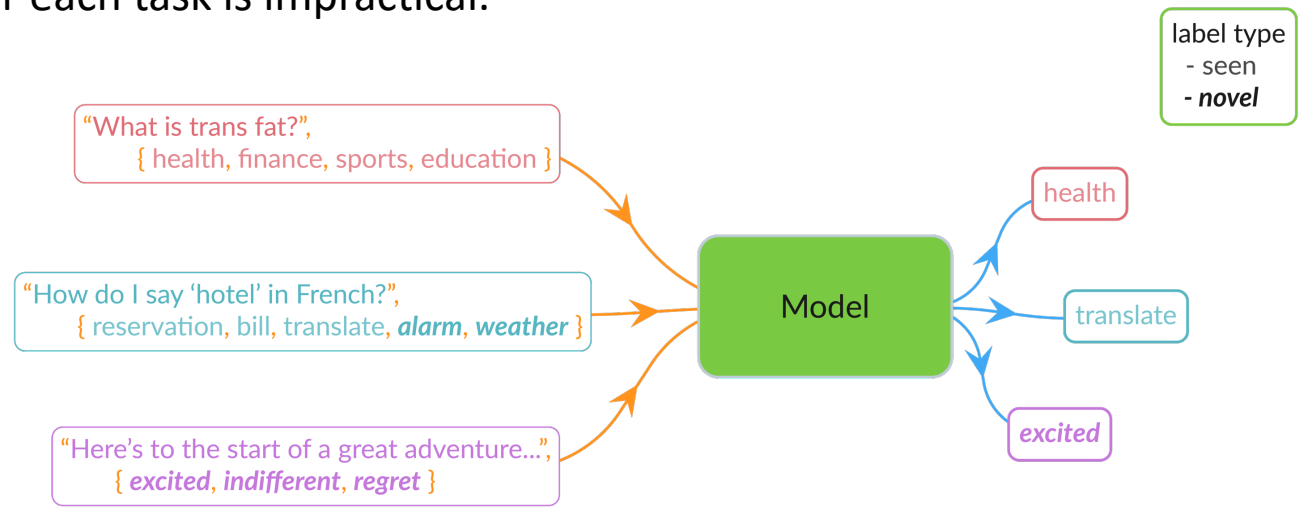
Label Agnostic Pre-training for Zero-shot Text Classification

Christopher Clarke[1], Yuzhao Heng[1], Yiping Kang[1], Krisztian Flautner[1], Lingjia Tang[1], Jason Mars[1]

[1]University of Michigan, Ann Arbor, MI.

