Template-based Approach to Zero-shot Intent Recognition

Dmitry Lamanov[1], Pavel Burnyshev[1], Ekaterna Artemova[1,2], Valentin Malykh[1], Andrey Bout[1], Irina Piontkovskaya[1]

[1]Huawei Noah's Ark lab, [2]HSE University

Generalized Zero-Shot paradigm for intent recognition

Task: recognize user intent by utterance

Zero-shot scenarios:

- **Generalized zero-shot learning** [Xian et al., 2018]
  - train with *seen* intents
  - infer with both *seen* and *unseen*
- **Dataless classification**
  - train with seen intent labels, no annotated utterances are available
  - infer with both *seen* and *unseen*

Please set up an alarm tomorrow in the morning	*Add Alarm*
I am so damn bored?	*Find Events*
OK. Can you find some classic music? I'd like to listen Beethoven	*Lookup Music*
Sounds great. Can you play this song?	*Play Song*

HUAWEI

Sentence pair modeling approach

Problem formulation:

- Let X be the set of utterances, $S = \{y_1, \ldots, y_k\}$ – set of seen intents, $U = \{y_{k+1}, \ldots, y_n\}$ – set of unseen intents.
- Training data $\{x_i, y_j\}$.
- At the test time in the GZS setup model choses an intent from both seen and unseen $y_j \in S \cup U$.

Our approach:

- Model encodes y_j and x_i concatenated by $[SEP]$ token. Representation of the $[CLS]$ token fed into classification head to get prediction $P(1|y_j\ x_i)$
- At the test time we loop over all intents $y_j \in S \cup U$ and select intent with maximum probability of positive class
- Leverages contextualized pretrained encoders like: RoBERTa, BERT, TOD-BERT
- Lexicalization of intent labels: utilize simple grammar templates to convert into natural-sounding sentences
- Negative sampling strategy: sampling hard negative utterances for a fixed intent $(y_j, {x_l}^+)$ and $(y_j, {x_m}^-)$

HUAWEI

Lexicalization templates

Most of the intent labels take such form:

$VERB + NOUN^{+}$ like *book hotel*

$NOUN^{+}$ like *flight status*

Lexicalized expression:

template + $VERB$ + *a/an* + $NOUN^{+}$

Default verb: "get "

Examples:

- the user wants to *book* a hotel
- does the user want to get a flight status

ID	Template
	declarative templates
d_1	*the user wants to* the user wants to book a hotel
d_2	*tell the user how to* tell the user how to book a hotel
	question templates
q_1	*does the user want to* does the user want to book a hotel
q_2	*how do I* how do I book a hotel

HUAWEI

Datasets

- **Schema-Guided Dialog (SGD)**
  [Rastogi et al., 2020]
  Preprocessing: keep utterances where users express an intent.
  Original splitting

- **MultiWoZ 2.2**
  (Multi-domain Wizard of Oz)
  [Budzianowski et al., 2018]

- **CLINC**
  [Larson et al., 2019]

	SGD	MultiWoz	CLINC
# Utterances	59.4K	27.5K	23.7K
# Domains	16	7	10
# Intents	46	11	112

HUAWEI

Comparison with other methods

Method	SGD				MultiWoZ				CLINC			
	Unseen		Seen		Unseen		Seen		Unseen		Seen	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
SEG (Yan et al., 2020)	0.372	0.403	0.613	0.636	0.371	0.414	0.652	0.646	-	-	-	-
RIDE+PU (Siddique et al., 2021)	0.590	0.573	0.832	0.830	0.569	0.521	0.884	0.885	**0.798**	0.573	0.908	0.912
ZSDNN + CTIR (Si et al., 2021)	0.603	0.580	0.809	0.878	0.468	0.437	0.827	0.892	0.561	0.493	0.904	0.871
CapsNet + CTIR (Si et al., 2021)	0.567	0.507	0.897	0.912	0.481	0.404	0.903	0.906	0.530	0.572	0.866	0.883
SP RoBERTa (ours)	0.698	0.732	0.917	0.925	0.606	0.686	0.903	0.919	0.661	0.742	**0.946**	**0.954**
SP RoBERTa + templates (ours)	**0.750**	**0.805**	**0.931**	**0.934**	**0.624**	**0.722**	**0.941**	**0.948**	0.692	**0.766**	0.927	0.931

Dataless classification

Synthetic utterances: paraphrased lexicalized intent labels

Example:

get alarms $\xrightarrow{\text{lexicalize}}$ *tell the user how to get alarms* $\xrightarrow{\text{paraphrase}}$ *What's the best way to get an alarm?*

