DiaASQ: A Benchmark of Conversational Aspect-based Sentiment Quadruple Analysis

Bobo Li[1], Hao Fei[2], Fei Li[1], Yuhan Wu[1], Jinsong Zhang[1], Shengqiong Wu[2],
Jingye Li[1], Yijiang Liu[1], Lizi Liao[3], Tat-Seng Chua[2] and Donghong Ji[1]
[1] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education,
School of Cyber Science and Engineering, Wuhan University
[2] Sea-NExT Joint Lab, National University of Singapore [3] Singapore Management University

TL;DR

This work first proposes the ABSA task in a conversational scenario, named DiaASQ, establishing a benchmark by providing an annotated dataset and extraction model.

Project

Code && Data

▶ 1. Abstract

DiaASQ: We propose a new task called Conversational Aspect-based Sentiment Quadruplet Analysis (DiaASQ), analyzing multi-party dialogues to identify and extract aspect-level sentiment quadruplets. These quadruplets capture opinions expressed about specific aspects related to a particular target, along with their corresponding sentiment polarity.

Keywords: **Aspect-based Sentiment Analysis**, **Conversational Sentiment**

Contribution:

- We pioneer the research of dialogue-level aspect-based sentiment analysis
- We release a dataset for the DiaASQ task in both Chinese and English languages, which is of high quality and at a large scale.
- We introduce an end-to-end model to benchmark the DiaASQ task.

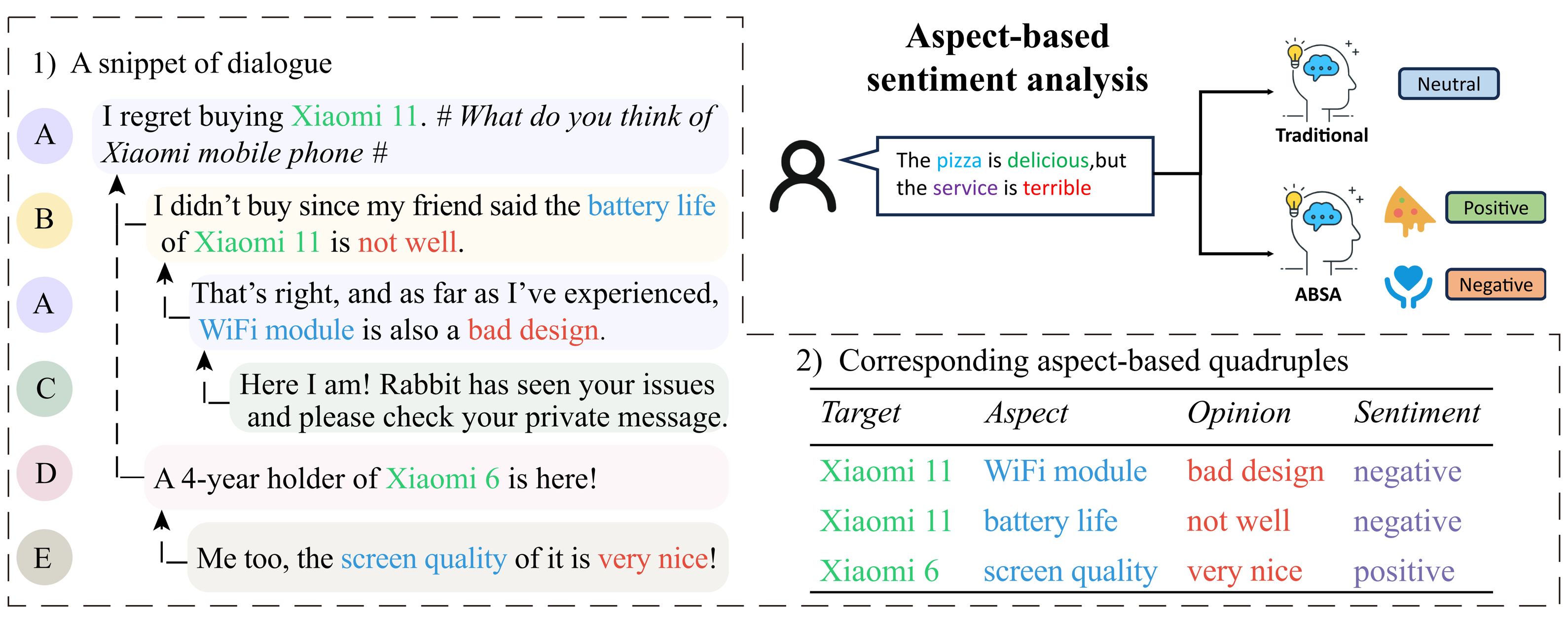

Figure 1: Illustration of the conversational aspect-based sentiment quadruple analysis (DiaASQ). The dialogue utterances produced by the corresponding speakers (marked at left) are organized into replying structure.

