

# OATS: Opinion Aspect Target Sentiment Quadruple Extraction Dataset for Aspect-Based Sentiment Analysis

---

Siva Uday Sampreeth Chebolu<sup>1</sup>, Franck Deroncourt<sup>2</sup>, Nedim Lipka<sup>2</sup>, Thamar Solorio<sup>1,3</sup>

<sup>1</sup>University of Houston, <sup>2</sup>Adobe Research, <sup>3</sup>MBZUAI





# Corpus of **O**pinion **A**spect **T**arget **S**entiment (OATS) for ABSA

## Motivation

- Limited datasets for joint detection evaluation
- Missing implicit opinion expression annotations
- Missing aggregated review level opinions

<b>Dataset</b>	<b>#sents</b>	<b>#pos</b>	<b>#neg</b>	<b>#neu</b>	<b>#Total</b>
Rest-15	1562	1710	701	85	2496
Rest-16	2024	2293	877	125	3295
<b>Total</b>	<b>3586</b>	<b>4003</b>	<b>1578</b>	<b>210</b>	<b>5791</b>

Statistics of the dataset with quadruple annotations

# Corpus of **O**pinion **A**spect **T**arget **S**entiment (OATS) for ABSA

## Annotation Procedure

- **Step 1:** 3 Annotators annotate portion of data
  - 50 reviews from each domain
  - Clear disagreements
  - Compute IAA F1-score
- **Step 2:** Full data
  - Divided among three annotators

Dataset	Tgt. Span	Op. Ph.	Asp-Sen	Quadruple
Amazon_FF	72.57	69.72	85.43	<b>65.78</b>
Coursera	78.26	71.26	79.63	68.56
Hotels	74.78	72.05	<b>87.32</b>	73.84

### Inter Annotator Agreement F1-scores for OATS

# Corpus of Opinion Aspect Target Sentiment (OATS) for ABSA

## Sentence Level Opinion Quadruples

Sent1: Well organized.

{ NULL, COURSE#QUALITY, Well Organized, Positive }

Sent2: Interactive .

{ NULL, FACULTY#RESPONSE, Interactive, Positive }

Sent3: I felt like a journalist !

Sent4: Thank Margaret & Denis for organizing this course .

{ Margaret & Denis, FACULTY#GENERAL, Thank, Positive }

Sent5: Was such a pleasure to participate.

{ NULL, COURSE#GENERAL, pleasure to participate, Positive }

## Review Level Aggregated Opinions

Review: Well organized . Interactive . I felt like a journalist ! Thank Margaret & Denis for organizing this course . Was such a pleasure to participate.

{ COURSE#QUALITY, Positive },  
{ FACULTY#RESPONSE, Positive },  
{ COURSE#GENERAL, Positive },  
{ FACULTY#GENERAL, Positive }

# Corpus of Opinion Aspect Target Sentiment (OATS) for ABSA

## Overall statistics

Domain	#Rev	#Sent	#Rev Ops	#Sent Ops	#Total Ops
Amazon_FF	1794	8913	4326	8260	12586
Coursera	1702	8278	5350	7875	13225
Hotels	1497	7963	7416	11335	18751
<b>Total</b>	<b>4993</b>	<b>25154</b>	<b>17092</b>	<b>27470</b>	<b>44562</b>

## #Opinions from the Sentiment Polarity perspective

Domain	Sentence-Level ABSA			Review-Level ABSA			
	#pos	#neg	#neu	#pos	#neg	#neu	#conf
Amazon_FF	5577	1187	234	2900	606	74	82
Coursera	4403	1008	213	2910	721	129	71
Hotels	6952	1207	169	4557	817	110	53
<b>Total</b>	<b>16932</b>	<b>3402</b>	<b>616</b>	<b>10367</b>	<b>2144</b>	<b>313</b>	<b>206</b>



# Corpus of **O**pinion **A**spect **T**arget **S**entiment (OATS) for ABSA

## Baseline Experiments for OATS

### Data Split

- Train (80%), Validation (10%), and Test (10%)

