MVP: Multi-view Prompting Improves Aspect Sentiment Tuple Prediction

Source: Acl 2023
Advisor: JIA-LING KOH
Speaker: FAN-CHI-YU
Date:2023/09/05

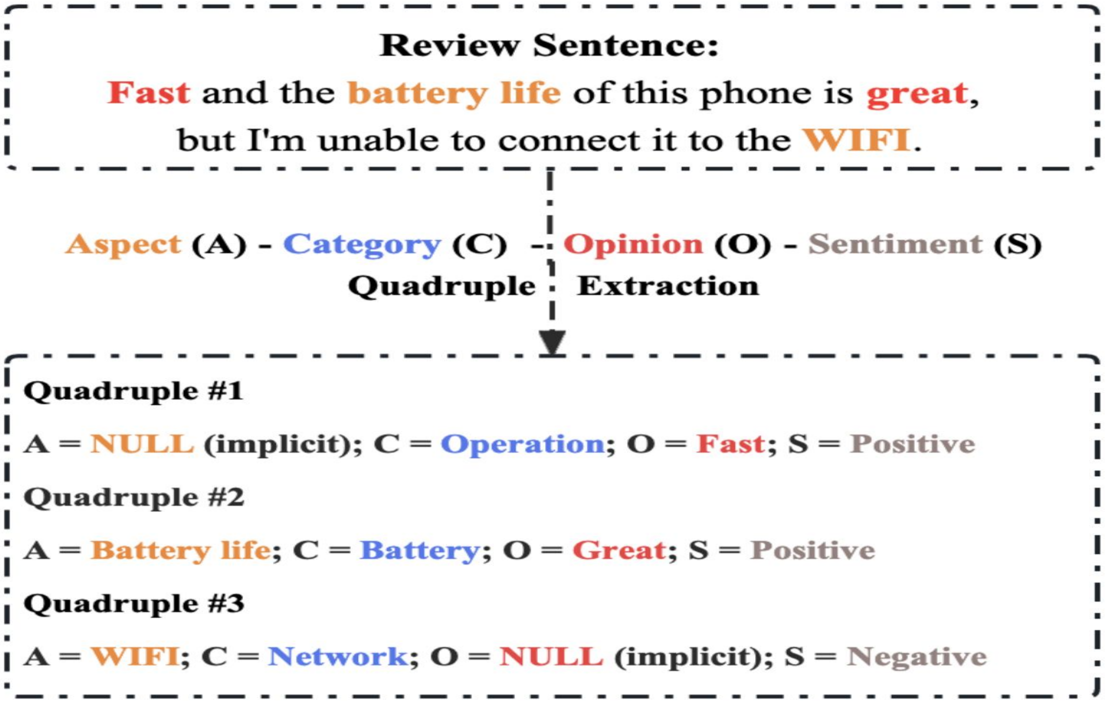
Outline

- Introduction
- Method
- Experiment
- Conclusion

Introduction

Aspect Based Sentiment Analysis(ABSA)

Aims to **predict** tuples of **sentiment elements** of interest for a given text.

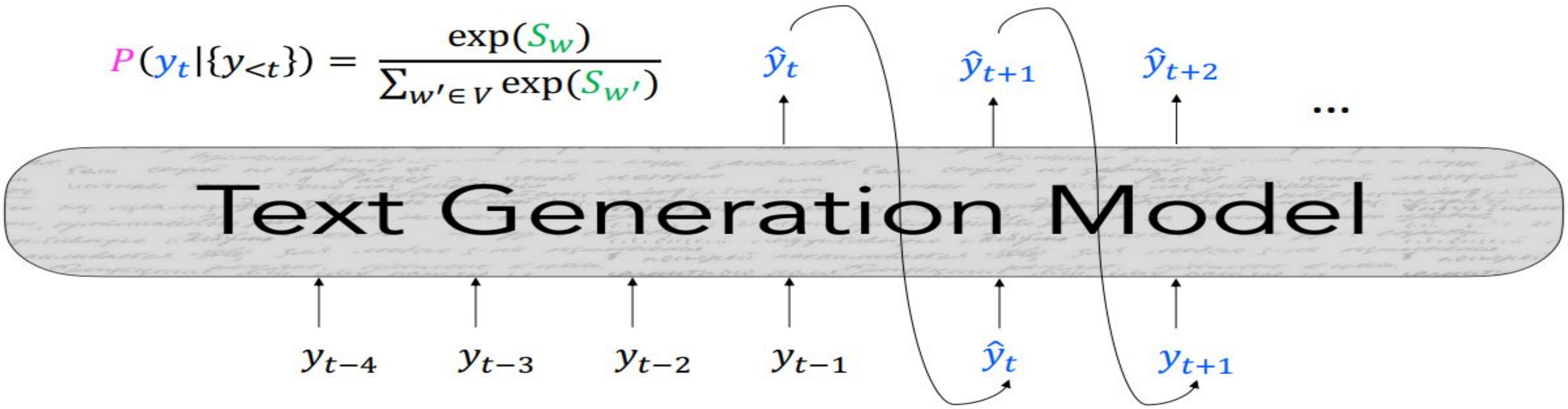


Aspect Based Sentiment Analysis(ABSA)

