MuSe-CarASTE: A comprehensive dataset for aspect sentiment triplet extraction in automotive review videos
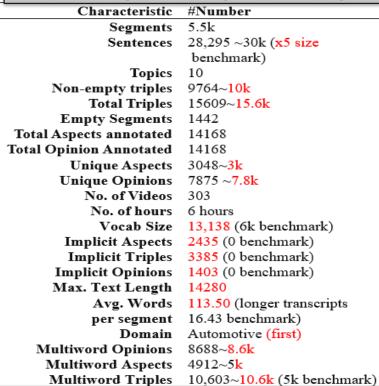
Atiya Usmani, Saeed Hammood Al Samhi, Muhammad Jaleed Khan, John Breslin, Edward Curry
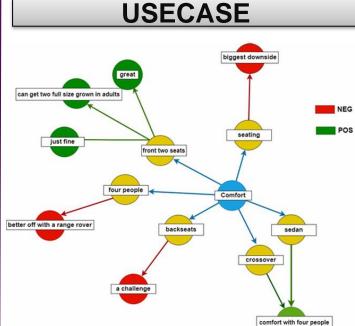
KEY CONTRIBUTIONS

- **A new benchmark for Aspect Sentiment Triplet Extraction.**
- **First** ASTE Dataset in Automotive Domain.
- **Largest** ASTE Repository till date with annotations for over **28,295** sentences.
- Complex **implicit** non-verbatim labels.
- **Domain: Aspect-based sentiment analysis, ASTE, Opinion Mining, Recommender System.**
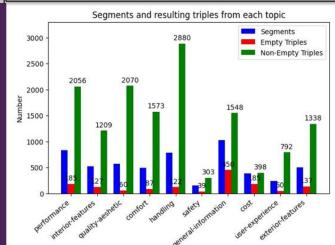
DATASET PROPERTIES

Characteristic	#Number
Segments	5.5k
Sentences	28,295 ~30k (x5 size benchmark)
Topics	10
Non-empty triples	9764~10k
Total Triples	15609~15.6k
Empty Segments	1442
Total Aspects annotated	14168
Total Opinion Annotated	14168
Unique Aspects	3048~3k
Unique Opinions	7875 ~7.8k
No. of Videos	303
No. of hours	6 hours
Vocab Size	13,138 (6k benchmark)
Implicit Aspects	2435 (0 benchmark)
Implicit Triples	3385 (0 benchmark)
Implicit Opinions	1403 (0 benchmark)
Max. Text Length	14280
Avg. Words per segment	113.50 (longer transcripts 16.43 benchmark)
Domain	Automotive (first)
Multiword Opinions	8688~8.6k
Multiword Aspects	4912~5k
Multiword Triples	10,603~10.6k (5k benchmark)

An ASTE Example

$$(\text{gearbox, rubbish, NEG})$$
$$(\text{steering, light, POS})$$
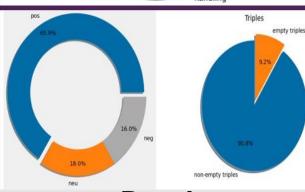
USECASE

BASELINES

- **BARTABSA-** is a **pointer-based** indice generation method
- **BMRC-** Machine Comprehension based
- **GAS-** LLM (T5) based generation
- **Span-ASTE** Tagging-based span prediction method

DATA DISTRIBUTION

Segments and resulting triples from each topic

DATASET ANNOTATION

1. **FORM ANNOTATION GUIDELINES**
2. **GET ANNOTATIONS**
3. **VERIFY ANNOTATIONS**

$$TPA = \frac{No.\ of\ agreeing\ triples > \tau}{Total\ No.\ of\ triples\ in\ the\ dataset * 100}$$

Table 7: Simple Triple-Wise Percentage Agreement between annotators.

Similarity Threshold τ	Triple Percentage Agreement %	Aspect Opinion Pair Agreement %
0.60	79.74138	79.74138
0.65	73.41954	74.28161
0.70	67.52874	67.81609

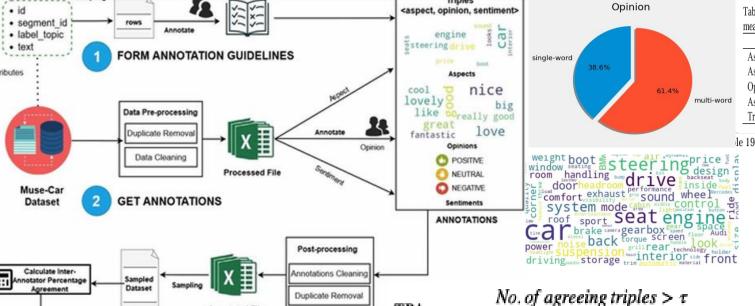Sentiment Agreement Percentage in Triples: 73.069

Results

Table 14: Detailed Result of baseline model (Zhang et al., 2021) on our dataset using precision, recall, and F1 measures up to 4 decimal places on dev file of whole dataset corresponding to the original MuSe-Car dataset

	precision	recall	F1
Aspect	0.7143	0.7196	0.7169
Aspect-Sentiment Pair	0.7611	0.7667	0.7639
Opinion	0.9126	0.9193	0.9156
Aspect-Opinion Pair	0.9272	0.9341	0.9306
Triple	0.9300	0.9232	**0.9266**

Table 19: Results of baseline models on on our dataset[1] and SemEval Datasets[2] using F1 scores.

Model	Dataset	Aspect (F1)	Opinion (F1)	Aspect Opinion Pair (F1)	Aspect Sentiment Pair (F1)	Triple (F1)
Generative-ABSA (Zhang et al., 2021)	14 lap	0.63	0.61	0.52	0.48	0.43
	14 res	0.66	0.71	0.69	0.63	0.65
	15res	0.66	0.71	0.60	0.61	0.56
	16res	0.67	0.75	0.67	0.62	0.63
	Muse-ASTE	**0.31**	**0.34**	**0.26**	**0.26**	**0.218**
BMRC (Chen et al., 2021)	14 lap	0.76	0.73	0.67	0.66	0.59
	14 res	0.82	0.84	0.76	0.76	0.71
	15res	0.72	0.78	0.66	0.658	0.61
	16res	0.82	0.83	0.76	0.73	0.68
	Muse-ASTE	**0.786**	**0.751**	**0.757**	**0.761**	**0.568**
BART-ABSA (Yan et al., 2021)	14 lap	0.79	0.84	0.68	0.69	0.60
	14 res	0.85	0.84	0.75	0.78	0.71
	15res	0.78	0.61	0.56	0.54	0.50
	16res	0.85	0.84	0.75	0.76	0.68
	Muse-ASTE	0.438	0.513	0.290	0.404	0.249

PROJECT LINK: