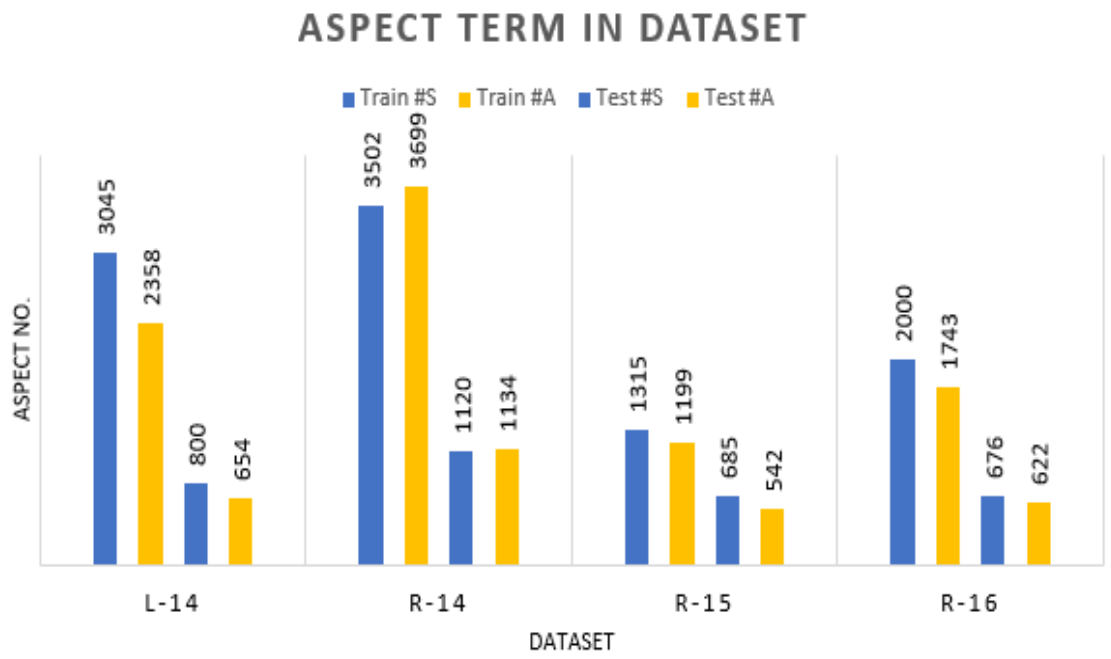


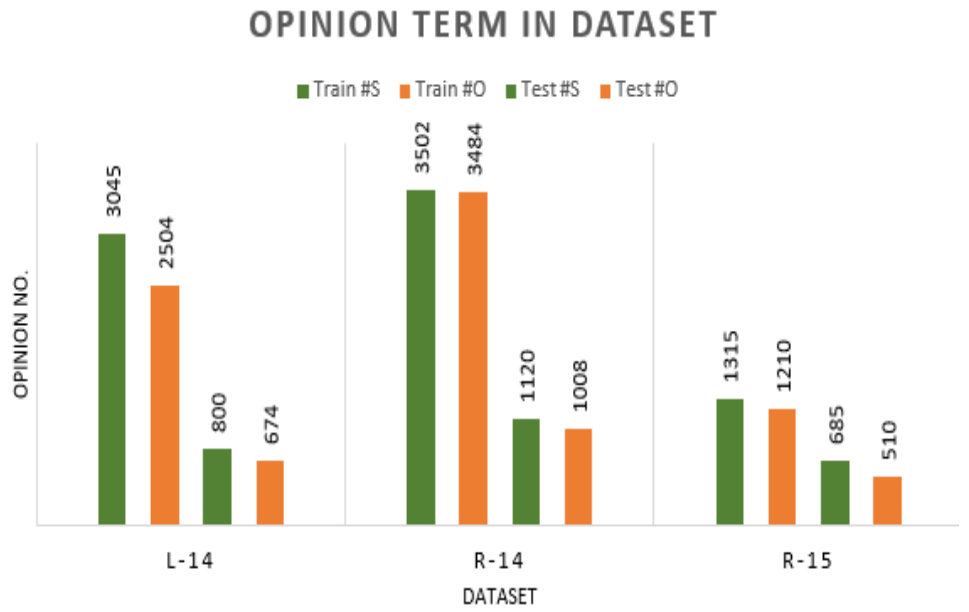
Dataset Analysis (1/3)