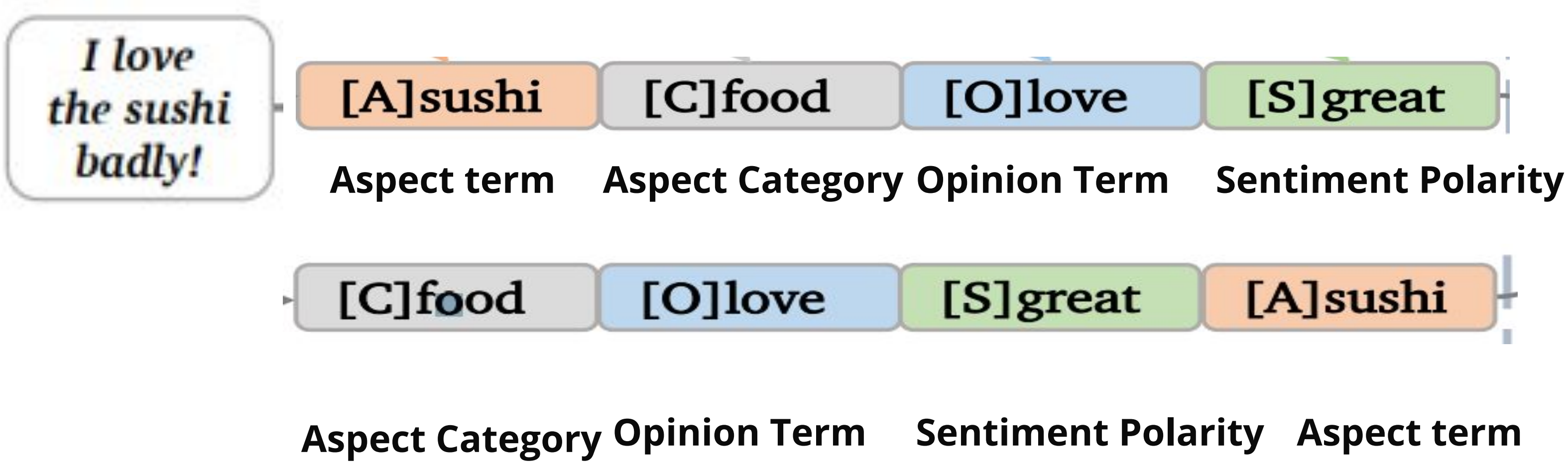
The battery life of this phone is great.

Type	Abbr	Task Name	Input	Output
Single	ATE	Aspect Term Extraction	S	a (battery life)
	OTE	Opinion Term Extraction	S	o (great)
	ACD	Aspect Category Detection	S	c (battery)
	AOCE	Aspect Opinion Co-Extraction	S	a (battery life), o (great)
	AOOE	Aspect-Oriented Opinion Extraction	S + a (battery life)	o (great)
	ABSC	Aspect-Based (Aspect-level) Sentiment Classification	S + a (battery life)	s (positive)
	COSC	Category-Oriented Sentiment Classification	S + c (battery)	s (positive)
Pair	AOPE	Aspect Opinion Pair Extraction	S	(a, o) (battery life, great)
	ASPE	Aspect Sentiment Pair Extraction	S	(a, s) (battery life, positive)
	CSPE	Category Sentiment Pair Extraction	S	(c, s) (battery, positive)
Triplet	ACSTE (TASD)	Aspect Category Sentiment Triplet Extraction or Target Aspect Sentiment Detection	S	(a, c, s) (battery life, battery, positive)
	AOSTE (ASTE)	Aspect Opinion Sentiment Triplet Extraction or Aspect Sentiment Triplet Extraction	S	(a, o, s) (battery life, great, positive)
Quad	ACOSQE	Aspect Category Opinion Sentiment Quadruple Extraction	S	(a, c, o, s) (battery life, battery, great, positive)

Neural Language Generation



Introduction(Aspect Sentiment Tuple Prediction)

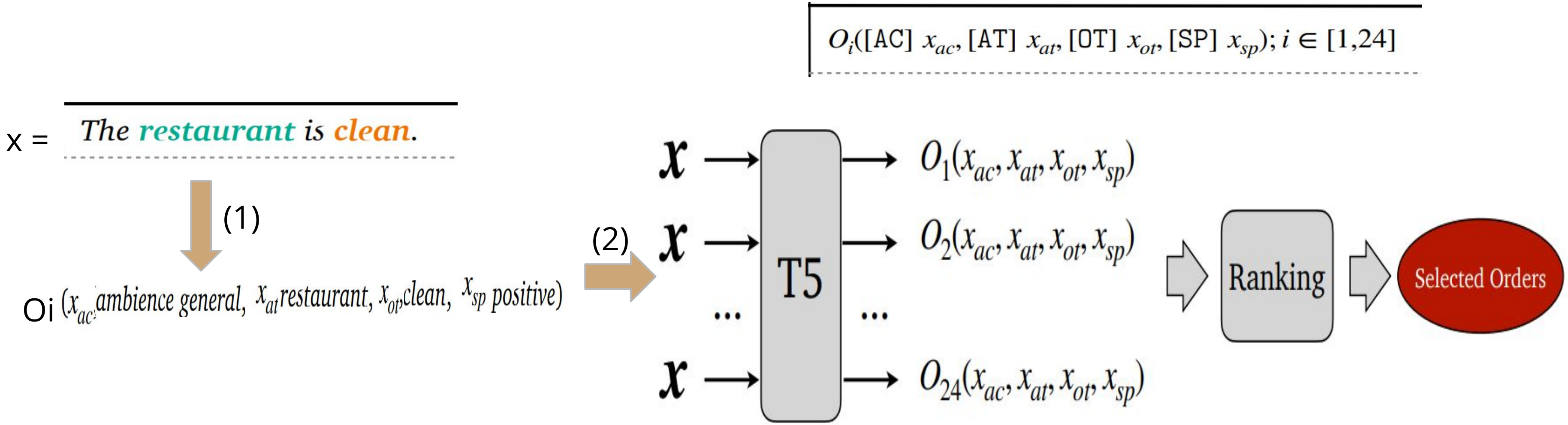


Paraphrase(Nature Language Way)

Designs semantic templates filled with **fixed-order** elements of tuples as generation targets.

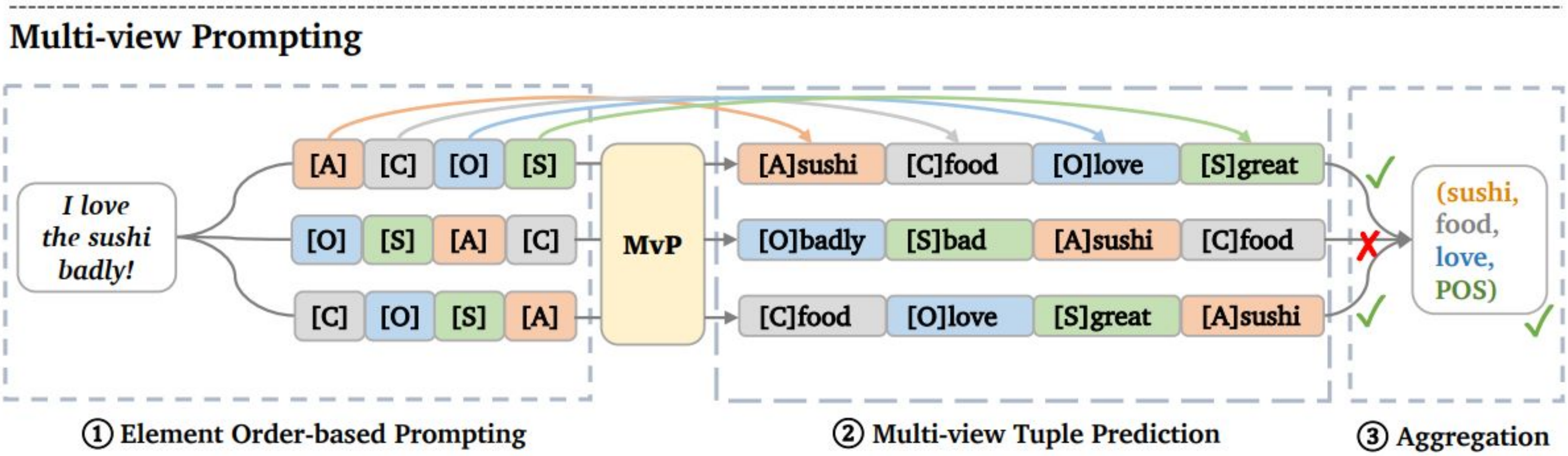
Input-1	*The pasta yesterday was delicious!*
Label-1	(c, a, o, p): (food quality, pasta, delicious, POS)
⇩	⇩
Target-1	Food quality is great because pasta is delicious

DLO(Order base)



