The Benefits of Label-Description Training for Zero-Shot Text Classification

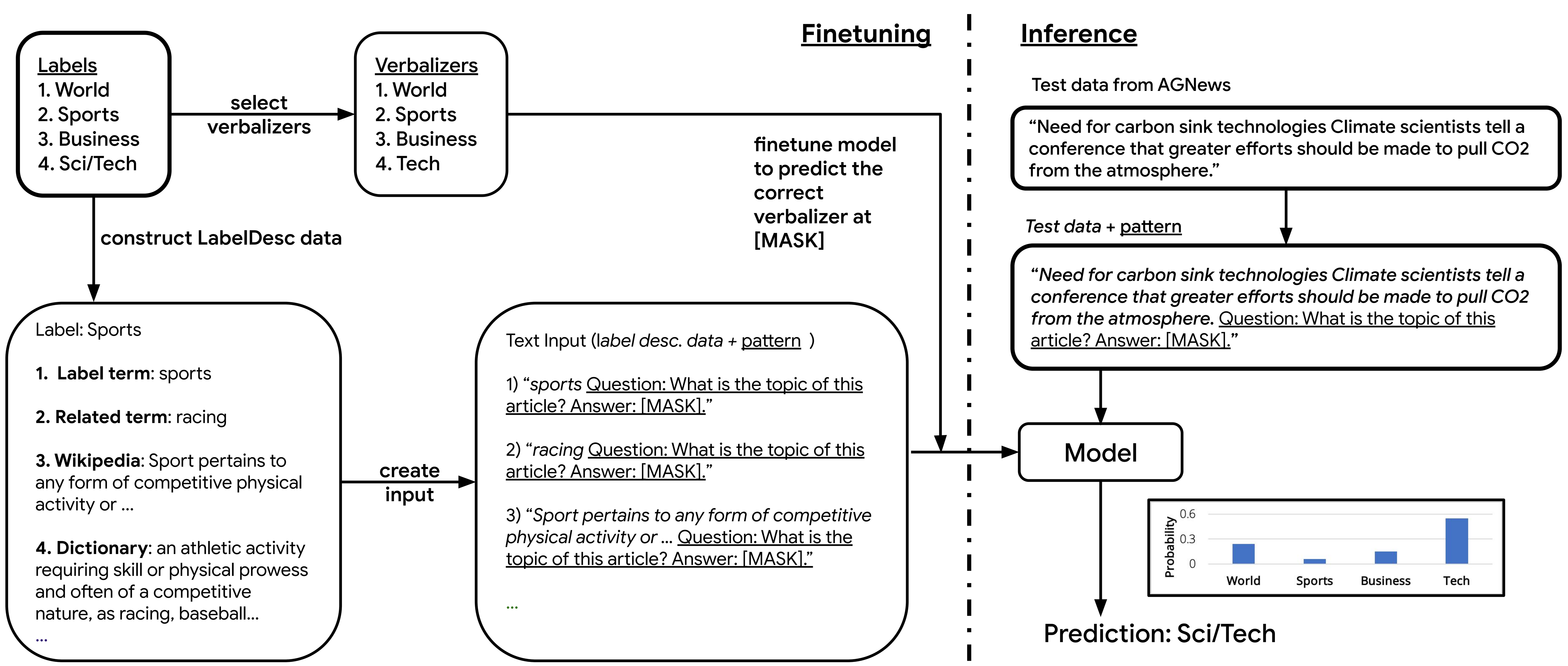
Lingyu Gao[1] Debanjan Ghosh [2†] Kevin Gimpel [1†]

[1]Toyota Technological Institute at Chicago [2]Educational Testing Service † Co-senior authors

Motivation and Contributions

- Pattern-verbalizer approach for zero-shot text classification
  - Choose words (**verbalizers**) for labels
  - Append a **pattern** to the text with a [MASK]
  - Choose most probable verbalizer at [MASK] using masked language modeling (MLM) head
  - Example: *Overpriced, salty and overrated! The restaurant is [MASK].*
- Effective but sensitive to choice of patterns/verbalizers!
- Solution: train on LabelDesc data, which has ***descriptions of labels***, rather than annotated texts
  - Topic: terms related to label, a definition, & a sentence from Wikipedia
  - Sentiment: related terms and hand-crafted templates
- Results
  - 17-19% accuracy gains across 9 topic/sentiment datasets
  - more robust to pattern/verbalizer choices
  - robust across domains

Overview



Evaluation and Results

test acc (%)	AGNews	Yahoo	DBPedia	Yelp-2	SST-2	Amz-2	IMDB
LabelDescTraining	84.6±0.3	59.9±0.3	82.4±1.2	84.8±0.6	88.2±0.2	89.6±0.4	83.4±0.4
Chu et al. (2021a)	68.8	57.8	81.9	67.3	65.0	66.8	-
Chu et al. (2021b)	75.1	60.0	88.6	-	-	-	-
van de Kar et al. (2022)	79.2	56.1	80.4	92.0	85.6	92.0	86.7