	SemEval-2014			SemEval-2015		SemEval-2016		Other Dataset		
	Restaurant	**Laptop**	**Twitter**	**Restaurant**	**Laptop**	**Restaurant**	**Laptop**	**Yelp**	**Amazon**	**Twitter**
Train	3502	2313	6248	1315	1739	2000	3045	10k	10k	30k
Test	1120	638	692	685	277	676	800	10k	10k	8k



Dataset Analysis (2/3)





Dataset Analysis (3/3)

Dataset		Positive	Negative	Neutral
SemVal-2014 (Restaurant)	Train	2164	805	633
	Test	728	196	196
SemVal-2014 (Laptop)	Train	987	866	460
	Test	341	128	169
SemVal-2014 (Twitter)	Train	1561	1560	3127
	Test	173	173	346
SemEval-2015 (Restaurant)	Train	912	256	36
	Test	326	182	34
SemEval-2016 (Restaurant)	Train	1240	439	69
	Test	469	117	30
Yelp	Train	10k	10k	10k
Amazon	Train	10k	10k	10k
Twitter	Train	10k	10k	10k

Dataset	Domains
Amazon	Book, DVD, Electronic, and Kitchen Appliance
SemEval	Laptop, Restaurant, Twitter

Evaluation Metrics

Precision (P): the number of true positive (TP) results divided by the sum of true positive (TP) and false positive (FP) results predicted by the method

Recall (R): is the number of true positive results divided by the sum of true positive and false negative results

F1 Score: is calculated using precision and recall

$$P = \frac{TP}{TP + FP}$$

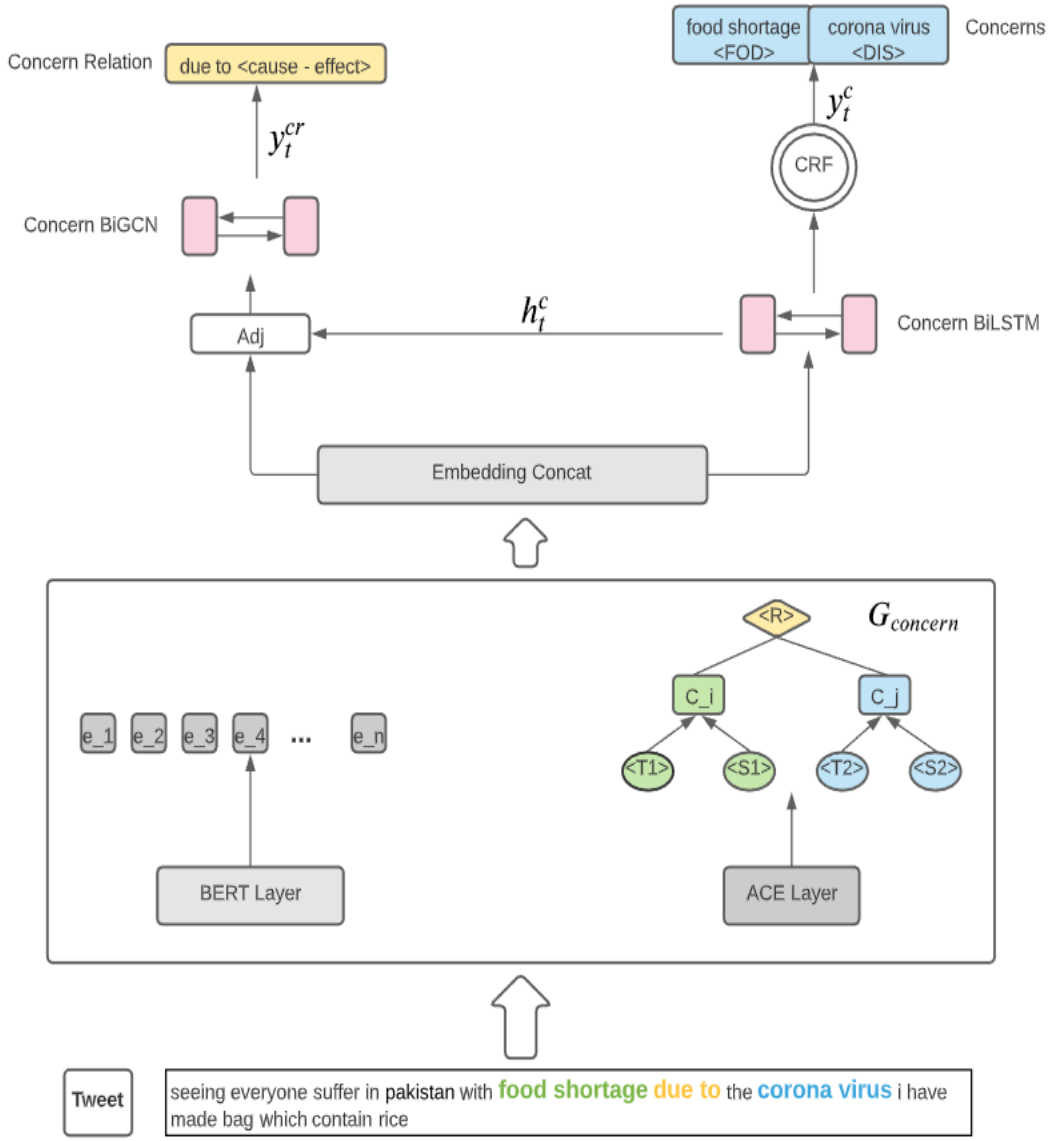
$$R = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{P * R}{P + R}$$