Method

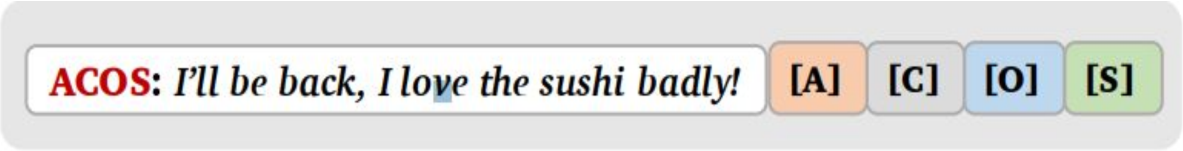
Method



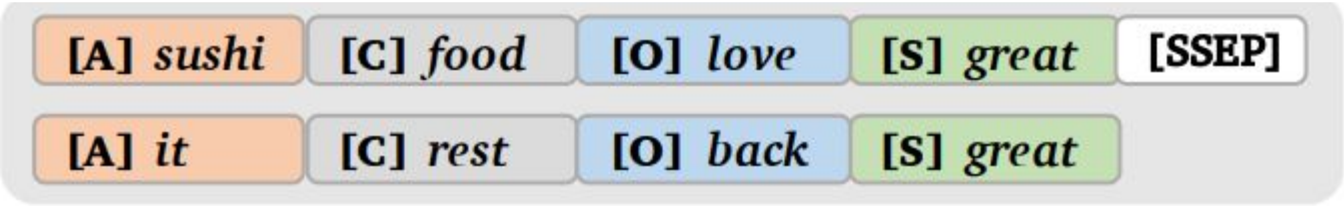
Multi-view Prompting
I love the sushi badly!
[A] [C] [O] [S]
[O] [S] [A] [C]
[C] [O] [S] [A]
MvP
[A]sushi [C]food [O]love [S]great
[O]badly [S]bad [A]sushi [C]food
[C]food [O]love [S]great [A]sushi
(sushi, food, love, POS)
① Element Order-based Prompting
② Multi-view Tuple Prediction
③ Aggregation

Element Order-based Prompt

Using element markers to represent the structure of information allows tokens to focus more on order.



If there have **multiple sentiment tuples** for an input sentence, we utilize [SSEP] to concatenate their final target sequence



Multi-view Training (Element Order Selection)

Choosing the potentially better-performing orders based on the average entropy of the candidate permutations on the training set.

{0.8,0.5....} ✓

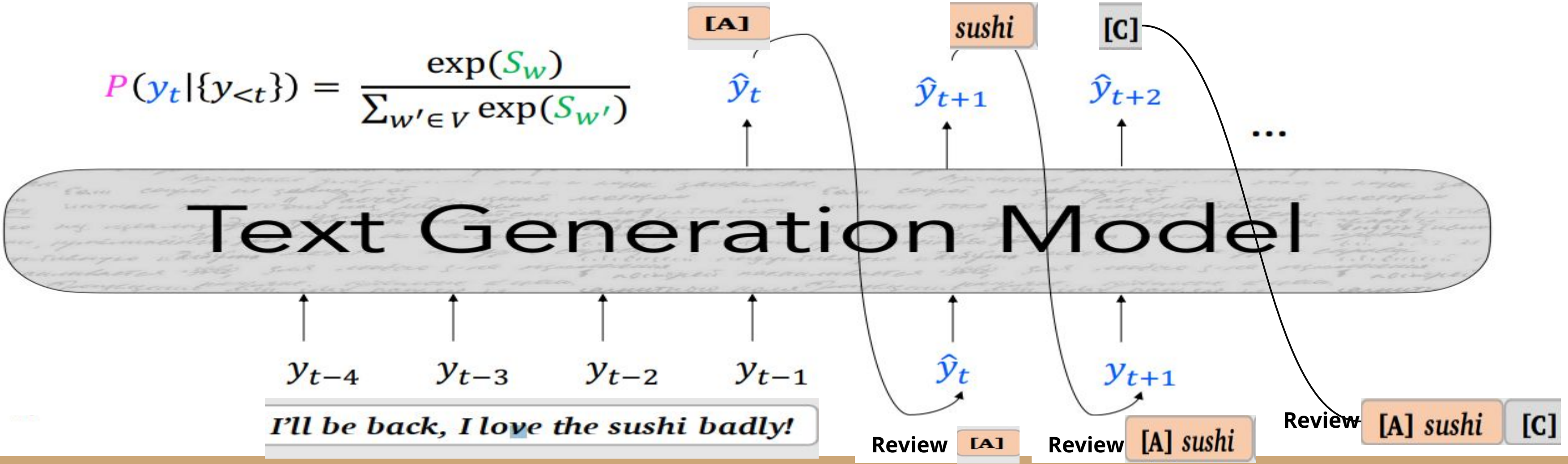
{0.8,0.6...} ✓

{0.6,0.7} ✓

$$S_{p_i} = \frac{\sum_D p(\boldsymbol{y}_{\boldsymbol{p_i}} | \boldsymbol{x})}{|D|} \quad (1)$$

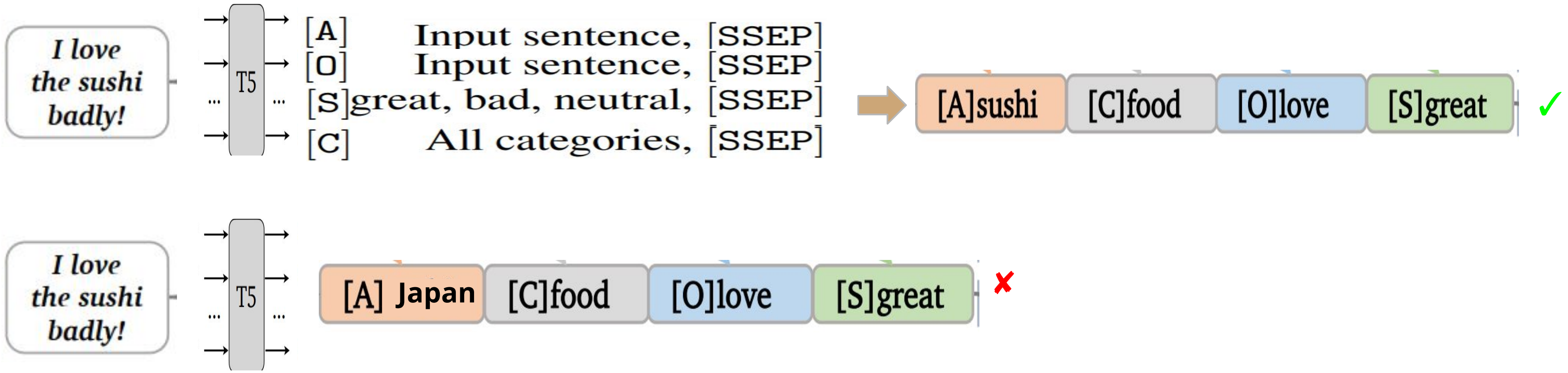
Multi-view Training (Training)