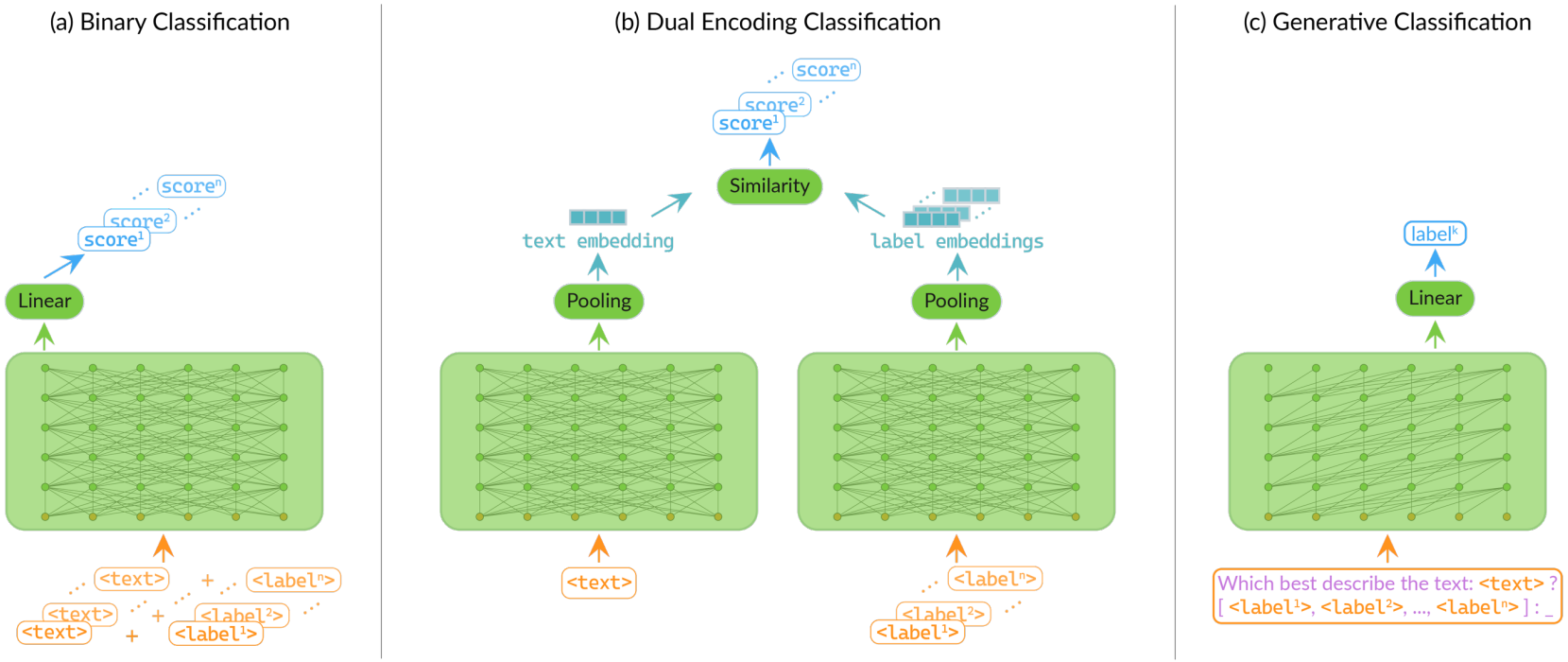
Background

- Text classification is the process of categorizing text into sets of organized groups where each set consists of similar content in a well-defined manner.
- Supervised approaches to text classification typically assume the presence of a pre-defined set of labels to which a given text can be classified. However, in real-world settings:
  - The label space is constantly evolving.
  - The type of task and domain vary greatly.
  - The use of dedicated models for each task is impractical.



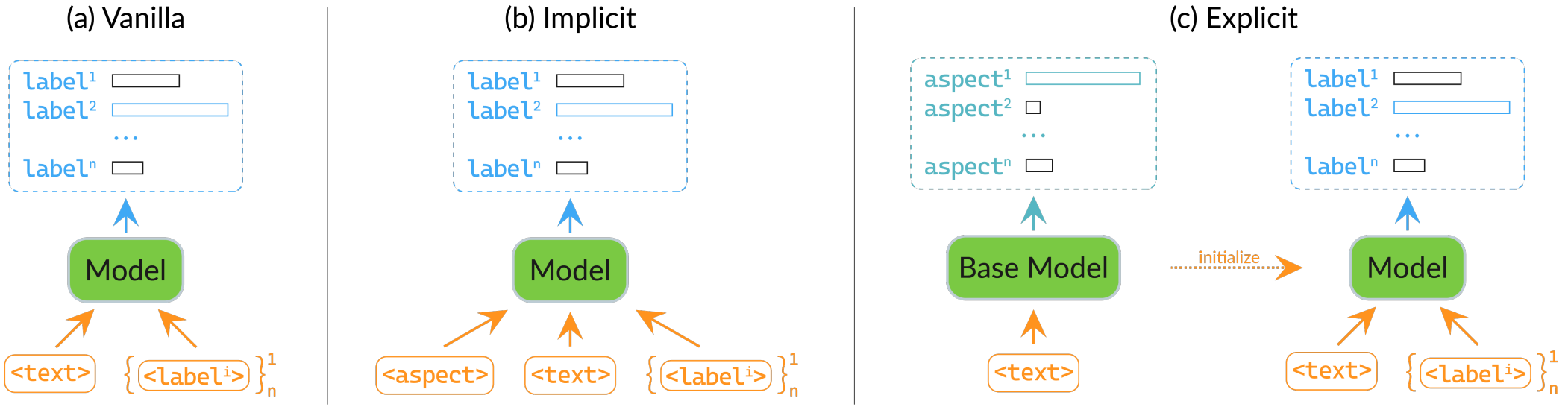
Zero-shot Learners

- Zero-shot Learners are models capable of predicting unseen classes. When applied to text classification, these models aim to associate a piece of text with a given label without the need for having been trained on that text label pair.