### Evaluation Metric

- F1-score

### BERT-based Methods

- BDTF (BERT)
- TAS-BERT-\*
- BMRC
- ABSA-BERT-pair-\*
- QACG-BERT-NLI-M

### Generative Approaches

- GAS
- Paraphrase
- T5-ABSA
- Template-\*
- GEN-SCL-NAT



# Corpus of **O**pinion **A**spect **T**arget **S**entiment (OATS) for ABSA

## Results for Quadruple & Triplet Extraction Tasks

ASQP <sup>2</sup>				TASD			
Method	A	C	H	Method	A	C	H
GAS-T5	19.62	<b>22.23</b>	26.33	TAS-BERT-BIO	45.12	44.41	45.92
GEN-NAT-SCL	20.36	20.20	28.58	TAS-BERT-TO	47.51	42.77	45.76
Template-DLO	20.39	21.96	26.59	T5-ABSA	<b>51.61</b>	<b>44.57</b>	<b>49.78</b>
Paraphrase	20.84	19.78	<b>34.51</b>	GAS	43.04	41.53	50.69
Template-ILO	<b>21.01</b>	21.24	24.62	Paraphrase	44.89	40.24	49.81

ASTE <sup>1</sup>			
Method	Am_FF	Crse	Htl
BDF	27.81	43.98	42.02
GAS	46.46	40.15	45.96
Paraphrase	46.64	39.30	46.37
BMRC	<b>49.48</b>	<b>50.85</b>	<b>50.71</b>



# Corpus of **O**pinion **A**spect **T**arget **S**entiment (OATS) for ABSA

## Results for Tuple Extraction Tasks

R-ASD			
Method	A	C	H
ABSA-BERT-pair-NLI-B	88.22	91.31	91.84
ABSA-BERT-pair-QA-B	<b>88.42</b>	<b>91.51</b>	<b>91.93</b>
QACG-BERT-NLI-M	87.15	90.41	90.05
-----			
T5-ABSA	58.26	40.76	54.09
GAS-T5	65.23	46.61	<u>58.61</u>
Paraphrase	<u>66.46</u>	<u>47.88</u>	56.11

S-ASD			
Method	A	C	H
ABSA-BERT-pair-NLI-B	93.25	95.92	96.82
ABSA-BERT-pair-QA-B	<b>93.61</b>	<b>96.69</b>	<b>97.18</b>
QACG-BERT-NLI-M	92.13	95.81	95.68
-----			
T5-ABSA	56.85	48.81	59.85
GAS-T5	<u>57.57</u>	<u>56.79</u>	<u>66.91</u>
Paraphrase	57.25	55.27	66.38

27 \* 4 \*  
2000 =  
216000

# Corpus of **O**pinion **A**spect **T**arget **S**entiment (OATS) for ABSA

## Analysis

Domain	ET / IT	EO / IO	ET-EO	ET-IO	IT-EO	IT-IO
Amazon_FF	2,999 / 5,261	6,780 / 1,480	2,491	508	4,298	972
Coursera	5,163 / 2,712	6,185 / 1,690	4,222	941	1,963	749
Hotels	8,654 / 2,820	10,193 / 1,281	7,927	727	2,266	554

Distribution of Explicit and Implicit Targets (ET and IT) & Opinion Phrases (EO and IO)

Domain/Task	TOWE	TSD	ASD	TASD	ASTE	ASQP
Amazon_FF	31.29	65.27	57.25	44.89	46.64	20.84
Coursera	38.64	67.75	55.27	40.24	39.3	19.78
Hotels	35.75	62.34	66.38	49.81	46.37	34.51

Performance Analysis on six main ABSA joint tasks using the Paraphrase method

# Corpus of **O**pinion **A**spect **T**arget **S**entiment (OATS) for ABSA

## Significance

- First manually annotated corpus with all elements of ABSA
- Has review-level aggregated annotations for aspect category and sentiment tuples
- Has  $\approx 44$ K opinions (27K Quadruples + 17K Tuples)
- Facilitates cross-task transfer for underperforming subtasks
- Helps evaluate all the 14 ABSA subtasks including the Quad-Prediction

Thank you for you attention!  
*Any questions?*