$$\mathcal{L}_{NLL} = -\mathbb{E}\log p(\boldsymbol{y}|\boldsymbol{x}) \ = -\mathbb{E}\sum_{t=1}^{T}\log p(\boldsymbol{y_t}|\boldsymbol{x}, \boldsymbol{y}_{<t}) \tag{2}$$



Multi-view Inference(Schema Constrained Generation)

We injects target schema **knowledge into the decoding** process.



Multi-view Inference(Results Aggregation)

$$T'_{\mathrm{MvP}} = \{t|t \in \bigcup_{i=1}^{m} T'_{p_i} \text{ and } (\sum_{i=1}^{m} \mathbb{1}_{T'_{p_i}}(t) \geq \frac{m}{2})\}$$

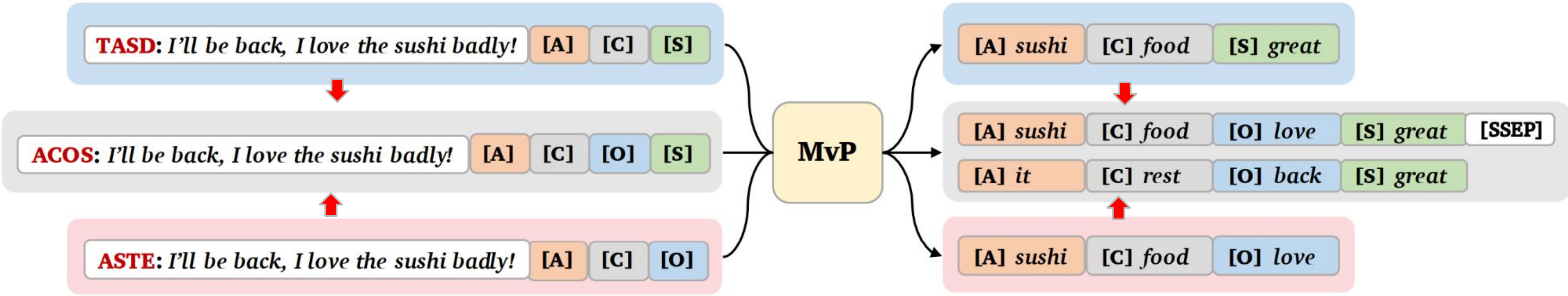
Voting

[A]sushi*3>3/2 ✓
[C]food*3 >3/2 ✓
[O]love*2>3/2 ✓
[S]gread*2>3/2 ✓

[O]badly*1<3/2 ✗
[O]bad*1<3/2 ✗

Multi-task learning

- MVP (multi-task) obtains generalized ability among diverse tasks
- Let Main task can learn well



Experiment

Experiment(Datasets)

- **SemEval Datasets (Semantic analysis and Annotated datasets):**

review	aspects_term	cat_obj	opinions_term	sematic polarity
<<null>> this unit is `` pretty `` and styli...	['unit']	{LAPTOP}	['pretty']	POS
<<null>> for now i ' m okay with upping the ex...	['device']	{LAPTOP}	['<<null>>']	NEU
<<null>> seems unlikely but whatever , i ' ll ...	['<<null>>']	{LAPTOP}	['<<null>>']	NEU
<<null>> this version has been my least favori...	['version']	{LAPTOP}	['least', 'favorite']	NEG
<<null>> - biggest disappointment is the track...	['track', 'pad']	{HARDWARE}	['disappointment']	NEG
<<null>> should not of bought this chromebook .	['chromebook']	{LAPTOP}	['<<null>>']	NEG
<<null>> after about 5 / 10 minutes of use the...	['screen']	{DISPLAY}	['crazy']	NEG
<<null>> i can not stand the trackpad or the k...	['trackpad']	{KEYBOARD}	['<<null>>']	NEG
<<null>> the keyboard is stiff and unresponsiv...	['keyboard']	{KEYBOARD}	['stiff']	NEG
<<null>> the chromebook r 11 was hardly used a...	['chromebook', 'r', '11']	{LAPTOP}	['<<null>>']	NEG
<<null>> one day it just would not power up .	['<<null>>']	{BATTERY}	['<<null>>']	NEG

Experiment

Generative methods

- Paraphrase
  a. Designs semantic templates filled with **fixed-order** elements of tuples as generation targets.
- DLO:
  a. The **order-free** property of the quadruplet based on templates.

Multi-tasking methods.

- Lego-ABSA:
  a. Designs task prompts similar to T5.Assembling task prompts, like assembling **Lego** bricks.

