A Generative Language Model for Few-Shot Aspect-Based Sentiment Analysis

Ehsan Hosseini-Asl, Wenhao Liu, Caiming Xiong

salesforce AI RESEARCH

Overview

Motivation

- Sentiment analysis is an important task in natural language processing.
- Aspect-based sentiment analysis, which involves extracting aspect term, category, and predicting their corresponding polarities.
- In recent works, pre-trained language models are often used to achieve state-of-the-art results, especially when training data is scarce.
- We are interested in few-shot settings.

Proposal

- Recasting aspect-based sentiment analysis as a simple, causal (unidirectional) language modeling task
- The model learns to accomplish the tasks via language generation without the need of training task-specific layers which is essential for few-shot learning

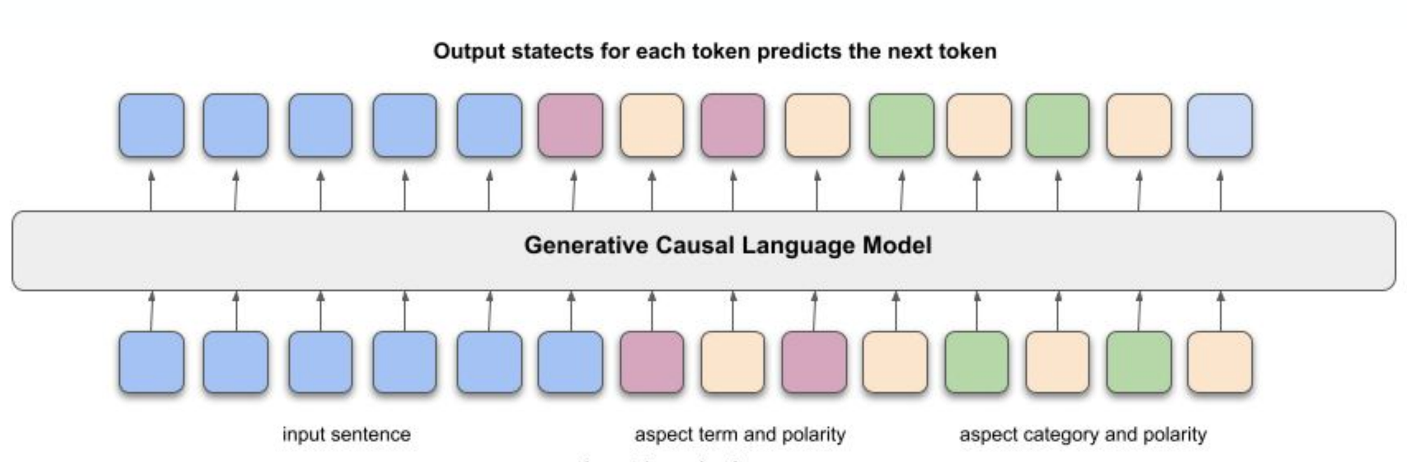
Algorithm

- A single training sequence consists of the concatenation of review sentence, aspect terms, term polarities, aspect category, category polarities.

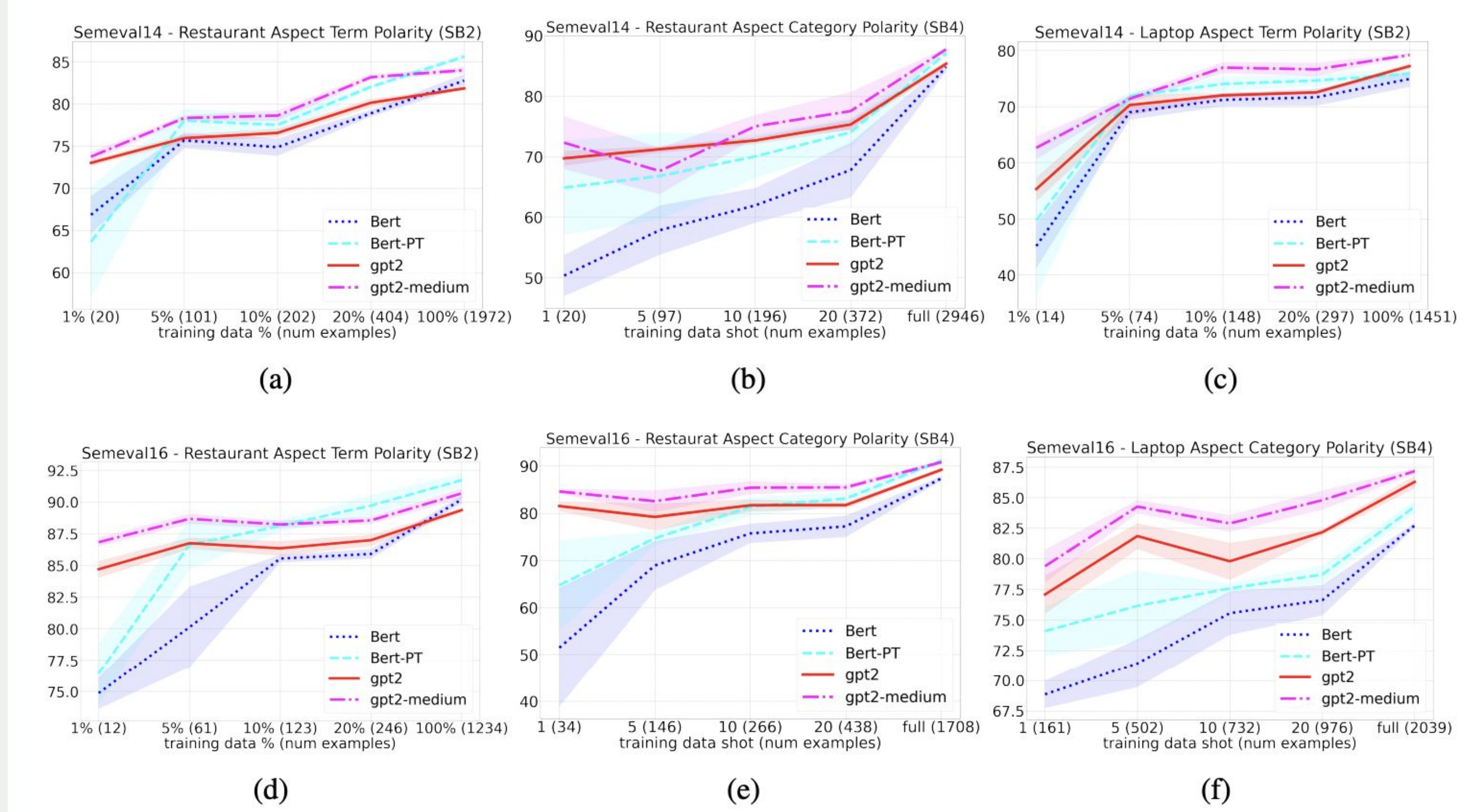
Sentence S^k	[review] *review sentence* [endofreview]
Aspect term T^k	[term] $term_1$ $polarity_1$, $term_2$ $polarity_2$, … $term_I$ $polarity_I$ [endofterm]
Aspect category C^k	[category] $category_1$ $polarity_1$, $category_2$ $polarity_2$, … $category_J$ $polarity_J$ [endofcategory]
Aspect term single and joint task training sequence (LM_{term})	[review] *review sentence* [endofreview] [term] $term_1$ $polarity_1$, … [endofterm]
Aspect category single and joint task training sequence ($LM_{category}$)	[review] *review sentence* [endofreview] [category] $category_1$ $polarity_1$, … [endofcategory]
Multi-task training sequence (LM_{multi})	[review] *review sentence* [endofreview] [term] $term_1$ $polarity_1$, … [endofterm] [category] $category_1$ $polarity_1$, … [endofcategory]

- Generative model is trained by minimizing the negative log likelihood over the joint sequence

$$p_\theta(x) = \prod_{t=0}^{n} p_\theta(x_t|x_{<t}) \qquad \mathcal{L}_D = -\sum_{k=1}^{K}\sum_{t=1}^{n} \log p_\theta(x_t^k|x_{<t}^k)$$



Single-task few-shot polarity prediction



Single-task polarity prediction: predicting the polarity of aspect terms or aspect categories

Inference: the input to the model (LM) comprises of k-th sentence and the corresponding aspect term or category

$$pt_i^k = LM_{term}(S^k, t_i^k) \qquad pc_j^k = LM_{category}(S^k, c_j^k)$$

LM_{term} : a single-task model that trained on aspect term dataset
$LM_{category}$: a single-task model that trained on to aspect category dataset

Joint- and Multi-task prediction

Joint-task: generating pairs of aspect term and term polarity, or pairs of aspect category and their polarity.

Multi-task: generating all pairs of aspect terms and aspect categories and their polarities.

