Open Aspect Target Sentiment Classification with Natural Language Prompts

University of Massachusetts Amherst

LEXALYTICS

Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, Alfred Hough

Main Idea

Review: Mike and other staff were very nice and polite.

Prompt: The front desk service is excellent.

What if we could design an ATSC model that could...

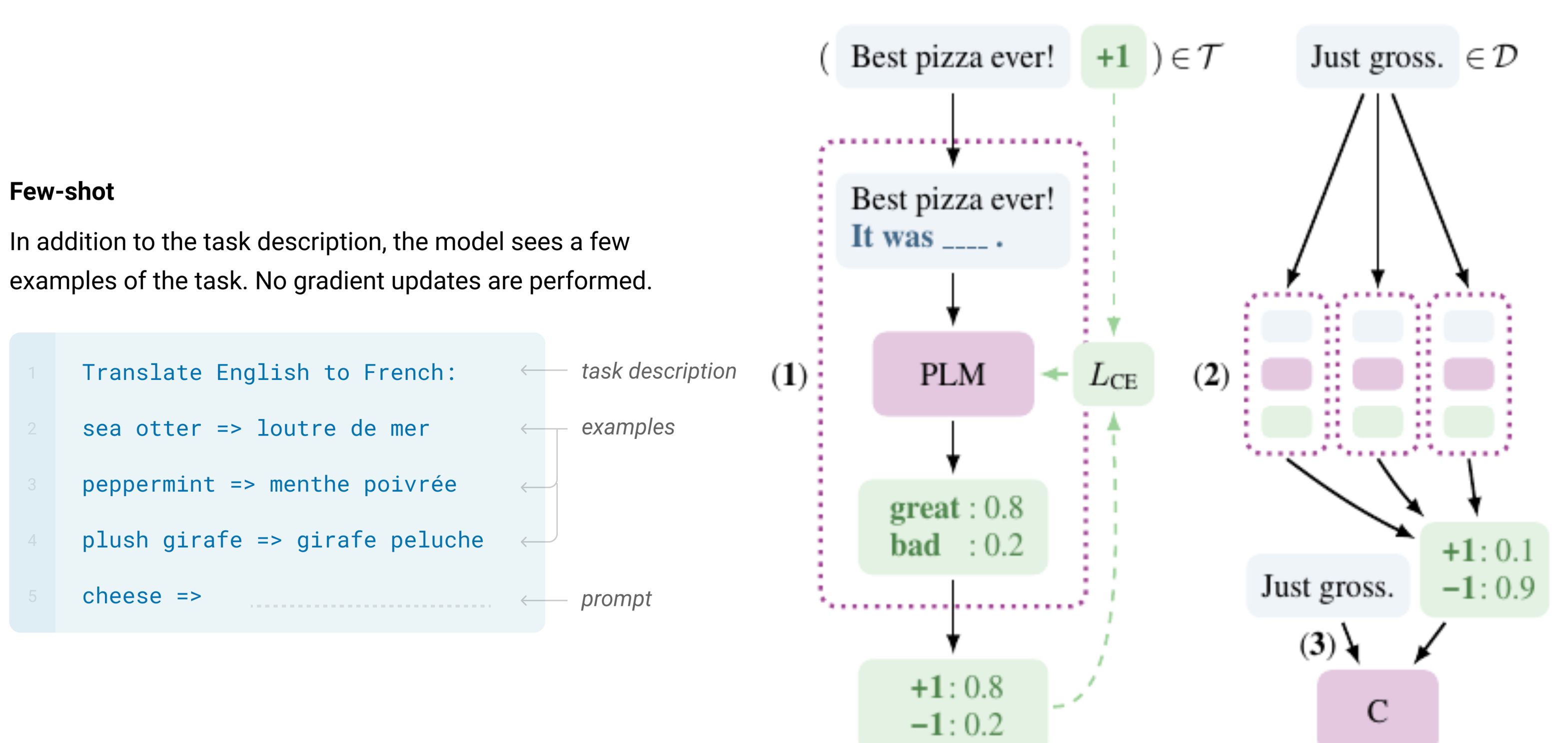
1) Predict whether the second sentence **naturally follows** the first?