▶ 2. Dataset Construction

The dataset is constructed by systematically crawling tweets from digital bloggers, followed by a series of preprocessing steps including filtering, normalizing, pruning, and annotating the collected dialogues, resulting in a final corpus of 1,000 dialogues. Additionally, to enhance its multilingual usability, the dataset is further translated and projected into the English language.

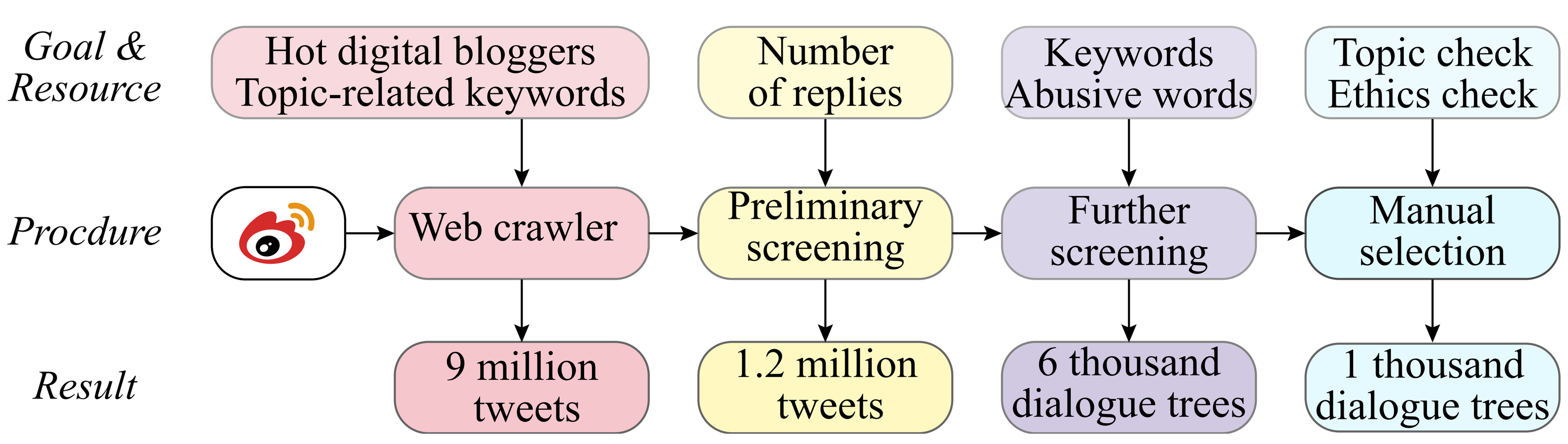

Figure 2: The workflow of data acquisition and preprocesssing.

Table 1: Data statistics.

		Dialogue			Items			Pairs			Quadruples		
		Dia.	Utt.	Spk.	Tgt.	Asp.	Opi.	$\text{Pair}_{t\text{-}a}$	$\text{Pair}_{t\text{-}o}$	$\text{Pair}_{a\text{-}o}$	Quad.	Intra.	Cross.
ZH	Total	1,000	7,452	4,991	8,308	6,572	7,051	6,041	7,587	5,358	5,742	4,467	1,275
	Train	800	5,947	3,986	6,652	5,220	5,622	4,823	6,062	4,297	4,607	3,594	1,013
	Valid	100	748	502	823	662	724	621	758	538	577	440	137
	Test	100	757	503	833	690	705	597	767	523	558	433	125
EN	Total	1,000	7,452	4,991	8,264	6,434	6,933	5,894	7,432	4,994	5,514	4,287	1,227
	Train	800	5,947	3,986	6,613	5,109	5,523	4,699	5,931	3,989	4,414	3,442	972
	Valid	100	748	502	822	644	719	603	750	509	555	423	132
	Test	100	757	503	829	681	691	592	751	496	545	422	123

As depicted in Figure 3, our dataset is characterized by its multilingual and multi-sentiment attributes within dialogue scenarios, making it one of the most comprehensive and challenging datasets available for the ABSA task.

	ASTE	TOWE	MAMS	CASA	DiaASQ
Target	✗	✗	✗	✓	✓
Aspect	✓	✓	✓	✗	✓
Opinion	✓	✓	✓	✓	✓
Polarity	✓	✗	✓	✓	✓
Dialogue-level	✗	✗	✗	✓	✓
Multi-sentiment	✗	✗	✓	✗	✓
Multilingual	✗	✗	✗	✗	✓

Figure 3: A comparison between our DiaASQ dataset and existing popular ABSA datasets.