Inference: the model input relies on the k-th review sentence only, and the model generates pairs in token-by-token (autoregressive) generation,

$$T^k = LM_{term}(S^k) \qquad C^k = LM_{category}(S^k) \qquad [T^k; C^k] = LM_{multi}(S^K)$$

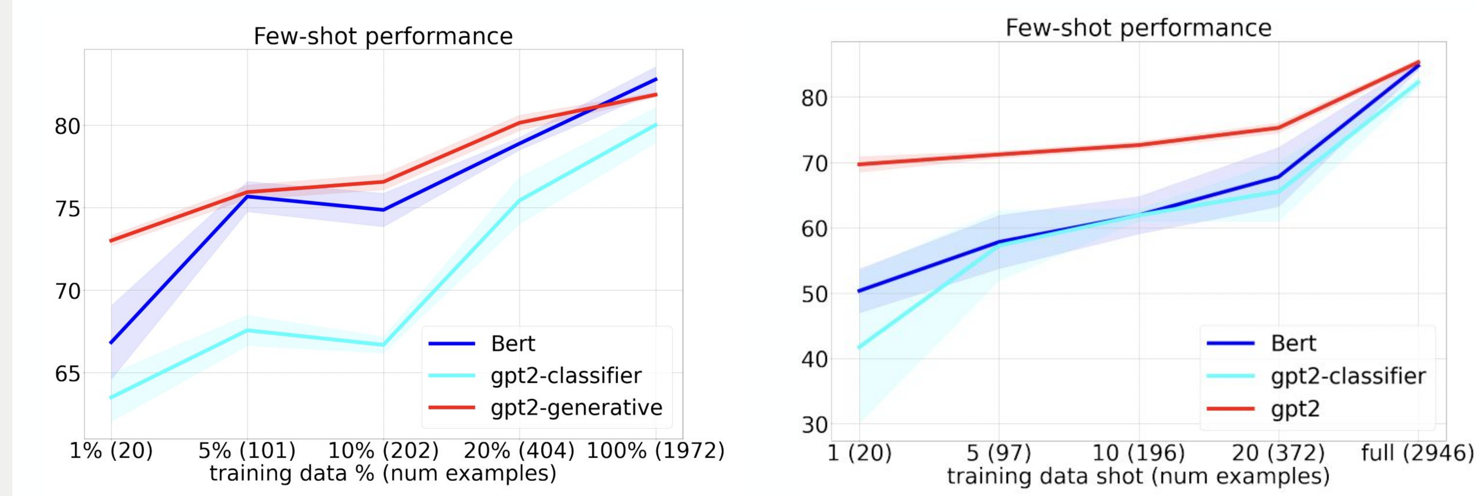
LM_{term} : a joint-task model that trained on aspect term dataset
$LM_{category}$: a joint-task model that trained on aspect category dataset
LM_{multi} : a multi-task model that trained on aspect term and aspect category dataset

Results

Method	Training Task	Model	Restaurant Joint Accuracy	Restaurant SB1 (F1)	Laptop Joint Accuracy	Laptop SB1 (F1)
Discriminative	Single (SB1)	MGAN	-	71.48	-	71.42
		BERT	-	74.1	-	79.28
		BERT-DK	-	77.02	-	83.55
		BERT-MRC	-	74.21	-	81.06
		BERT-PT	-	77.97	-	84.26
		BERT-PSUM	-	-	-	85.94
		BERT-HSUM	-	-	-	**86.09**
Generative	Joint (SB1&2)	GPT2 (base)	$56.47_{\pm 0.82}$	$77.59_{\pm 0.32}$	$50.65_{\pm 1.04}$	$72.61_{\pm 1.03}$
		GPT2 (medium)	$60.07_{\pm 0.52}$	$\mathbf{81.52}_{\pm 0.8}$	$53.55_{\pm 0.43}$	$75.94_{\pm 0.17}$
	Multi (SB1-4)	GPT2 (base)	$49.84_{\pm 1.03}$	$77.92_{\pm 0.53}$	-	-
		GPT2 (medium)	$54.43_{\pm 0.47}$	$\mathbf{82.04}_{\pm 0.21}$	-	-

SB1: aspect term extraction sub-task
Restaurant domain: joint- and multi-task model still outperforms previous single-task models
Previous Bert-based models: trained to solve single-task aspect term extraction only, on aspect term extraction