Problem

- Zero-shot models greatly underperform compared to their supervised counterparts on unseen data.
- We theorize that this poor generalization of these zero-shot models is due to their lack of aspect-level understanding during their training process.
- To alleviate this, we introduce two new simple yet effective pre-training strategies, **Implicit and Explicit pre-training** which specifically inject aspect-level understanding into the model.



Universal Text Classification Dataset (UTCD)

- In order to test the zero-shot generalization of these NLP models we introduce v1 of the Universal Text Classification Dataset (UTCD) consisting of **~6M/800K** train/test examples.
- UTCD is a large-scale compilation of 18 classification datasets spanning 3 main categories of Sentiment, Intent/Dialogue, and Topic classification. UTCD focuses on the task of zero-shot text classification where the candidate labels are descriptive of the text being classified.
- In the construction of UTCD we enforce the following principles:
  - **Textual labels:** In UTCD, we mandate the use of textual labels. While numerical label values are often used in classification tasks, descriptive textual labels such as those present in the datasets across UTCD enable the development of techniques that can leverage the class name which is instrumental in providing zero-shot support.
  - **Diverse domains and Sequence lengths:** In addition to broad coverage of aspects, UTCD compiles diverse data across several domains such as Banking, Finance, Legal, etc. each comprising varied length sequences (long and short).

Results

Model	Training Strategy	Sentiment			Intent			Topic			Average
		Amazon Polarity	Fin. Phrase Bank	Yelp	Banking 77	SNIPS	NLU Eval	Multi EURLEX	Patent	Consumer Finance	
BERT Seq-CLS*	individual	96.0	97.2	84.8	88.6	99.0	88.9	94.8	64.1	82.6	88.4
	full	93.1	24.9	79.0	84.7	97.3	87.4	81.4	50.2	76.9	75.0
Binary BERT	vanilla	**80.7**	**68.9**	58.5	**51.4**	82.9	71.6	28.7	13.6	22.3	53.2
	implicit (ours)	80.1	66.0	**59.8**	51.3	82.5	**73.1**	30.3	15.2	23.4	53.5
	explicit (ours)	76.1	66.7	56.0	49.8	**83.8**	69.6	**44.5**	**19.5**	**30.2**	**55.1**
Bi-Encoder	vanilla	69.9	**71.7**	46.5	9.4	70.4	**71.1**	33.5	**11.7**	18.4	44.7
	implicit (ours)	**79.6**	64.0	**56.8**	**21.1**	**72.5**	61.9	**35.4**	9.6	11.3	**45.8**
	explicit (ours)	71.5	63.6	52.1	9.7	71.9	70.0	27.4	9.3	**27.0**	44.7
GPT-2†	vanilla	88.3	71.1	70.9	**22.8**	52.2	61.7	22.3	23.5	12.6	47.3
	implicit (ours)	89.3	61.4	**71.9**	16.5	33.7	**63.1**	18.6	**25.8**	12.2	43.6
	explicit (ours)	**89.7**	**75.9**	71.5	22.4	**54.1**	60.7	**23.5**	21.6	**13.9**	**48.2**
BART‡	Zero-shot	91.0	40.2	75.2	42.2	61.4	40.1	19.8	8.9	24.6	44.8
GPT-3‡	Zero-shot	54.4	52.8	77.0	23.7	13.9	37.9	-	-	-	43.3

Highlights: By conducting Implicit and explicit pre-training, we are able to outperform baselines on generalizing to unseen data on 6, 6, and 8 of the 9 datasets in out-of-domain UTCD across Binary BERT, Bi-encoder, and GPT-2.

Conclusion

HuggingFace Dataset

GitHub Code

Hugginface Models

Highlights: We release with our paper all our data, models, and code for training and evaluating zero-shot text classification!