Lego-ABSA

The proposed approach can train on simple tasks and transfer to difficult tasks by assembling task prompts, like assembling **Lego** bricks

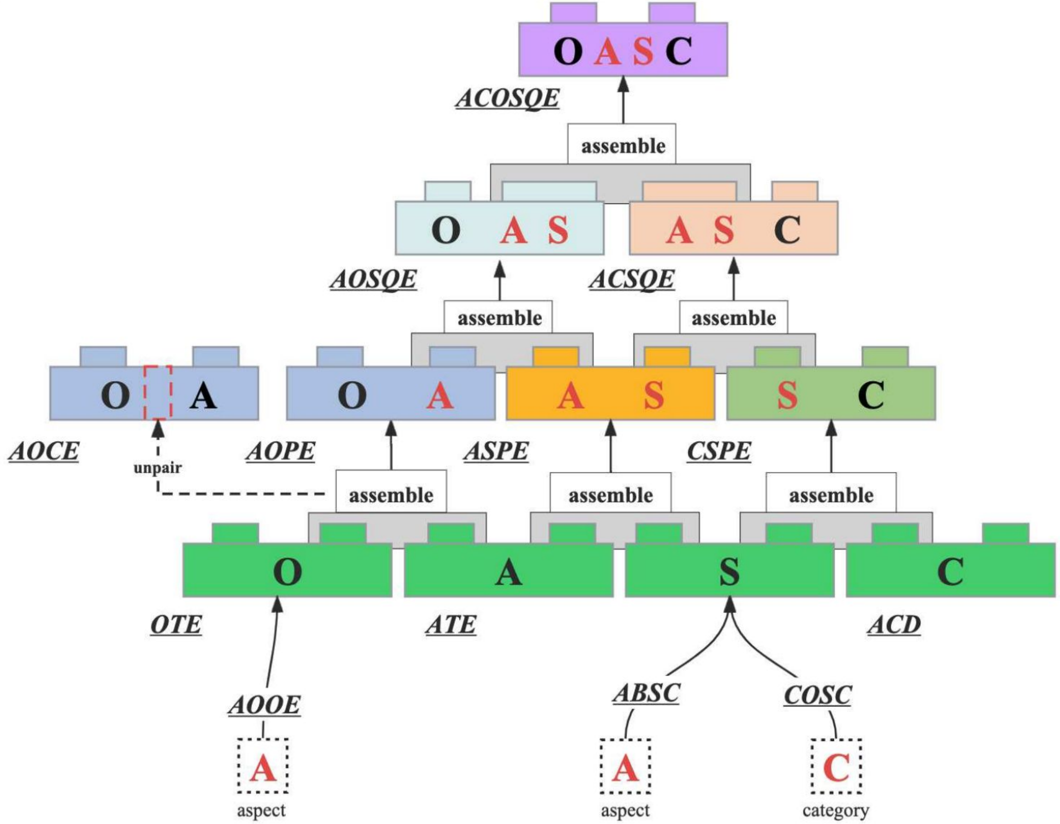


Fig. 3: Assembling the sentiment analysis task like building with Lego blocks.

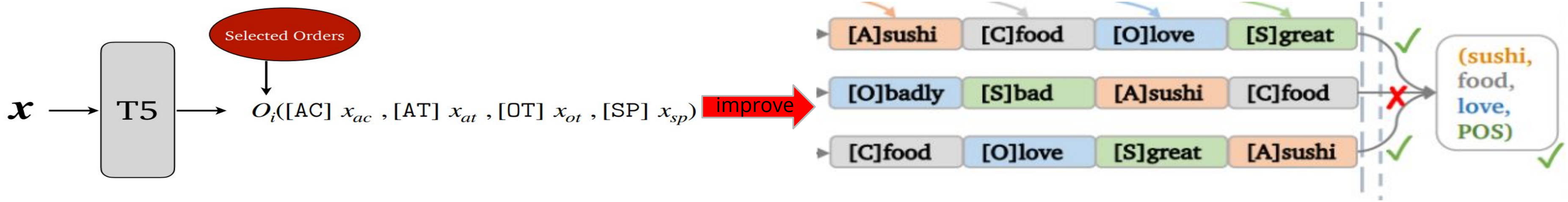
Experiment

Methods	ASQP		ACOS		TASD		ASTE				AVG
	R15	**R16**	**Lap**	**Rest**	**R15**	**R16**	**L14**	**R14**	**R15**	**R16**	
Paraphrase (Zhang et al., 2021b)	46.93	57.93	43.51	61.16	63.06	71.97	61.13	72.03	62.56	71.70	61.20
DLO (Hu et al., 2022)	48.18	59.79	43.64	59.99	62.95	71.79	61.46	72.39	64.26	73.03	61.75
LEGO-ABSA† (Gao et al., 2022)	46.10	57.60	-	-	62.30	71.80	62.20	73.70	64.40	69.90	-
MvP	51.04	**60.39**	**43.92**	**61.54**	64.53	**72.76**	63.33	74.05	65.89	**73.48**	63.09
MvP (multi-task)†	**52.21**	58.94	43.84	60.36	**64.74**	70.18	**65.30**	**76.30**	**69.44**	73.10	**63.44**

