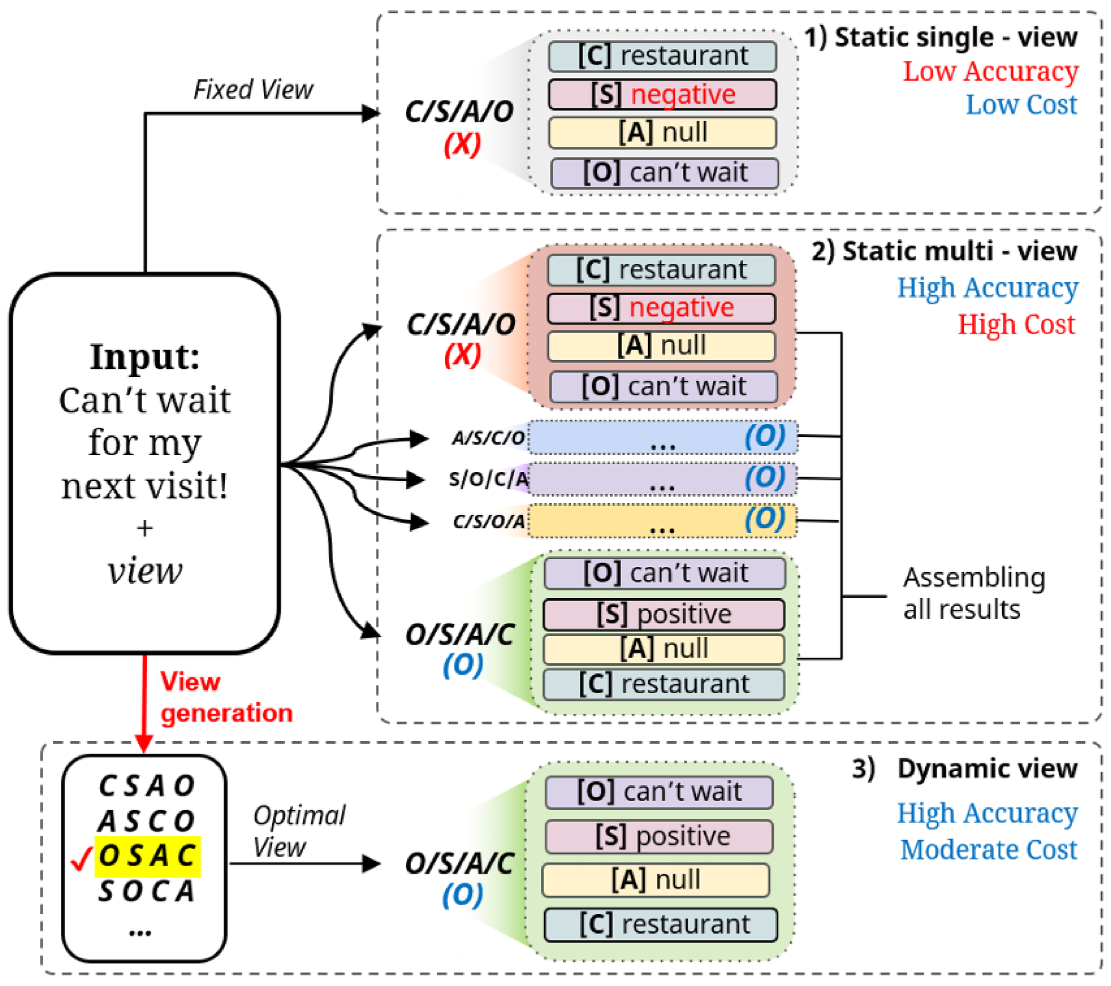
Dynamic Order Template Prediction for Generative Aspect-Based Sentiment Analysis

Yonghyun Jun, Hwanhee Lee
Language Intelligence lab, Chung-Ang University

Motivation & Problem



- Fixed order ABSA templates (ex. C S A O) are <u>order sensitive</u>;
- Multi view prompting (*MvP*) ensembles many orders ⇒ <u>high latency</u> & <u>domain specific</u> tuning.
- Dynamic Order Templates (*DOT*) **<u>choose only the views each sentence needs</u>** via entropy.