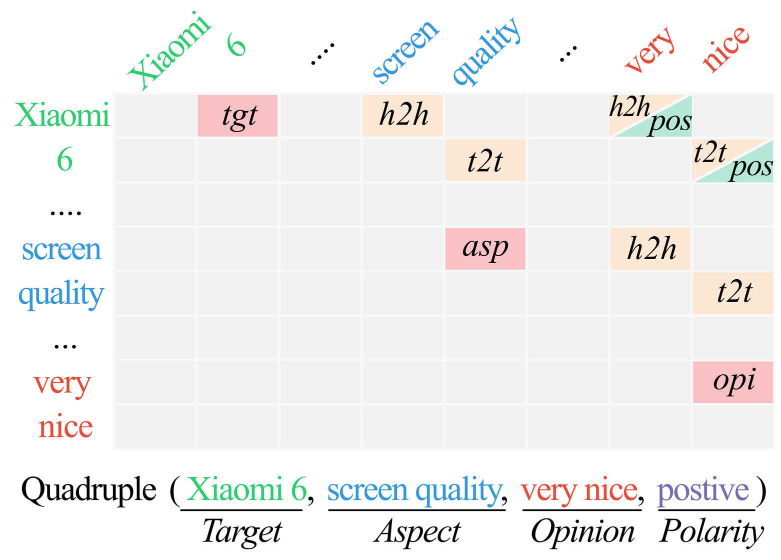

Figure 4: Tagging scheme for quadruple extraction.

▶ 3. Model Framework

To benchmark the task, we propose an end-to-end framework for DiaASQ extraction, which involves four main steps:

- Encoding the original dialogue text using pre-trained language models.
- Enhancing dialogue structure within threads, speakers, and replies information through multi-view interactions.
- Enhancing relative position between token pairs using Rotational Position Encoding (RoPE).
- Introducing a novel grid-based schema to encode and decode quadruplets with token-pair relations.

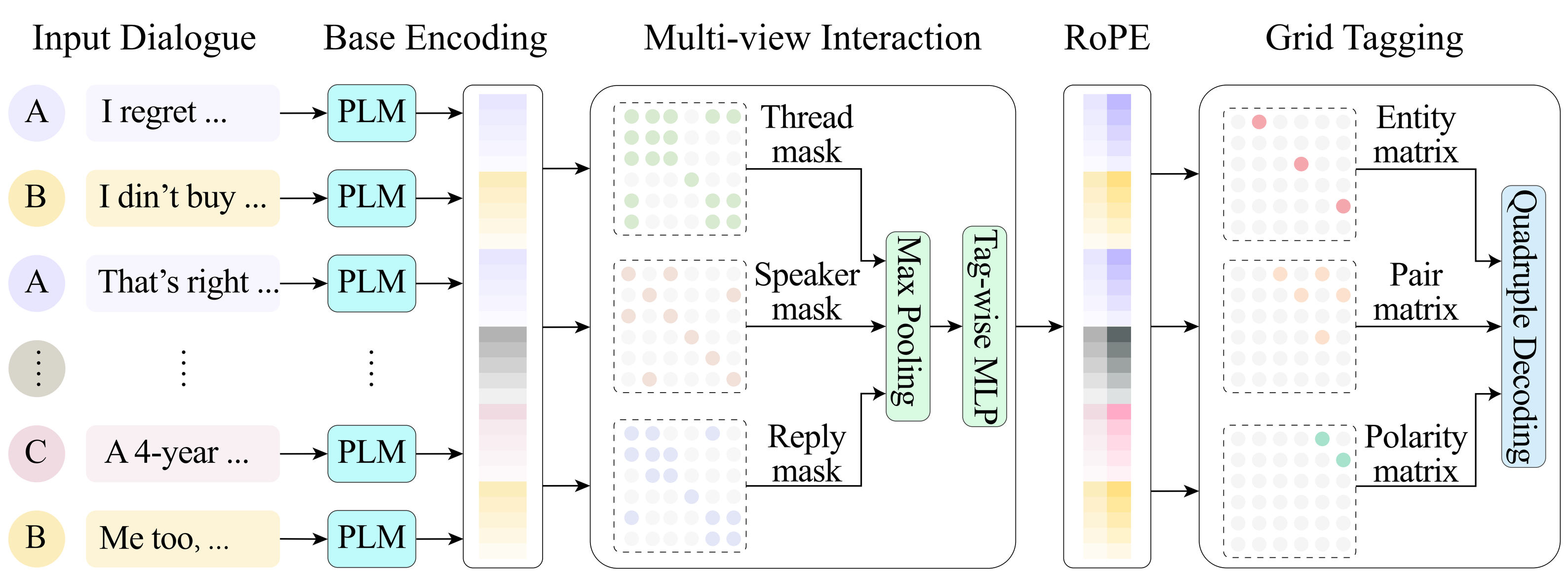

Figure 5: The overall framework of our DiaASQ model.

