Generalized Zero Shot Learning for Intent Classification and Slot Filling

Korolev Alexey
scientific advisors: Bondarenko Ivan, Batura Tatyana

Big Data AI
Novosibirsk State University

19 May

Content

Introduction to DST

During the dialogue, we recognize the user intends and fill the appropriate slot.
For different services, we have different intends and slots.
So if we include new service we need to provide train dialogues and train again system.

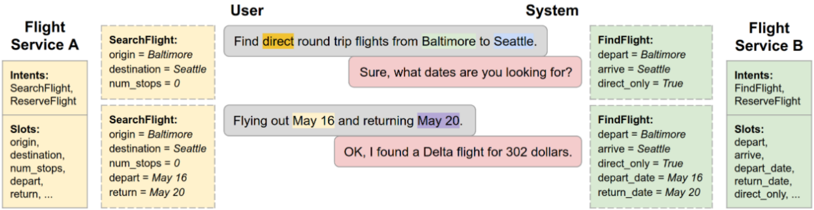


Figure: Example of shema dialogue system.

GZSL

Generalized Zero-shot learning aims to recognize objects whose class may not have been seen during training.
For classes, we have some description. So GZSL methods trying to extract knowledge between class description and class example.

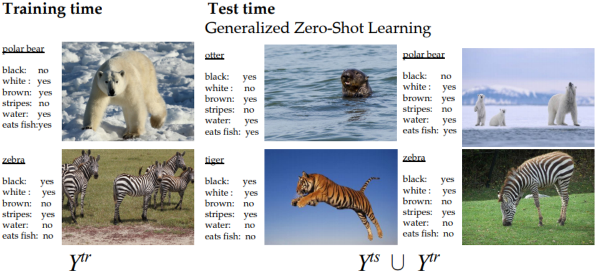


Figure: Flow visualisation[1]

[1]B. Schiele Y. Xian C. H. Lampert and Z. Akata. "Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly". In: (2018). URL:

BERT approaches

BERT is a popular architecture for many NLP tasks. There are o lot of works in DST with BERT. For example in this work with architecture similar to this work..[2]
$[h_0, \dots, h_T] = BERT([e_1, \dots, e_T]); y_i = softmax(Wh_i + b);$

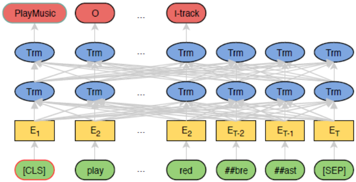


Figure: A high-level view of the observed model. The input query is "play the song little robin redbreast".

[2] W. Wang Q. Chen Z. Zhuo. "BERT for Joint Intent Classification and Slot Filling". In: (2019). URL: https://arxiv.org/pdf/1902.10909.pdf.

Our command creating a tool for Zero-shot task. Using this system you can perform zls task or evaluate your architecture on wild used datasets.

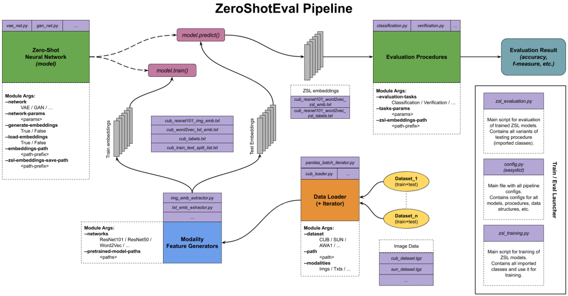


Figure: Strurcure of Zero-shot eval system[3]

[3] **https://github.com/ZSLresearch-team/ZeroShotEval**.

GZSL Nets

There's a lot of different approaches to perform GZSL task.
In ZeroShortEval inplement 2 nets: CADA-VAE and LISGAN.
CADA-VAE: based on Variation auto encoder.[4]
LISGAN: based on GAN[5]
for input to this nets, we take picture embedding(resnet-101) and class embedding. output: ZSL-embedding.
Also, there are Bayesian networks, knowledge graph etc.

---

[4] S. Sinha T. Darrell Z. Akata E. Schonfeld S. Ebrahimi. "Generalized Zero- and Few-Shot Learning via Aligned Variational Autoencoders". In: (2019). URL: https://arxiv.org/pdf/1812.01784.pdf.