Method

- **<u>Stage 1: View Training</u>** - Train # tuples (K) via entropy ranked views *(A, C, S)* in one view ↔ one tuple scaffold.

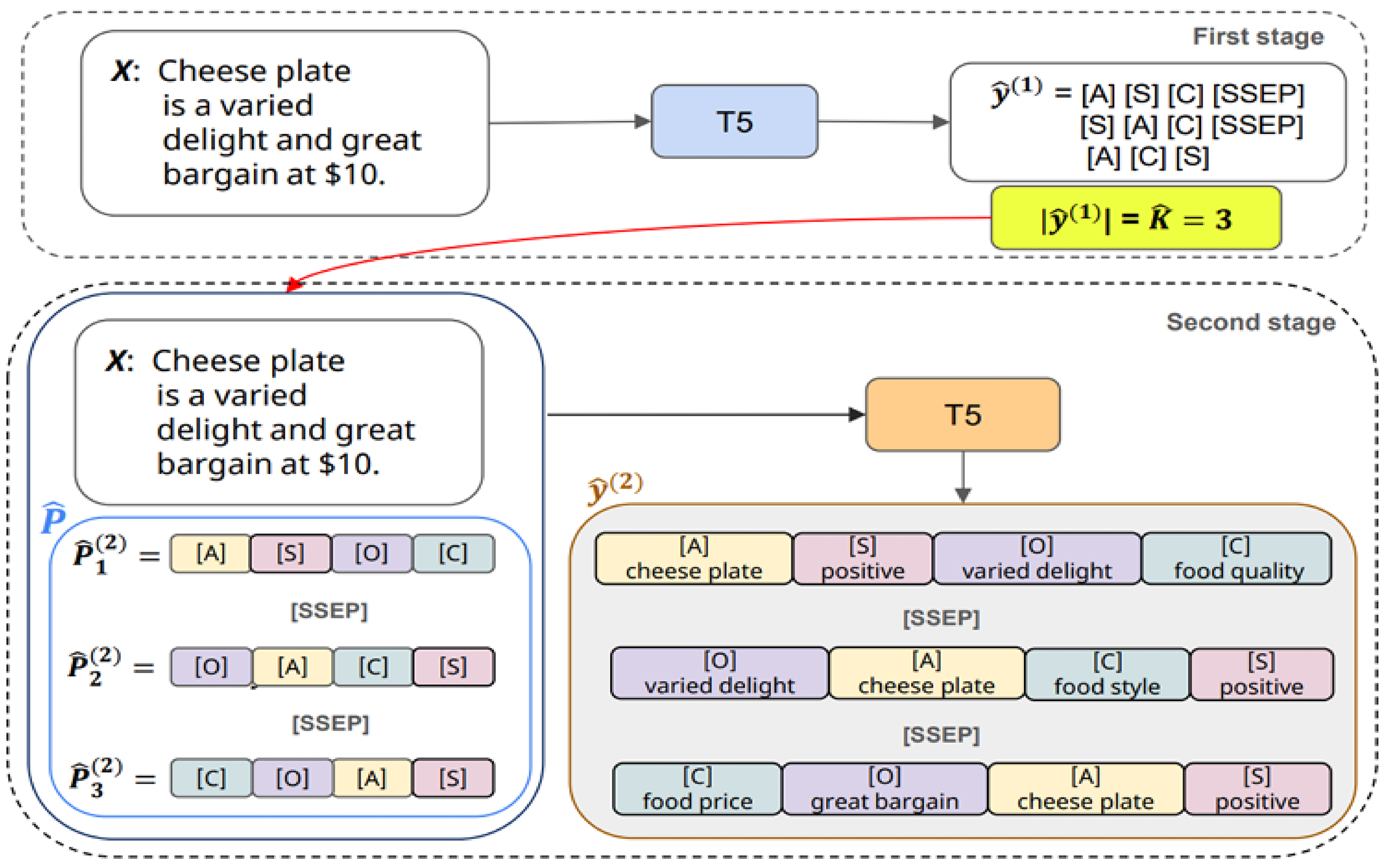
$$\mathcal{E}_{i,v} = -\sum P(v|x_i) \log P(v|x_i)$$ ➢ Entropy score calculation

$$y_i^{(1)} = P_{i,1}^{(1)} \text{ [SSEP] } P_{i,2}^{(1)} \text{ [SSEP] } \dots P_{i,K_i}^{(1)}$$ ➢ First stage target

- **<u>Stage 2: Tuple Training</u>** - Reuse scaffolds to output full *(A C O S)* quadruples with one-to-one alignment.

$$y_i^{(2)} = P_{i,1}^{(2)} \otimes tuple_1 \text{ [SSEP] } \dots P_{i,K_i}^{(2)} \otimes tuple_{K_i}$$ ➢ Second stage target

- Constrained decoding keeps output well formed.