▶ 4. Experiments

We conducted experiments on two datasets and arrived at the following conclusions:

- Our model demonstrates superior performance compared to other models, providing evidence of its superiority over baselines.
- However, the absolute scores for quadruple extraction remain relatively low, highlighting the challenges of our task.

Table 2: Main results of the DiaASQ task.

		Span Match (F1)			Pair Extraction (F1)			Quadruple (F1)	
		T	A	O	T-A	T-O	A-O	Micro	Iden.
ZH	CRF-Extract-Classify	**91.11**	75.24	50.06	32.47	26.78	18.90	8.81	9.25
	SpERT	90.69	76.81	54.06	38.05	31.28	21.89	13.00	14.19
	ParaPhrase	/	/	/	37.81	34.32	27.76	23.27	27.98
	Span-ASTE	/	/	/	44.13	34.46	32.21	27.42	30.85
	w/o PLM	/	/	/	28.36	24.81	22.50	8.96	11.21
	Ours	90.23	**76.94**	**59.35**	**48.61**	**43.31**	**45.44**	**34.94**	**37.51**
	w/o PLM	85.52	75.21	47.15	34.72	26.16	30.73	14.21	17.55
EN	CRF-Extract-Classify	88.31	71.71	47.90	34.31	20.94	19.21	11.59	12.80
	SpERT	87.82	74.65	54.17	28.33	21.39	23.64	13.07	13.38
	ParaPhrase	/	/	/	37.22	32.19	30.78	24.54	26.76
	Span-ASTE	/	/	/	42.19	30.44	**45.90**	26.99	28.34
	w/o PLM	/	/	/	27.26	20.63	44.62	13.84	14.17
	Ours	**88.62**	**74.71**	**60.22**	**47.91**	**45.58**	44.27	**33.31**	**36.80**
	w/o PLM	83.02	68.89	53.87	32.53	31.09	35.59	15.68	19.57

In-depth Analysis: We conduct an in-depth analysis and gain a deep understanding of the strengths of our method:

- Our model excels in cross-utterance quadruple extraction, outperforming baselines even for high cross-utterance levels, benefiting from dialogue-specific interaction features and RoPE.
- Dialogue-level distance encoding enhances conversational discourse understanding compared to alternatives such as relative position encoding and global position encoding.

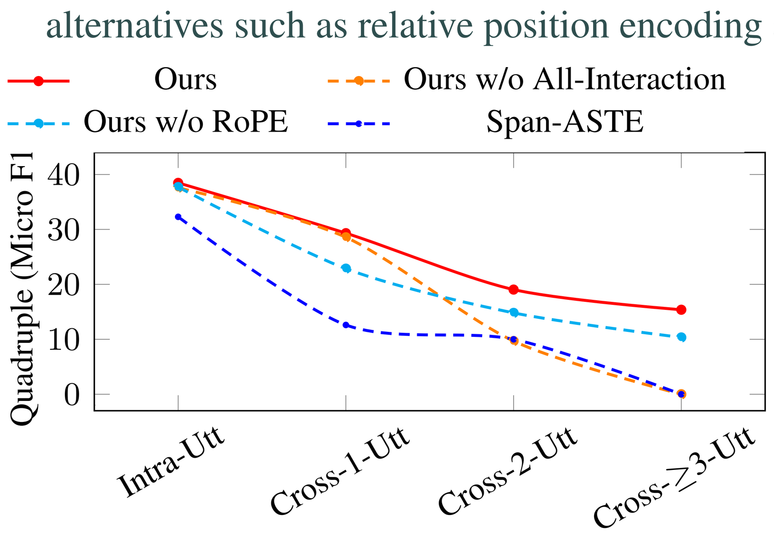

Figure 6: Results on different cross-utterance levels.

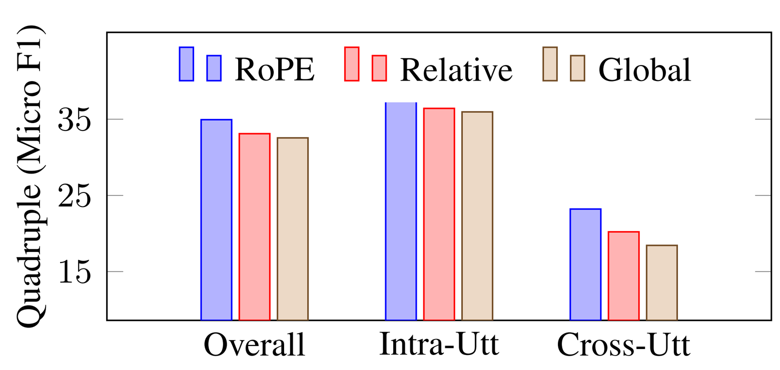

Figure 7: Influences of using difference distance-encoding methods.