Current Progress - Graph-based Concern and Relation Extraction

RQ 1:

- Inherent Structure (CG)
- Feature Synchronization



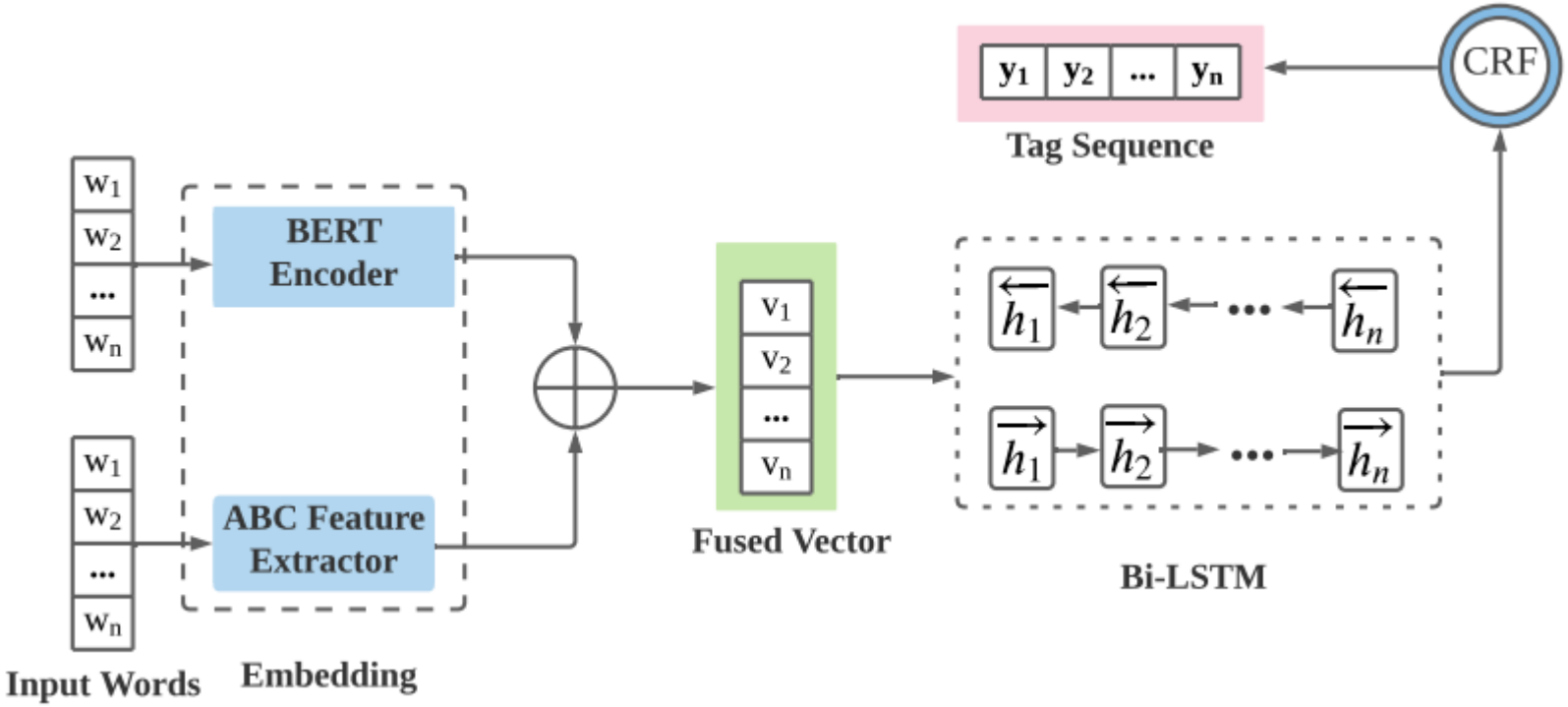
Experimental Results

Model	Manual-labelled Tweets			Auto-labelled Tweets		
	Precision	Recall	F1	Precision	Recall	F1
One-Decoder	0.160	0.160	0.160	0.316	0.316	0.316
Multi-Decoder	0.150	0.150	0.150	0.340	0.340	0.340
NovelTagging	0.273	0.336	0.302	0.570	0.593	0.582
SPTree	0.424	0.349	0.383	0.434	0.366	0.397
JointER	**0.644**	0.369	0.469	0.405	0.314	0.354
SPERT	0.239	**0.675**	0.339	0.310	**0.839**	0.421
Proposed Model	0.545	0.630	**0.567**	**0.638**	0.642	**0.592**

Current Progress - Feature Selection-based Aspect Extraction

RQ1:

- Linguistic Features
- Feature Selection



Experimental Results

- Dataset: SemEval-2014, SemEval-2015, and SemEval-2016

	Laptop			Restaurant		
Method	P	R	F1	P	R	F1
RandomForest	0.700	0.533	0.606	0.719	0.614	0.663
MultinomialNB	0.537	0.733	0.620	0.563	0.766	0.649
SVM	0.737	0.587	0.654	0.761	0.695	0.726
RNN-based	0.810	**0.757**	0.782	0.828	0.804	0.816
DLIREC	0.819	0.671	0.738	0854	0.827	0.840
PSO	0.855	0.667	0.749	**0.871**	0.821	0.845
Proposed	**0.889**	0.755	**0.807**	0.856	**0.840**	**0.847**

	SemEval-15	SemEval-16
Method	F1	F1
RandomForest	0.513	0.504
MultinomialNB	0.483	0.502
SVM	0.504	0.463
LSTM	0.683	0.704
DE-CNN	0.683	0.744
HAST	0.715	0.736
Proposed	**0.722**	**0.746**

Timetable for Completion

Activities / Date	2020									2021												2022												2023		
	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3
1. PhD Course																																				
1.1. Research Method																																				
2. Finalise Research Scope																																				
3. Develop Research Questions																																				
4. RQ1: Graph-based Aspect Extraction																																				
4.1 Literature Review																																				