Experiment

Methods	ASQP $(x_{ac}, x_{at}, x_{ot}, x_{sp})$		ACOS $(x_{ac}, x_{at}, x_{ot}, x_{sp})$		TASD (x_{ac}, x_{at}, x_{sp})		ASTE (x_{at}, x_{ot}, x_{sp})				AVG
	R15	**R16**	**Lap**	**Rest**	**R15**	**R16**	**L14**	**R14**	**R15**	**R16**	
Paraphrase (Zhang et al., 2021b)	46.93	57.93	43.51	61.16	63.06	71.97	61.13	72.03	62.56	71.70	61.20
DLO (Hu et al., 2022)	48.18	59.79	43.64	59.99	62.95	71.79	61.46	72.39	64.26	73.03	61.75
LEGO-ABSA† (Gao et al., 2022)	46.10	57.60	-	-	62.30	71.80	62.20	73.70	64.40	69.90	-
MvP	51.04	**60.39**	**43.92**	**61.54**	64.53	**72.76**	63.33	74.05	65.89	**73.48**	63.09
MvP (multi-task)†	**52.21**	58.94	43.84	60.36	**64.74**	70.18	**65.30**	**76.30**	**69.44**	73.10	**63.44**

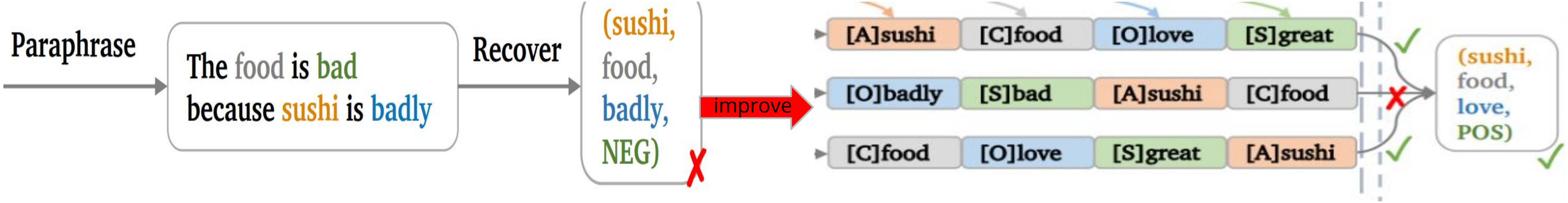
The MVP can **improve lot** at the opinion prediction



Experiment

Methods	ASQP		ACOS		TASD		ASTE				AVG
	$(x_{ac}, x_{at}, x_{ot}, x_{sp})$		$(x_{ac}, x_{at}, x_{ot}, x_{sp})$		(x_{ac}, x_{at}, x_{sp})		(x_{at}, x_{ot}, x_{sp})				
	R15	R16	Lap	Rest	R15	R16	L14	R14	R15	R16	
Paraphrase (Zhang et al., 2021b)	46.93	57.93	43.51	61.16	63.06	71.97	61.13	72.03	62.56	71.70	61.20
DLO (Hu et al., 2022)	48.18	59.79	43.64	59.99	62.95	71.79	61.46	72.39	64.26	73.03	61.75
LEGO-ABSA† (Gao et al., 2022)	46.10	57.60	-	-	62.30	71.80	62.20	73.70	64.40	69.90	-
MvP	51.04	**60.39**	**43.92**	**61.54**	64.53	**72.76**	63.33	74.05	65.89	**73.48**	63.09
MvP (multi-task)†	**52.21**	58.94	43.84	60.36	**64.74**	70.18	**65.30**	**76.30**	**69.44**	73.10	**63.44**