2) Predict whether the first **entails** the second?

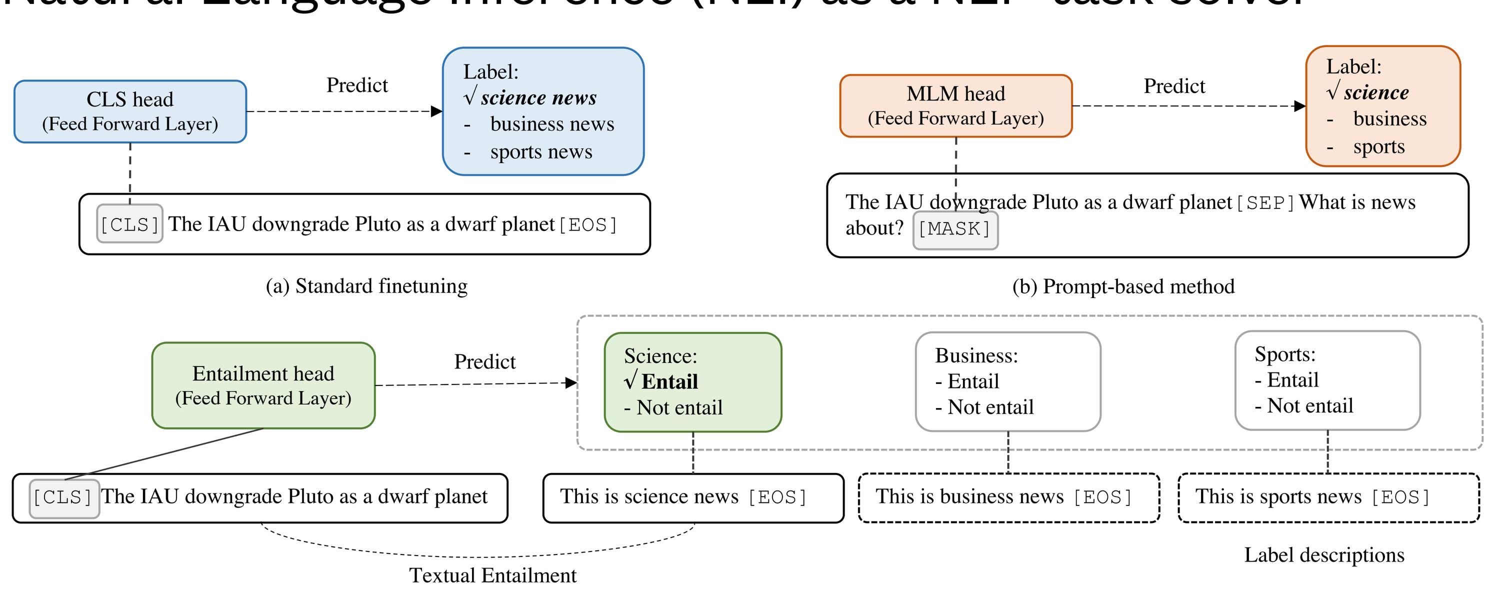
Related Work

Aspect Target Sentiment Classification (ATSC)

