Dynamic Order Template Prediction for Generative Aspect-Based Sentiment Analysis

The 63rd Annual Meeting of the Association for Computational Linguistics

ACL 2025 VIENNA

Yonghyun Jun

LILAB

Department of Artificial Intelligence

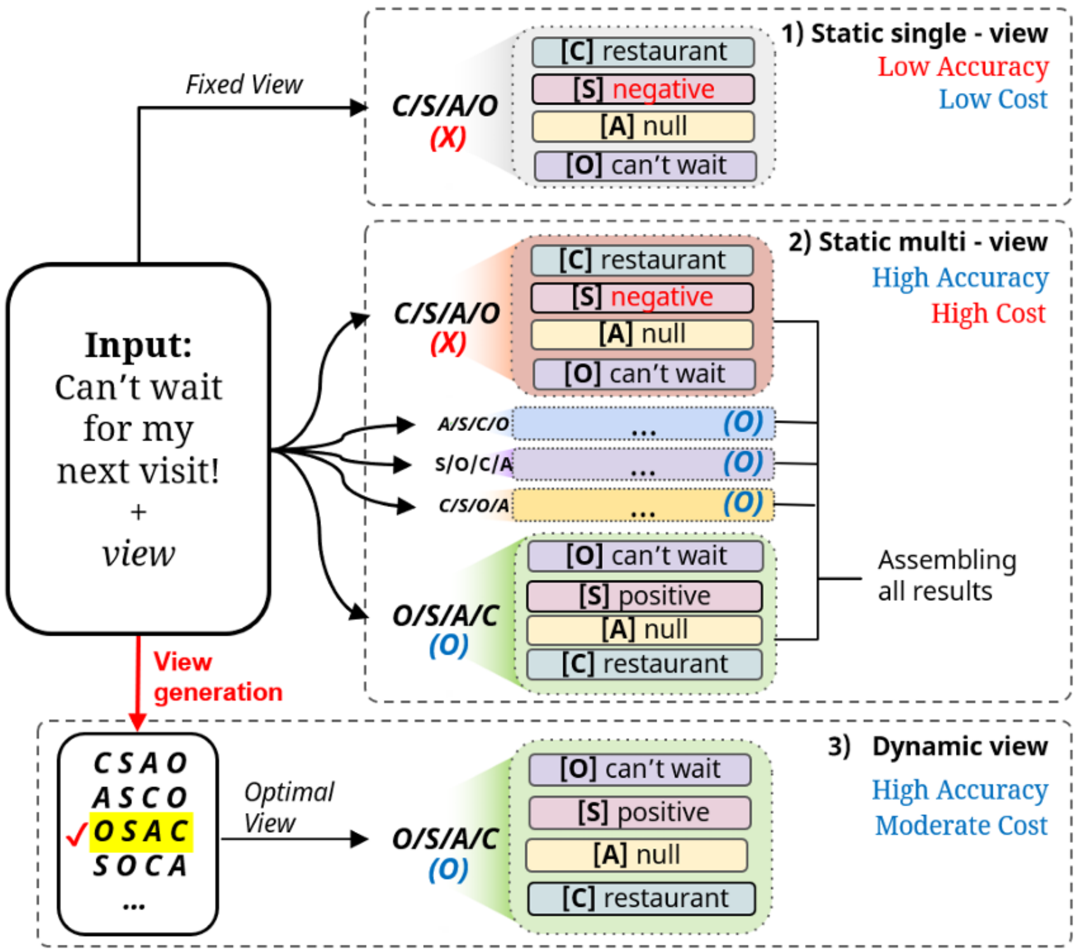
Chung-Ang University

Reference Citation

Outline

- Introduction
- Method
- Experimental Setup
- Results & Analysis
- Conclusion & Limitation

Introduction



- Fixed order ABSA templates (A C O S) are **order sensitive**; permutations can flip predictions.
- Multi view prompting (MvP) ensembles many orders ⇒ **high latency** & **domain specific** tuning.
- Our Solution: Dynamic Order Templates (DOT) **choose only the views each sentence needs** via entropy.
- Benefits: **Fewer views** ⇒ faster inference; **Dynamic selection** ⇒ higher accuracy

Method – First Stage

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

➢ Entropy score calculation

$$y_i^{(1)} = P_{i,1}^{(1)} \text{ [SSEP] } P_{i,2}^{(1)} \text{ [SSEP] } \ldots \; P_{i,K_i}^{(1)}$$

➢ First stage target

$$\mathcal{L}_1 = -\sum_{i=1}^{|B|} \sum_{t=1}^{T} \log p(\boldsymbol{y}_{i,t}^{(1)} | \boldsymbol{x}_i, \boldsymbol{y}_{i,<t}^{(1)})$$

➢ First stage loss function

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

Method – Second Stage

$$y_i^{(2)} = P_{i,1}^{(2)} \otimes tuple_1 \,[\mathrm{SSEP}] \ldots P_{i,K_i}^{(2)} \otimes tuple_{K_i}$$

