

Task and Data

- Multi-Perspective Question Answering (MPQA)
 - A comprehensive analysis
 - Presented a new methodology
 - Utilized the most extensive subset of MPQA items compared to previous studies (Wu et al., 2022; Xia et al., 2021; Marasovic and Frank, 2018; Katiyar and Cardie, 2016)

Annotation Type	Original	Our Version
Agents	14,595	14,562
Targets	8,413	8,397
ESE	13,793	13,654
Sources in ESE	21,497	21,411
DS	15,437	15,076
Sources in DS	31,253	30,567
Attitudes in DS	10,336	9,973
OSE	16,906	15,959
Sources in OSE	22,542	20,794
Attitudes	10,308	10,292
Targets in Attitudes	9,466	9,056
Sentences	15,789	15,789

Opinion Analysis

- Fine-grained annotation scheme
- Detecting opinion expressions
- Opinion role labeling

Results

Type classification

Model Name	F1
Majority Baseline	63.7
BERT-based model	83.5
T5-based model	83.1
FLAN-T5-based model	83.8

Polarity classification

Model Name	F1
Majority Baseline	24.9
BERT-based model	86.1
T5-based model	87.0
FLAN-T5-based model	87.1

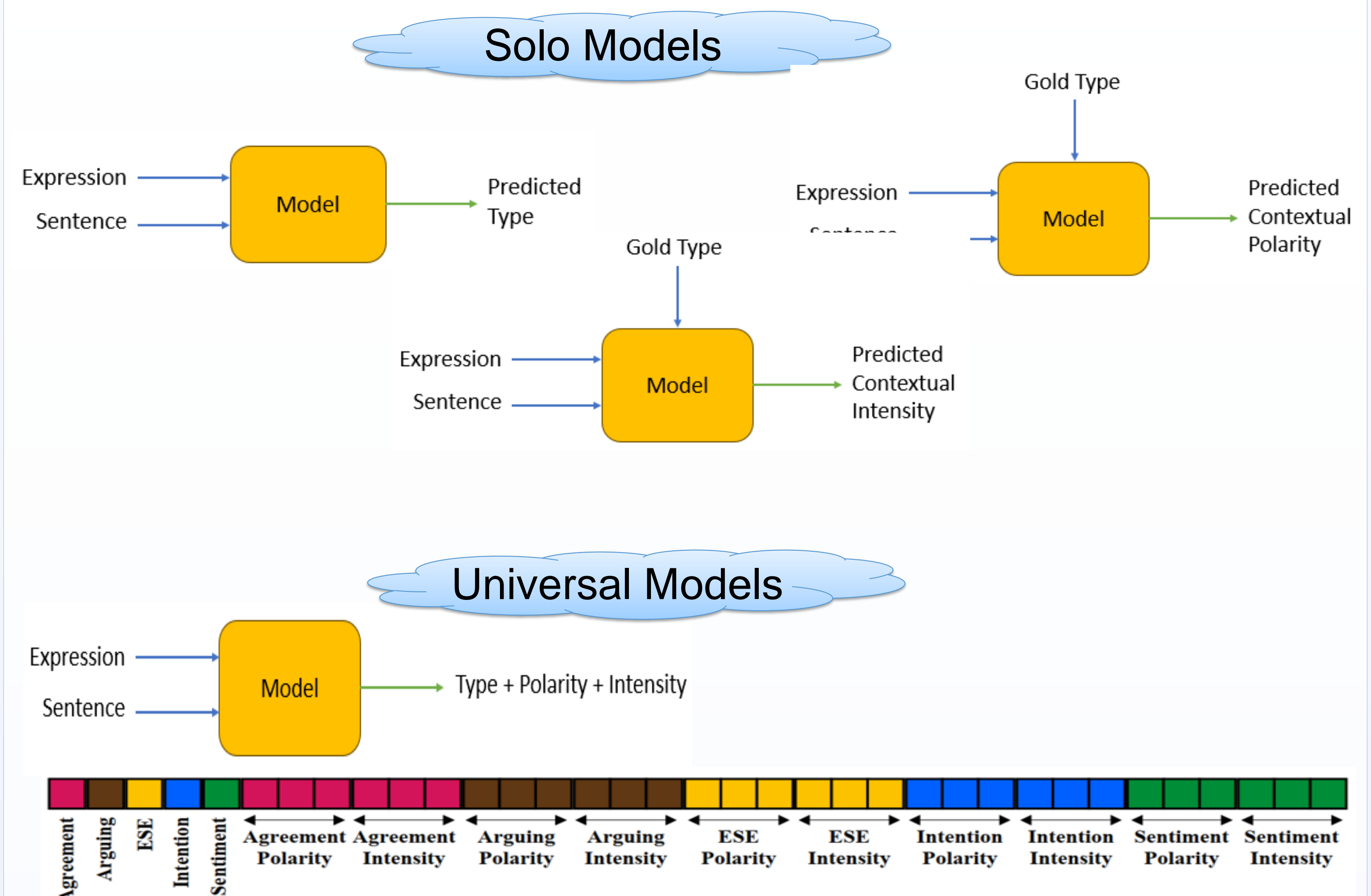
Intensity classification	Model Name	F1
	Majority Baseline	39.5
	BERT-based model	72.5
	T5-based model	72.5
	FLAN-T5-based model	73.0