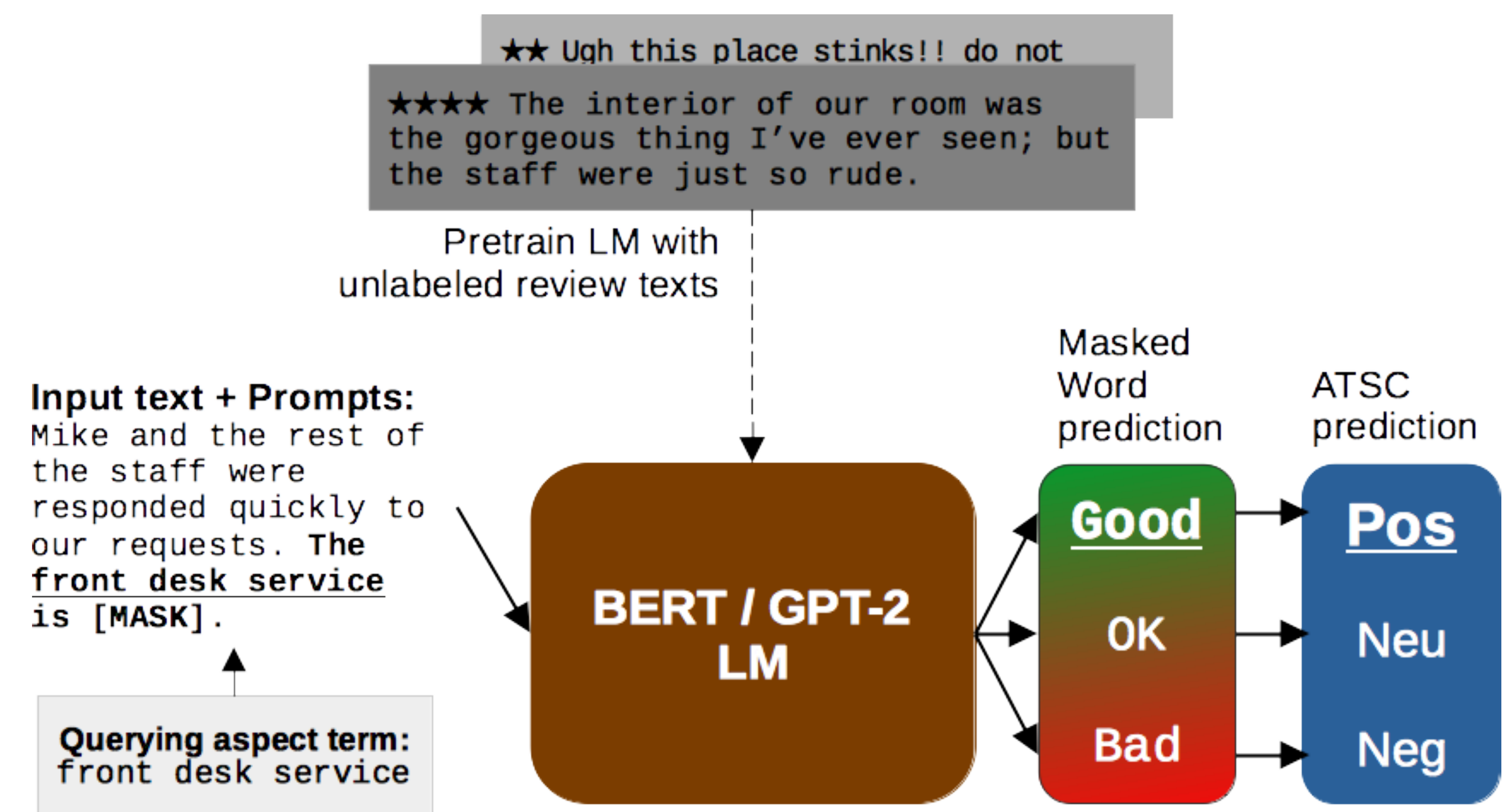
Language Model (LM) Prompting



Natural Language Inference (NLI) as a NLP task solver

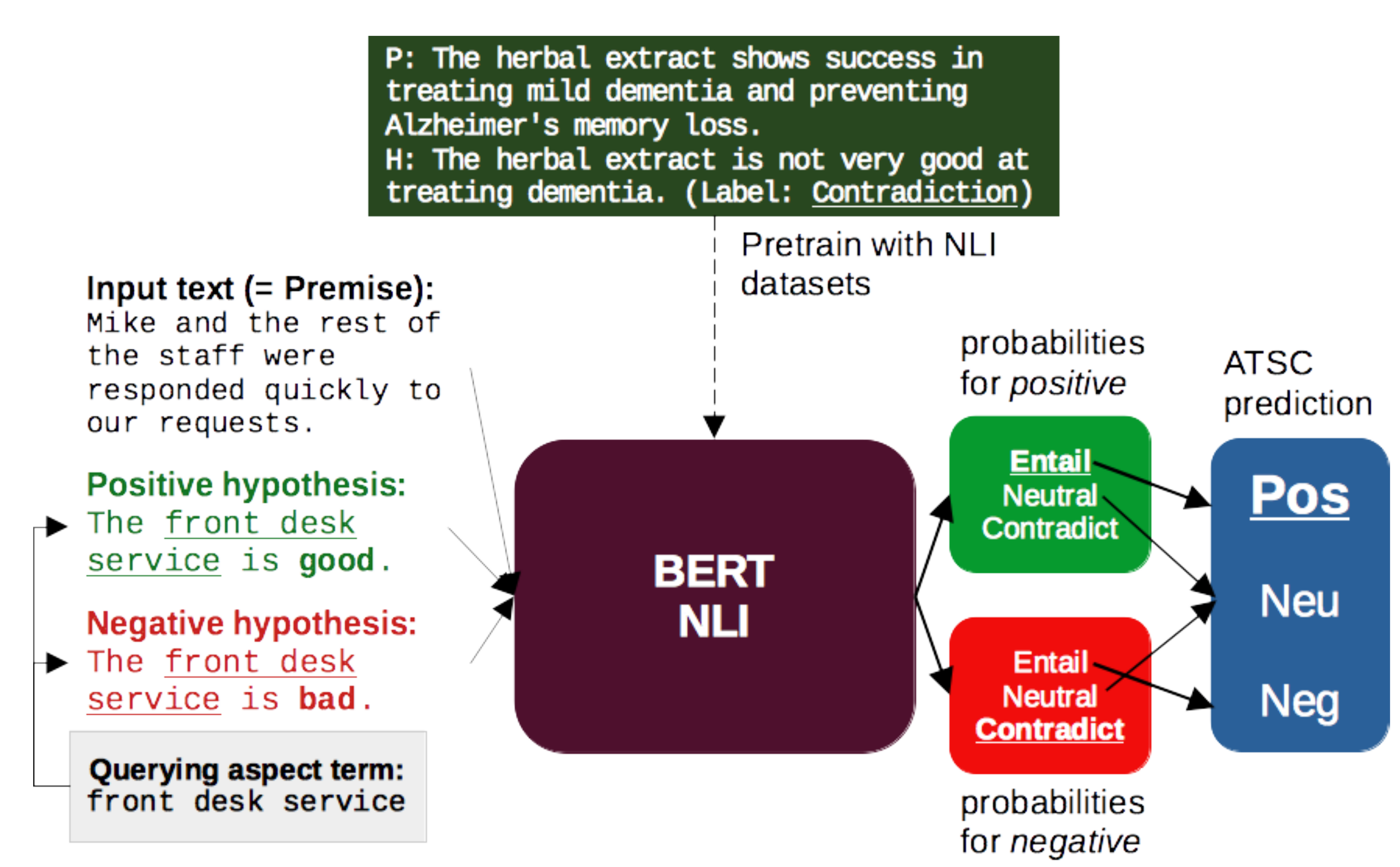


ATSC as a LM task



- Append cloze question prompts to the review text and predict how likely it is to observe *good*, *bad*, and *ok* as the next word.
- LMs are pretrained with unlabeled shopping review corpora

ATSC as a NLI task



- Treat the review as a premise, and the prompt sentence with the sentimental next word as a hypothesis
- Predict whether the review entails the prompt.
- BERT NLI is pretrained with MNLI

