Coach: A Coarse-to-Fine Approach for Cross-domain Slot Filling

ACL 2020

Zihan Liu

Genta Indra Winata

Peng Xu

Pascale Fung

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

CAiRE
Center for Artificial Intelligence Research

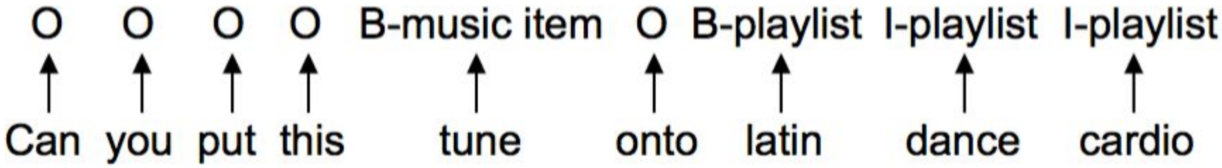
Outline

- Background
- Coach Framework
- Experiment Setup
- Results and Analysis
- Conclusion

Background

Cross-domain Slot Filling

- Slot Filling Example



- Cross-domain Slot Filling
  - Use zero or only a few labeled training samples in target domains.

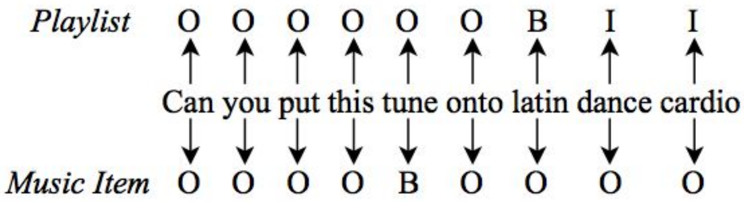
Challenge & Previous work

Challenge: Handle unseen slot types

- Some slot types in the target domain do not exist in the source domain

Previous work: Concept Tagger (CT) (Bapna, et al 2017)

- CT conducts slot filling for each slot type



Disadvantages

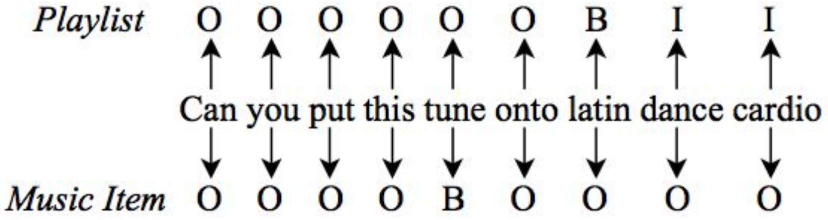
- Difficult to capture the whole slot entity
- An entity could have multiple predictions

- Bapna, A., Tür, G., Hakkani-Tür, D., & Heck, L. (2017). Towards Zero-Shot Frame Semantic Parsing for Domain Scaling. *Proc. Interspeech 2017*, 2476-2480.

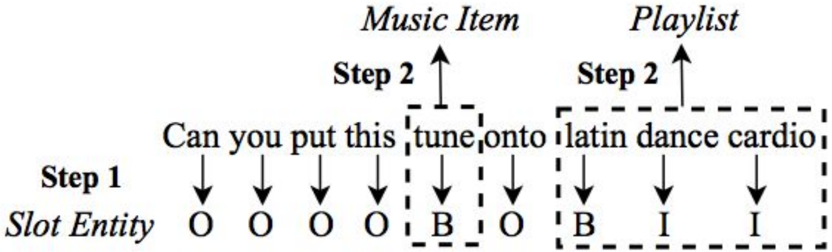
Coach Framework

Concept Tagger vs. Coarse-to-Fine

Concept Tagger



Coarse-to-Fine
(Coach)



Coarse-to-Fine

Step One (Coarse Step)