Train data: intent labels +	SGD				MultiWoZ				CLINC			
	Unseen		Seen		Unseen		Seen		Unseen		Seen	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
original utterances	0.687	0.716	0.916	0.922	0.594	0.705	0.903	0.912	0.639	0.731	0.894	0.903
synthetic utterances	0.666	0.688	0.746	0.778	0.615	0.642	0.621	0.713	0.580	0.613	0.608	0.654

Conclusion

Key contributions of the paper:

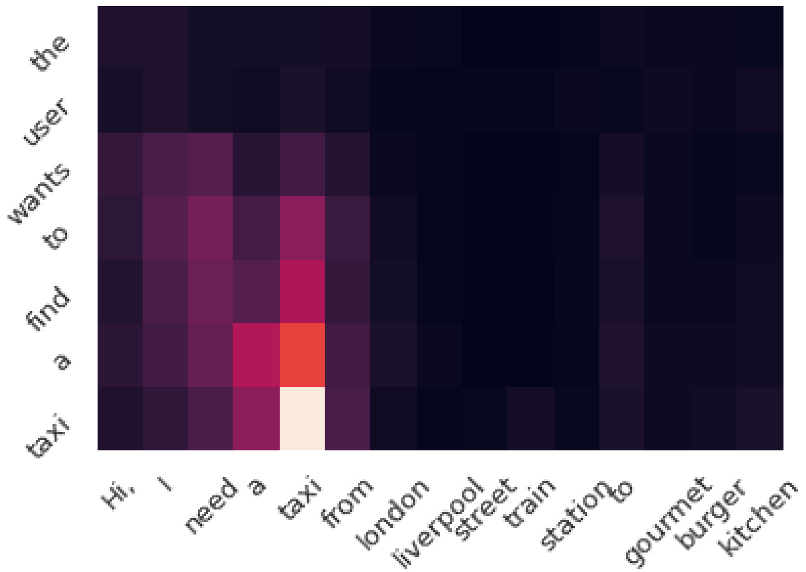
- Sentence pair modeling in GZS intent recognition established state-of-the-art results
- Lexicalization of intent labels yields significant improvement
- Training in dataless regime, task transferring experiments, ablation studies and error analysis

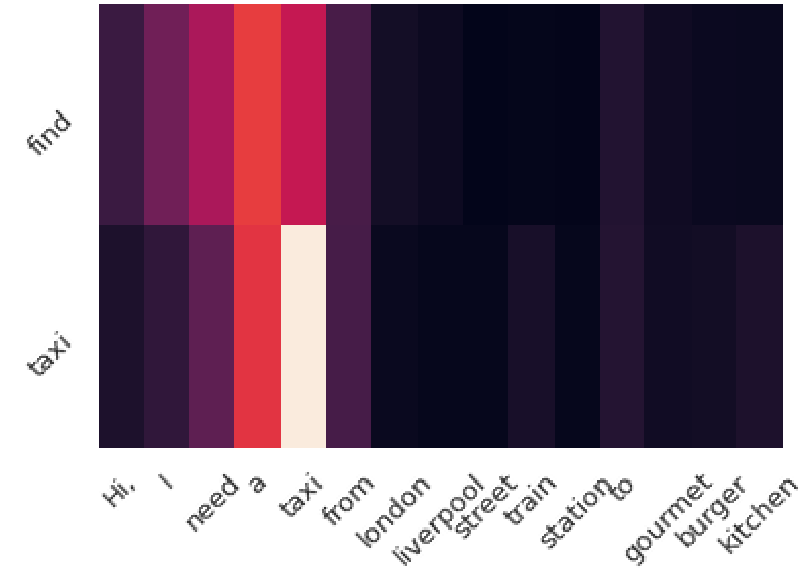
Advantages of the proposed approach

- Easy to implement. Doesn't use any extra data structure.
- Interpretability of the results

Limitations and future work

- Resource-greedy: requires to loop over all intents for given utterance
- Meaningful intent labels may not be available (num indices)





Thank you for your attention!

HUAWEI

References

- Yongqin Xian, et al. 2018. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. IEEE transactions on pattern analysis and machine intelligence, 41(9):2251–2265.
- Abhinav Rastogi, et al. 2020. Towards scalable multi-domain conversational agents: The schema-guided dialogue dataset. In The ThirtyFourth AAAI Conference on Artificial Intelligence, AAAI 2020
- Paweł Budzianowski, et al. 2018. MultiWOZ - a large-scale multi-domain Wizard-of-Oz dataset for task-oriented dialogue modelling. In Proceedings of the 2019 Conference on Empirical Methods in Natu ral Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)
- Stefan Larson et al. In Proceedings of the 2019 Conference on Empirical Methods in Natu ral Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)

HUAWEI