Results

Number of Training Examples

Model	Zero Acc	Zero MF1	16 Acc	16 MF1	64 Acc	64 MF1	256 Acc	256 MF1	1024 Acc	1024 MF1	Full (1850) Acc	Full (1850) MF1
BERT-ADA	–	–	–	–	–	–	–	–	–	–	79.19†	74.18†
BERT [CLS]	–	–	48.75	34.92	60.63	49.43	72.35	64.31	76.87	71.22	80.06	75.08
BERT NSP	–	–	48.24	31.35	60.91	49.27	72.38	64.64	76.77	71.12	80.25	75.46
BERT LM	**63.58**	46.17	69.05	58.60	72.80	65.54	**76.59**	70.65	**79.30**	74.80	**81.10**	76.83
	–	–	+20.30*	+23.68*	+11.89*	+16.11*	**+4.21***	+6.01*	**+2.43***	+3.58*	**+0.85***	+1.37*
GPT-2 LM	60.45	39.59	68.94	56.71	71.54	63.69	76.48	70.89	79.02	**74.88**	80.73	**77.13**
	–	–	+20.19*	+21.79*	+10.63*	+14.26*	+4.10*	+6.25*	+2.15*	+3.66*	+0.48	**+1.67***
BERT NLI	58.93	**54.91**	**72.88**	**68.06**	**74.95**	**70.84**	76.22	**71.65**	77.42	73.52	77.58	73.18
	–	–	**+24.13***	**+33.14***	**+14.04***	**+21.41***	+3.84*	**+7.01***	+0.55	+2.30	-2.67	-2.28

(a) Laptops

Model	Zero Acc	Zero MF1	16 Acc	16 MF1	64 Acc	64 MF1	256 Acc	256 MF1	1024 Acc	1024 MF1	Full (3602) Acc	Full (3602) MF1
BERT-ADA	–	–	–	–	–	–	–	–	–	–	87.14†	80.05†
BERT [CLS]	–	–	59.89	34.50	73.00	50.79	79.45	64.70	83.48	73.62	86.77	79.33
BERT NSP	–	–	61.05	32.46	74.73	53.00	79.34	65.51	83.61	74.15	87.09	79.98
BERT LM	70.86	48.17	71.99	56.65	77.79	63.30	81.10	69.27	85.12	76.60	**87.50**	**80.78**
	–	–	+10.94*	22.15*	+3.06*	+10.30*	+1.65*	+3.76*	+1.51*	+2.45*	**+0.41**	**+0.80**
GPT-2 LM	**71.40**	45.53	**75.41**	60.06	79.30	65.49	**82.27**	71.62	**85.28**	**77.38**	86.99	80.02
	–	–	**+14.36***	+25.56*	+4.57*	+12.49*	**+2.82***	+6.11*	**+1.67***	**+3.23***	-0.1	+0.04
BERT NLI	61.79	**57.93**	74.74	**65.58**	**79.33**	**69.44**	81.24	**71.94**	83.07	74.52	85.07	77.53
	–	–	+13.69*	**+31.08***	**+4.60***	**+16.44***	+1.79*	**+6.43***	-0.54	+0.37	-2.02	-2.45

(b) Restaurants

Model	In/Cross	16 Restaurants	16 Laptops	Full Restaurants	Full Laptops
BERT NSP	In	61.05	48.24	87.09	80.25
	Cross	49.55	47.46	81.29	78.56
BERT LM	In	71.99	69.05	87.50	81.10
	Cross	75.21	68.17	81.27	79.03
BERT NLI	In	74.74	72.88	85.07	77.58
	Cross	77.45	70.43	80.35	76.61

Table 2: Accuracies of BERT NSP, LM, and NLI trained with in-domain and cross-domain data.

Model	16 Acc	16 MF1	Full Acc	Full MF1
BERT NSP	64.73	33.22	82.45	70.91
BERT LM	76.67	56.77	84.31	74.14
BERT NLI	66.92	58.18	67.42	59.24

Table 3: Performance on ACSC without any extra training. Refer to Appendix E for the results with other training set sizes.

Future Work

- Adapt our prompts to **jointly perform** aspect term extraction and sentiment classification (Luo et al. (2020))
- Explore potential ways of **combining ATSC, LM, and NLI** tasks into an unified task
- Determine whether there are any strong **linguistic patterns** among the prompt models' predictions

References

[1] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis.
[2] Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. 2020. Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification.
[3] Timo Schick and Hinrich Schütze. 2020. Exploiting cloze questions for few-shot text classification and natural language inference.
[4] Sinong Wang, Han Fang, Madian Khabsa, Hanzi Mao, and Hao Ma. 2021. Entailment as few-shot learner.