Expression detection	Model Name	Type	Exact F1	Binary F1	Proportional F1	
	T5: Prefix + Sentence	ESE		54.4	74.2	72.6
		DS		70.6	80.2	79.8
		OSE		80.7	82.9	82.9
	FLAN-T5: Prefix + Sentence	ESE		51.6	70.3	68.4
		DS		65.8	75.2	74.7
OSE			78.1	80.0	79.9	

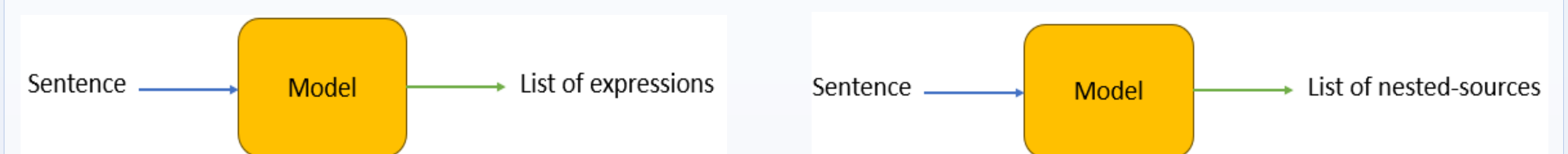
Source detection	Model Name	Comparison Method	Exact F1	Binary F1	Proportional F1	
	T5: Sentence	Last source in the list		73.9	81.5	81.3
		Entire ordered list of nested sources		71.0	83.7	83.4
	T5: Prefix + Sentence	Last source in the list		83.5	86.1	86.1
Entire ordered list of nested sources			81.8	87.7	87.6	

Proposed Method

- Classify each of the type, polarity, and intensity elements contextually



- Predict opinion expressions and agents containing ESE, DS and OSE



Error Analysis

- For Type, Polarity and Intensity
 - Analyzing the F1 of all/each type(s) in the TPI tasks on the validation set

Type	Count	F1 (Type)	F1 (Polarity)	F1 (Intensity)
All	All	85.1	86.7	73.5
Agreement	47	77.9	97.9	78.0
Arguing	389	71.7	94.6	82.5
ESE	1,851	92.6	81.4	67.0
Intention	69	84.1	98.5	88.0
Sentiment	618	71.2	95.1	85.1

- For expression prediction

- 71% do not overlap in any tokens
- An specific set of words cause discrepancies in the rest 29%

- For source prediction

- 94% do not overlap at all
- Inconsistency in nested-sources' lengths
- Most errors occur in the last source in nested-source structure

Conclusions

- A comprehensive opinion mining using pre-trained LLMs
- Re-examined MPQA 2.0 corpus
- Provide useful tools for working with MPQA data
- Set new high baselines for future research