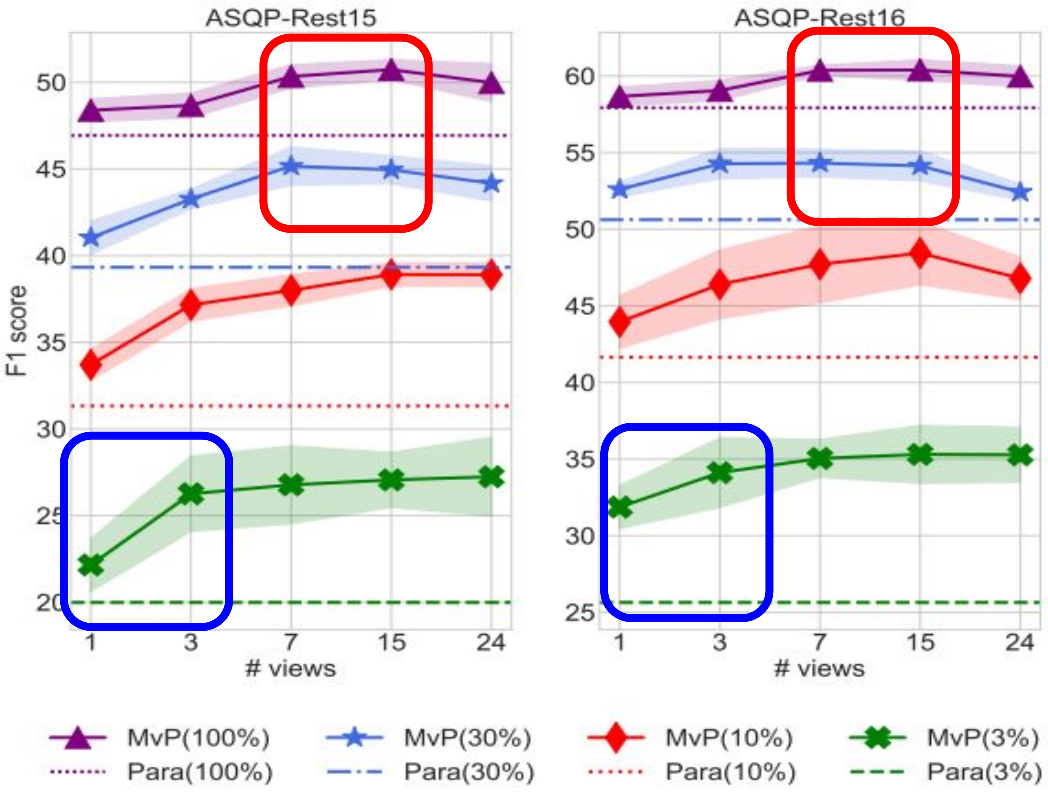
The MVP can **improve lot** at the opinion prediction



Experiment(Effect Analysis)

1. F1 **decreases** slightly after a certain number (between 7 and 15)

2. **Low-resource** setting **boost a lot**



Experiment(Ablation test)

- Constrained decoding significantly impacts text generation
- Random sampling outperforms Top-1 ranked, as it introduces variability and reduces model confusion.

	Methods	**ASTE (L14)** 1%	**ASTE (L14)** 10%	**ASTE (L14)** 100%	**ASQP (R15)** 1%	**ASQP (R15)** 10%	**ASQP (R15)** 100%
w/o Constrained decoding	MvP w/o cd	21.37	49.98	<u>63.27</u>	12.09	<u>37.87</u>	<u>50.92</u>
Random Select one	MvP (rand)	<u>27.32</u>	<u>51.02</u>	62.50	**13.56**	37.18	49.84
Select top one	MvP (rank)	25.98	49.98	62.48	13.38	37.45	49.98
	MvP	**28.37**	**52.33**	**63.33**	<u>13.46</u>	**38.48**	**51.04**

Cross-task transfer

- Transfer brings further significant improvements, from **triplets to quadruplets**

	Methods	Transfer Source	1%	2%	5%	10%	20%	AVG
ASQP (R15)	DLO (transfer)	ASTE (R15)	26.28	28.72	35.94	39.48	42.92	34.67
	MvP (transfer)	ASTE (R15)	**28.69**	**33.93**	**40.08**	**43.10**	**45.09**	**38.18**
ASTE (L14)	DLO (transfer)‡	ASQP (R16)	44.76	48.86	51.22	**56.43**	56.71	51.60
	MvP (transfer)‡	ASQP (R16)	**48.43**	**50.33**	**54.27**	56.34	**59.05**	**53.68**

Case Study1

Example 1 (ASQP task)

Sentence: *The restaurant offers an extensive wine list and an ambiance you won't forget.*

Gold: (wine list, drinks style_options, great, extensive), (ambiance, ambience general, great, won't forget)

Tuples from 15 views:		**Final output:**
(wine list, drinks style_options, great, extensive) * 10 (ambiance, ambience general, great, won't forget) * 15	Pick →	(wine list, drinks style_options, great, extensive) ✓ (ambiance, ambience general, great, won't forget) ✓
(restaurant, drinks style_options, great, extensive) * 5	Drop	

- MVP handles well after **filtering** unreasonable tuples predicted
- MVP only outputs tuples considered important in most views

Case Study2

Example 2 (ACOS task)

Sentence: *I do like the screen on this , images are clean and crisp , enjoying the 4 gigs of ram which allow me to have a few more tab open .*

Gold: (screen, display general, great, like), (ram, memory general, great, enjoying), (screen, display general, great, clean), (screen, display general, great, crisp)

Tuples from 15 views:

(screen, display general, great, like) * 15
(ram, memory general, great, enjoying) * 9
(images, display general, great, clean) * 8
(images, display quality, great, crisp) * 8

Pick

Final output:

(screen, display general, great, like) ✓
(ram, memory general, great, enjoying) ✓
(images, display general, great, clean) ✗
(images, display quality, great, crisp) ✗

(images, display quality, great, clean) * 6 (ram, memory operation_performance, great, enjoying) * 6
(images, display general, great, crisp) * 6 (images, display design_features, great, clean) * 1
(images, display design_features, great, crisp) * 1

Drop

Laptop dataset includes **121** categories that model get **confused** with similar **aspect categories**

Conclusion

Conclusion

1. **MVP** Improves aspect-level opinion information prediction by effective multi-view results aggregation.

2. **MVP** proposed Element order-based prompt learning method.

3. Multi-tasking model substantially outperforms task-specific models on a variety of ABSA tasks.