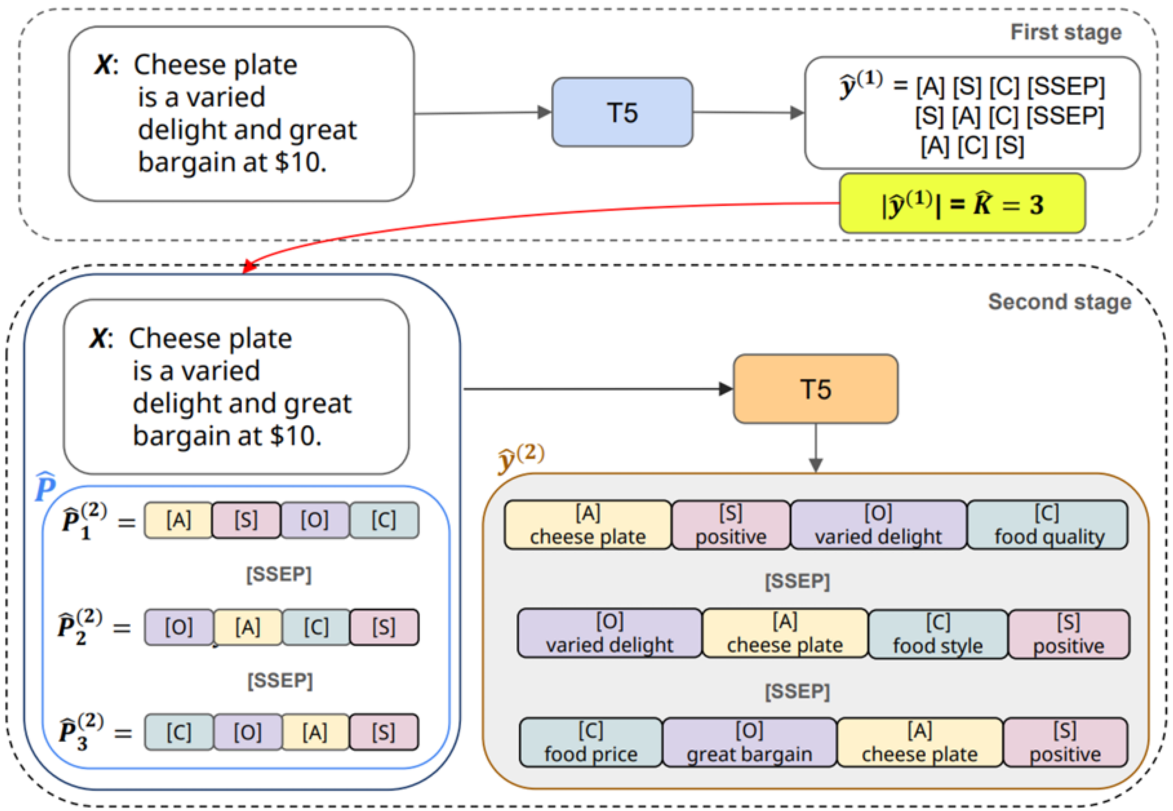
- Second stage target

$$\mathcal{L}_2 = -\sum_{i=1}^{|B|}\sum_{t=1}^{T} \log p(\boldsymbol{y_{i,t}^{(2)}} | \boldsymbol{x_i}, \boldsymbol{P_i}, \boldsymbol{y_{i,<t}^{(2)}})$$

- Second stage loss function

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

Method – Inference



- **<u>Two-stage inference:</u>** Predict # tuples → Predict exact tuples
- **Efficiency:** Skips unneeded permutations ⇒ far less compute than static multi-view.
- **Robustness:** Still benefits from multiple perspectives where necessary.

Experimental Setup

- Datasets (9): ASQP (R15, R16) , ACOS (Lap, Rest), MEMD (M-Rest, M-Laptop, Books, Clothing, Hotel)
- Baselines (13): 9 ABSA models — TAS BERT, Extract Classify, One ASQP,
  Seq2Path, AugABSA, SCRAP, Paraphrase, DLO, MvP
  — plus 4 LLMs (GPT 4o, Llama 3.1 8 B, Qwen 2.5 7 B, Mistral 7 B).
- Backbone: T5 base fine tuned in Stage 1, reused in Stage 2 for regularized warm start.
- Extras: Stop word filtering cleans labels, Constrained decoding enforces format

Results & Analysis – F1 score, Speed

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

Results & Analysis – 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)
- DOT retains relatively high F1-score; The others crash.

Conclusion & Limitation

Conclusion

- **Dynamic Order Templates (DOT)** unite single-view speed with multi-view robustness.
- **Outcome:** ≥ SOTA F1 on 9 datasets • **> 60 % faster** than static multi-view.
- **TL;DR:** Fewer views → better tuples → practical ABSA.

Limitations

- Two sequential models → not end-to-end; longer training.
- Stage-1 errors may cascade → higher variance across seeds.
- View count ≈ tuple count – simple, not always optimal.