Ablation: Generative vs. Discriminative training of language model

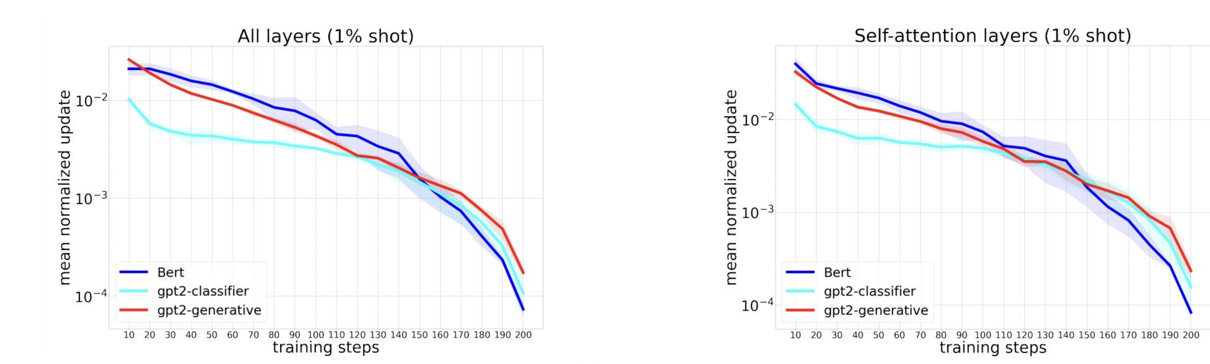


Restaurant aspect-term polarity prediction

Restaurant aspect-category polarity prediction

- when fine-tuning GPT2 model as a classifier on the downstream task using an classification layer, it under-performs BERT model on few and full-shot settings.

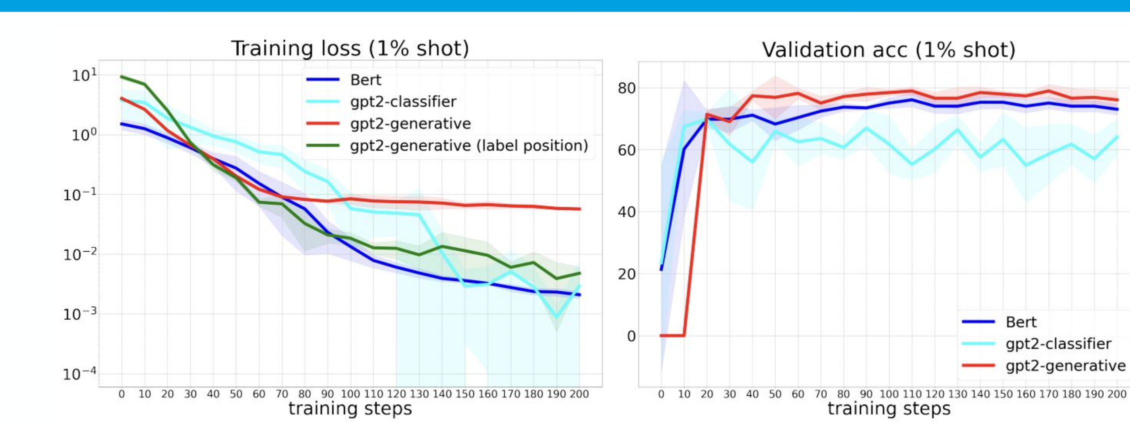
Ablation: Model parameter shift



Mean-normalized weight update of each layer: $\sum_{i=0}^{k} \frac{(w_i^l - w_{i-1}^l)}{w_0^l}$

- Bert model has higher variance for all layers, especially for the randomly-initialized classification layer.
- Mean-normalized update of BERT model is larger that gpt2-generative early during training, but is smaller at the end of training, where gpt2-generative achieves higher validation performance (ablation on training convergence)

Ablation: Training convergence



- BERT model converges faster than GPT2 in 1% few-shot settings, due to using a small classification head
- GPT2 converges more slowly, perhaps due to using language modeling loss, i.e. cross-entropy loss across all tokens of the input sequence

Conclusion: GPT2 language model exploits more supervision than BERT in few-shot setting

Multi-task Generation

Sentence	Task	Model Output
the sangria's - watered down.	aspect term	<|term|> sangria negative
	aspect category	<|category|> food neutral
	aspect term & category	<|term|> sangria negative <|category|> food negative
	groundtruth	<|term|> sangria negative <|category|> food negative
everyone who works there (the host, the bartender, the servers) is so helpful.	aspect term	host positive, bartender neutral, servers positive
	aspect category	<|category|> service positive
	aspect term & category	<|term|> host positive, bartender positive, servers positive <|category|> service positive
	groundtruth	<|term|> bartender positive, host positive, servers positive <|category|> service positive

Contact
- Ehsan Hosseini-Asl, ehsan.hosseiniasl@gmail.com
- Wenhao Liu, wenhao.liu@salesforce.com
- Caiming Xiong, cxiong@salesforce.com

Code:
https://github.com/salesforce/fewshot_absa

Selected References:

- [1] Xu, Hu, et al., "Bert post-training for review reading comprehension and aspect-based sentiment analysis.", *NAACL-HLT*, 2019.
- [2] Reddy, Natesh, et al. "Does bert understand sentiment? leveraging comparisons between contextual and non-contextual embeddings to improve aspect-based sentiment models." arXiv:2011.11673., 2020.
- [3] Liu, Jian, et al., "Solving aspect category sentiment analysis as a text generation task.", EMNLP, 2021.