$$[h_1, h_2, ..., h_n] = \text{BiLSTM}(\mathbf{E}(\mathbf{w})), \quad (1)$$

$$[p_1, p_2, ..., p_n] = \text{CRF}([h_1, h_2, ..., h_n]), \quad (2)$$

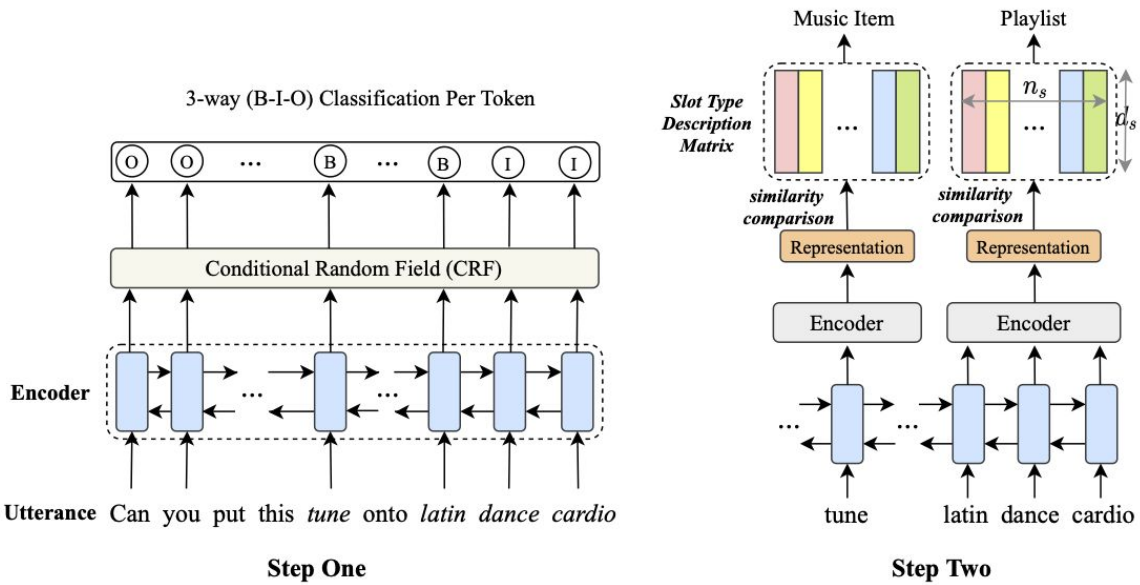
Step Two (Fine Step)

$$r_k = \text{BiLSTM}([h_i, h_{i+1}, ...h_j]), \quad (3)$$

$$s_k = M_{desc} \cdot r_k, \quad (4)$$

We sum the embeddings of the N slot description tokens to obtain slot representation

$$r^{desc} = \sum_{i=1}^{N} \mathbf{E}(t_i), \quad (5)$$

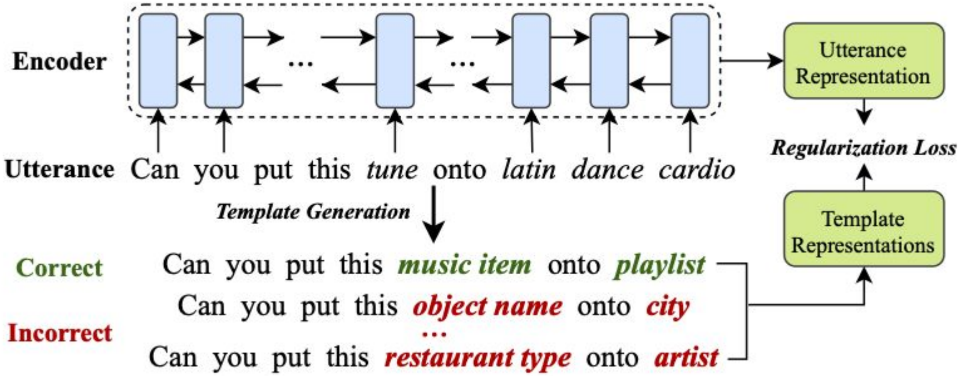


Template Regularization

$$e_t = h_t w_a, \;\; \alpha_t = \frac{exp(e_t)}{\sum_{j=1}^{n} exp(e_j)}, \;\; R = \sum_{t=1}^{n} \alpha_t h_t, \tag{6}$$

$$L^r = \mathrm{MSE}(R^u, R^r), \tag{7}$$

$$L^w = -\beta \times \mathrm{MSE}(R^u, R^w), \tag{8}$$



Experiment Setup

Datasets

SNIPS (Slot Filling)

- SNIPS contains 39 slot types across seven domains
- Each time, we choose one domain as the target domain and the other six domains as the source domains.

CoNLL-2003 & SciTech News (NER)

- Source domain: CoNLL 2003 English NER dataset
- Target domain: CBS SciTech News NER dataset
- These two datasets have the same four types of entities, namely, PER (person), LOC (location), ORG (organization), and MISC (miscellaneous).

Baselines

Concept Tagger (CT)

- Bapna et al. (2017) proposed a slot filling framework that utilizes slot descriptions to cope with the unseen slot types in the target domain.