- **<u>Two-stage inference:</u>** Predict # tuples → Predict exact tuples

Main Results

Methods	ASQP R15	ASQP R16	ACOS Lap	ACOS Rest	MEMD M-Rest	MEMD M-Lap	MEMD Books	MEMD Clothing	MEMD Hotel	Avg	Time(s)
TAS-BERT	34.78	43.71	27.31	33.53	-	-	-	-	-	-	-
Extract-Classify	36.42	43.77	35.80	44.61	-	-	-	-	-	-	-
One-ASQP (large)	-	-	41.56	60.69	-	-	-	-	-	-	-
Seq2Path	-	-	42.97	58.41	-	-	-	-	-	-	-
AugABSA	50.01	60.88	-	-	-	-	-	-	-	-	-
SCRAP	49.93	**62.48**	-	-	-	-	-	-	-	-	-
Paraphrase	46.93	57.93	43.51	<u>61.16</u>	57.38	35.07	39.30	43.00	68.79	50.34	40.63
DLO	48.18	59.79	43.64	59.99	57.07	<u>35.56</u>	<u>42.63</u>	43.35	**70.27**	51.16	260.74
MvP	<u>51.04</u>	60.39	<u>43.92</u>	**61.54**	<u>58.12</u>	35.25	42.57	**43.94**	69.06	<u>51.76</u>	2161.81
GPT-4o	40.45	47.29	24.77	46.53	35.11	20.69	30.39	40.27	24.84	34.48	-
LLaMa-3.1-8b	37.52	47.60	40.07	54.06	38.10	31.16	28.62	32.21	44.62	39.33	-
Qwen-2.5-7b	29.93	39.34	12.48	33.56	25.63	24.13	17.77	18.09	38.03	26.66	-
Mistral-7b	44.14	51.96	39.02	53.02	41.28	26.80	26.54	21.81	40.35	38.32	-
DOT (Ours)	**51.91**	<u>61.24</u>	**44.92**	59.25	**58.25**	**39.02**	**43.02**	<u>43.37</u>	<u>69.94</u>	**52.28**	298.17

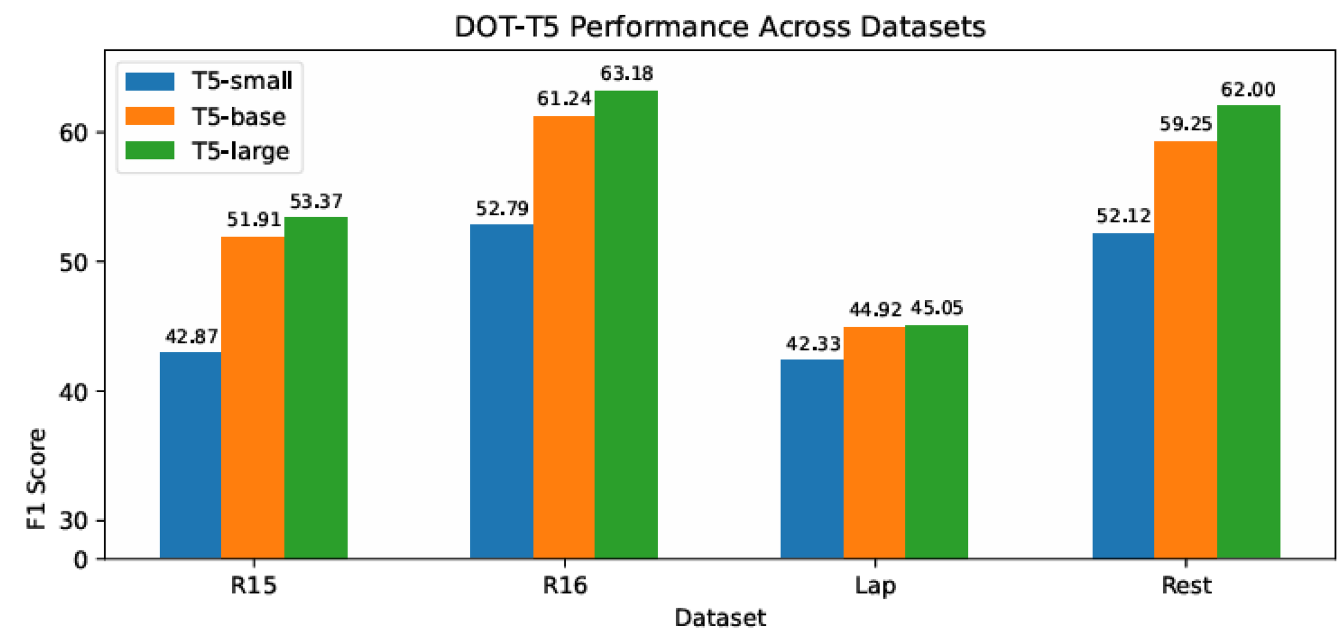
- **<u>Accuracy:</u>** #1 on 5 benchmarks, #2 on 3 → shows **<u>two stage specialization works</u>**.
- **<u>Speed:</u> <u>7× faster</u>** than static multi-view (MvP) method.