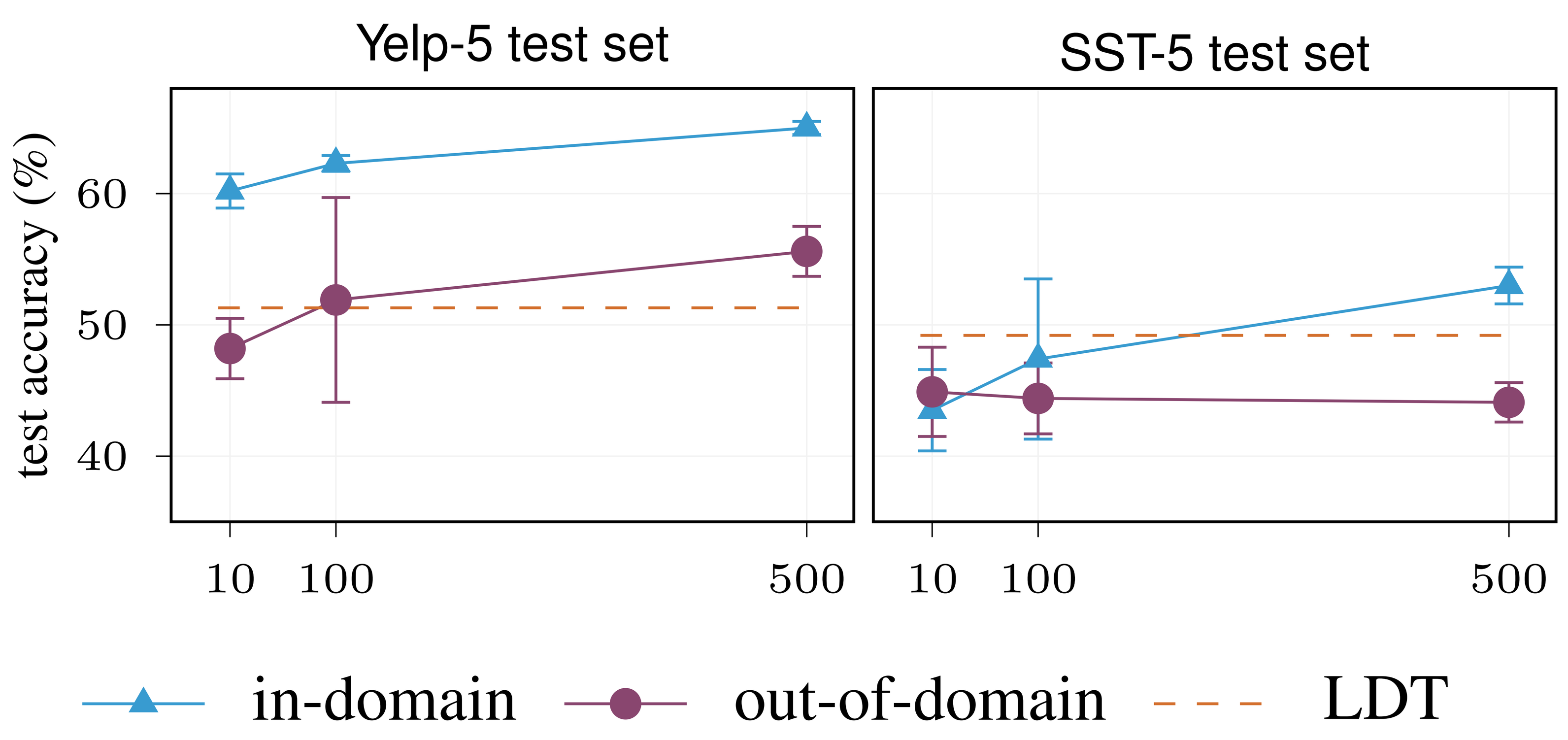
- **hyperparameters** (# of training steps, pattern for comparison, etc.) are tuned on 20 Newsgroups data
- Comparison against SOTA results (RoBERTa-base) using a single pattern with **LabelDescTraining**

	zero-shot	LDT	MLM_r	MLM_m	classifier
Avg.	58.8±11.3	77.7±2.3	73.4±6.1	65.4±6.0	71.5±2.8

- Test accuracies (%) with RoBERTa-large averaged across 9 datasets (the above + SST-5 and Yelp-5)

- **LDT**: **LabelDescTraining**
- **MLM_r**: c new verbalizers (c = # labels) are added to the vocab with random initialization of their embeddings
- **MLM_m**: Mismatched labels and verbalizers (to simulate a setting in which verbalizers are poorly chosen)
- **classifier**: Classifier without patterns

Multi-Domain Evaluation



LabelDescTraining improves over few-shot out-of-domain classification in multiple settings