Robust Zero-shot Tagger (RZT)

- Based on CT, Shah et al. (2019) leveraged example values of slots to improve robustness of cross-domain adaptation.

BiLSTM-CRF

- BiLSTM-CRF uses the same label set for the source and target domains and casts it as an entity classification task for each token. This baseline is only for the cross-domain NER.

- Bapna, A., Tür, G., Hakkani-Tür, D., & Heck, L. (2017). Towards Zero-Shot Frame Semantic Parsing for Domain Scaling. *Proc. Interspeech 2017*, 2476-2480.
- Shah, D., Gupta, R., Fayazi, A., & Hakkani-Tur, D. (2019, July). Robust Zero-Shot Cross-Domain Slot Filling with Example Values. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5484-5490).

Results & Analysis

Zero/Few-shot Results on SNIPS

Training Setting	Zero-shot				Few-shot on 20 (1%) samples				Few-shot on 50 (2.5%) samples			
Domain ↓ Model →	CT	RZT	Coach	+TR	CT	RZT	Coach	+TR	CT	RZT	Coach	+TR
AddToPlaylist	38.82	42.77	45.23	**50.90**	58.36	**63.18**	58.29	62.76	68.69	**74.89**	71.63	74.68
BookRestaurant	27.54	30.68	33.45	**34.01**	45.65	50.54	61.08	**65.97**	54.22	54.49	72.19	**74.82**
GetWeather	46.45	50.28	47.93	**50.47**	54.22	58.86	67.61	**67.89**	63.23	58.87	**81.55**	79.64
PlayMusic	32.86	**33.12**	28.89	32.01	46.35	47.20	53.82	**54.04**	54.32	59.20	62.41	**66.38**
RateBook	14.54	16.43	**25.67**	22.06	64.37	63.33	**74.87**	74.68	76.45	76.87	**86.88**	84.62
SearchCreativeWork	39.79	44.45	43.91	**46.65**	57.83	**63.39**	60.32	57.19	66.38	**67.81**	65.38	64.56
FindScreeningEvent	13.83	12.25	**25.64**	25.63	48.59	49.18	66.18	**67.38**	70.67	74.58	78.10	**83.85**
Average F1	30.55	32.85	35.82	**37.39**	53.62	56.53	63.17	**64.27**	64.85	66.67	74.02	**75.51**

Table 1: Slot F1-scores based on standard BIO structure for SNIPS. Scores in each row represents the performance of the leftmost target domain, and TR denotes template regularization.

- Coach outperforms CT and RZT, especially in the few-shot scenario.
- Template Regularization (TR) further improves the cross-domain performance.

SNIPS Results on Seen and Unseen Slots

- Both Coach and TR improve the performance on seen and unseen slots.
- **For unseen slot:** Our models are better able to capture the unseen slots since they explicitly learn the general pattern of slot entities.
- **For seen slots:** Baseline models might fail to recognize these seen slots in the target domain, while our approaches can adapt to the seen slot types more quickly in comparison.

Target Samples‡	0 samples		20 samples		50 samples	
	unseen	seen	unseen	seen	unseen	seen
CT	27.1	44.18	50.13	61.21	62.05	69.64
RZT	28.28	47.15	52.56	63.26	63.96	73.10
Coach	32.89	50.78	61.96	73.78	74.65	76.95
Coach+TR	**34.09**	**51.93**	**64.16**	**73.85**	**76.49**	**80.16**

Table 2: Averaged F1-scores for seen and unseen slots over all target domains. ‡ represent the number of training samples utilized for the target domain.

Zero/Few-shot Results on SciTech News

- Coach framework is also suitable for the case where there are no unseen labels in the target domain.
- Template regularization loses its effectiveness in this task, since the text in NER is relatively more open, which makes it hard to capture the templates for each label type.

Target Samples	0	50
CT (Bapna et al. (2017))	61.43	65.85
RZT (Shah et al. (2019))	61.94	65.21
BiLSTM-CRF	61.77	66.57
Coach	64.08	**68.35**
Coach + TR	**64.54**	67.45

Table 3: F1-scores on the NER target domain (CBS SciTech News).

Conclusion

- We introduce a new cross-domain slot filling framework to handle the unseen slot type issue.
- Moreover, template regularization is proposed to improve the adaptation robustness further.
- Experiments show that our model significantly outperforms existing cross-domain slot filling approaches, and it also achieves better performance for the cross-domain NER task, where there is no unseen label type in the target domain.

Thank you

All questions are welcome

Check our code

https://github.com/zliucr/coach