Transferability

Train	SemEval		Yelp	
Test	*SemEval*	*Yelp*	*Yelp*	*SemEval*
Paraphrase	52.38	38.52 (-11.86)	57.38	44.88 (-12.50)
MvP_3	55.62	34.42 (-21.20)	57.27	41.72 (-15.55)
MvP_9	56.89	35.02 (-21.87)	56.98	42.52 (-14.46)
MvP_{15}	**57.66**	35.21 (-21.45)	58.12	41.94 (-16.18)
DOT	57.47	**39.88** (-17.59)	**58.25**	**46.97** (-11.28)

- Zero shot cross domain test: SemEval(R15, R16, Rest) ↔ Yelp(M-Rest)
- **<u>DOT retains relatively high F1-score</u>**; The others crash.

Analysis



- **<u>F1 score for model scale:</u>** T5-small < T5-base < T5-large
- **<u>T5 base</u>** is the best acc/latency balance.

Model Configuration	Average F1
Full Model	54.33
w/o stage division	52.73 (-1.60)
w/o entropy score	52.53 (-1.80)
w/o multi view	52.31 (-2.02)
w/o stage division, entropy score	50.04 (-4.29)
w/o filtering, stage division, entropy score	45.80 (-8.53)

- **<u>Ablation Study:</u>** Our full model outperforms all ablated settings, demonstrating that **<u>each component of our design is essential.</u>**

Case Study

Case 1: Efficiency in Simple sentence

Input: Best mexican place for lunch in the financial district.

Target: [(mexican place, best, positive, restaurant general)]

Output: [(mexican place, best, positive, restaurant general)]

Case 2: One sentiment tuple, but complex

Input: The crowd is mixed yuppies, young and old.

Target: [(crowd, null, neutral, restaurant miscellaneous)]

Output: [(crowd, **mixed**, neutral, **ambience general**)]

Case 3: Complex sentence analysis

Input: If you ' re interested in good tasting (without the fish taste or smell), large portions and creative sushi dishes this is your place…

Target: [(null, good, positive, food quality), (portions, large, positive, food_style_options), (sushi dishes, creative, positive, food_style_options)]

Output: [(null, **good tasting**, positive, food quality), (portions, large, pos, food_style_options), (sushi dishes, creative, positive, food_style_options), (fish taste or smell, null, negative, food quality)]

- First-stage optimization reduces count errors, but richer view-selection logic is future work.

Conclusion

- *DOT* merges the best of **<u>single view efficiency</u>** and **<u>multi view robustness</u>**. Across 9 ABSA benchmarks we exceed—or reach—**<u>state of the art F1</u>** while **<u>cutting inference time</u>**.
- In short: Fewer views, Better tuples, Improved transferability.