4.2 Data Collection & Analysis																																				
4.3 Proposed Method Implementation																																				
4.4 Conference Paper 1																																				
4.5 Journal Paper 1																																				
5. RQ2: Aspect-Opinion Pair Extraction																																				
5.1 Literature Review																																				
5.2 Data Collection & Analysis																																				
5.3 Proposed Method Implementation																																				
5.4 Conference Paper 2																																				
5.5 Journal Paper 2																																				
6. RQ3: Cross-Domain Sentiment Analysis																																				
6.1 Literature Review																																				
6.2 Data Collection & Analysis																																				
6.3 Proposed Method Implementation																																				
6.4 Conference Paper 3																																				
6.5 Journal Paper 3																																				
7. Preparing Thesis																																				

Published Papers

1. **Shi, J., Li**, W., Yang, Y., Yao, N., Bai, Q., Yongchareon, S., & Yu, J. (2021). Automated Concern Exploration in Pandemic Situations-COVID-19 as a Use Case. In Knowledge Management and Acquisition for Intelligent Systems: 17th Pacific Rim Knowledge Acquisition Workshop, PKAW 2020, Yokohama, Japan, January 7–8, 2021, Proceedings 17 (pp. 178-185). Springer International Publishing.
2. **Jingli Shi**, Weihua Li, Sira Yongchareon, Yi Yang, Quan Bai. Graph-based Joint Pandemic Concern and Relation Extraction on Twitter. Expert Systems with Applications (2021) (submitted)
3. **Jingli Shi**, Weihua Li, Quan Bai and Takayuki Ito. BeeAE: Effective Aspect Extraction using Artificial Bee Colony. The Pacific Rim International Conference on Artificial Intelligence (PRICAI) 2021 (submitted)

Thank you

Q & A