[5] B. Schiele Z. Akata Y. Xian T. Lorenz. "Feature Generating Networks for Zero-Shot Learning". In: (2018). URL: https://arxiv.org/pdf/1712.00981v2.pdf.

Datasets

During literature reviews next datasets is frequent occurrence: DSTC 2, SNIPS.
Also occure: XSchema, MultiWOZ 2.1., ect
Also some special Dataset DSTC-8 (SGP-DST).[6]Special new dataset for ZSL approach in DST.

Metric ↓ Dataset →	DSTC2	WOZ2.0	MultiWOZ	SGD
No. of domains	1	1	7	**16**
No. of dialogues	1,612	600	8,438	**16,142**
Total no. of turns	23,354	4,472	113,556	**329,964**
Avg. turns per dialogue	14.49	7.45	13.46	**20.44**
Avg. tokens per turn	8.54	11.24	**13.13**	9.75
Total unique tokens	986	2,142	23,689	**30,352**
No. of slots	8	4	24	**214**
No. of slot values	212	99	4,510	**14,139**

Figure: Datasets statistics.

[6] S. Sunkara R. Gupta P. Khaitan A. Rastogi X. Zang. "Schema-Guided Dialogue State Tracking Task at DSTC8". In: (2020). URL: https://arxiv.org/pdf/2002.01359.pdf.

Data

We use DSTC-8 - special dataset containing many services with different intents.

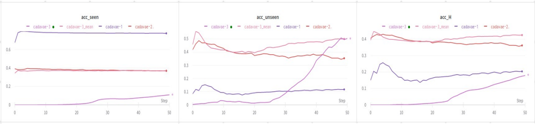
split:

1. In the first setup, we unite intent with the same name, but different thesis
2. In the second setup, we consider similar intends in different services as different.
3. In the second setup, we consider similar intends in different services as different, but add service descriptions.

Encoders:

- Roberta
- XLM Roberta
- Universal sentence encoder

Results CADAVAE

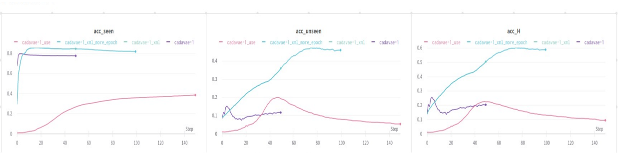


Name (4 visualized)	Unseen	Seen	H
cadavae-3	0.3087	0.1689	0.2183
cadavae-3_mean	0.558	0.3734	0.4474
cadavae-1	0.1528	0.7972	0.2565
cadavae-2.	0.4705	0.3949	0.4294

Figure: different scenarios.

1 scenario give best seen accuracy because we unite similar intents from different services.
3 scenario give the worst result, but with the summed description it gives a much better result.
2 scenario are quite balanced. Badly seen accuracy can be related to the similarity of intends in different services.

Results CADAVAE



Name (4 visualized)	Unseen	Seen	H
cadavae-1_use	0.1978	0.2626	0.2257
cadavae-1_xml_more_epocl	0.4738	0.8296	0.6032
cadavae-1_xml	0.36	0.8462	0.5051
cadavae-1	0.1528	0.7972	0.2565

Figure: different encoders.

Results Lisgan

Name (6 visualized)	H	Seen	Unseen
lisgan-3-xlm	0.0003423	0.6808	0.0001712
lisgan-3	0.01255	0.6749	0.006333
lisgan-1-xlm	0.0003423	0.6844	0.0001712
lisgan-1	0.0006844	0.6772	0.0003423
lisgan-2_use	0.0006844	0.6772	0.0003423
lisgan-2_xlm	0.0003423	0.6844	0.0001712

Figure: different encoders.

Conclusion

- During this work, we test opportunity use ZSL architectures(LISGAN, CADAVAE) from computer vision tasks on NLP It shows the universality of checked architectures.
- During work, we try different splittings and different encoders. Provide experiments results and their interpretations.
- Develop ZeroShotEval framework.
- Thesis is accepted in ISSC 2021.

Thank you for attention.