An Application Survey of Zero-Shot Learning on Natural Language Processing

Bowen Xu
Institute of Applied Electronics
Beijing, China
xubowen@pku.edu.cn

ABSTRACT

Zero-shot learning plays an important role facing the difficulty of laking annotated data. The technique or idea of zero-shot learning has been widely used in many traditional fields such as computer vision and natural language processing (NLP). While there are not any literatures to review the applications of zero-shot learning on NLP. In this paper, we make a comprehensive survey on that, aiming to give the researchers on zero-shot learning a full view of it. We first introduce the works of zero-shot learning on neural machine learning, natural language understanding, and knowledge graph constructing in a birds' eye view, and then detailly compare the methods and their performance in each application.

1 APPLICATIONS

Zero-short Learning is a useful means when training data and respective labels are hard to achieve [1]. It has been widely combined with traditional neural methods of Computer Vision, Natural Language Processing (NLP), etc. In this paper, we make a comprehensive survey of the applications of zero-shot learning on NLP. We will introduce the application schemes and basic ideas of methods in the next few subsections. The applications in neural machine translation will be introduced more detailly, so that readers could comprehend the basic principles and methods of how to apply the zero-shot idea to some field. Then we will describe and compare the methods and their performances in each application in detail in section xx and section xx.

1.1 Neural Machine Translation

Traditionally, neural machine translation (NMT) relies on large amounts of the parallel corpus. To translate a zero-shot source word to a target language, the basic assumption is that the mapping of representations between different linguistic words is linear. Similar semantics of different linguistic words under the same context have similar representations [2]. Based on that assumption, we can infer the representations, i.e. word embeddings as the terminology of NLP, of words unseen before if the words in similar semantic or with high relevance are learned by supervised learning. With the seen words, we can learn a mapping from the source language to the target language. Using the same mapping, the target linguistic representations of unseen words can be predicted well.

A simpler problem than translation is dictionary induction, mapping words from a source language, as an entry, to equivalent words of a target language. Word embeddings of each language are learned, then given a seed dictionary a mapping is trained to connect the two linguistic vector spaces. A zero-hot word can be translated through the mapping [3].

To consider the more complex problem, neural machine translation, which aims to translate a source sentence to a target one. The state-of-the-art deep-learning-based methods rely on big data. For the lack of parallel corpus, training becomes extremely hard. Zero-shot multilingual translation model assumes that the cross-lingual sentence pairs are rare in some languages. Although traditional pivot-based translation is used to address zero-shot translation, Firat et al. proposed a multilingual neural model to gain the ability of zero-shot learning, which is deficient in the traditional one-to-one and many-to-one translation strategies [4]. It is implicated that the zero-shot translation model can learn the common, underlying structures of multiple languages. Johnson et al. propose a more general method to translate multiple languages [5]. All the language-pairs are trained to learn shared parameters, which allows the models to generalize in the corpus-missing languages. Al-Shedivat et al. give the definition of zero-shot consistency to measure the feasibility of zero-shot learning [6], and hereby take a different approach from a probabilistic perspective aiming to improve the training procedure of [5].

Zheng et al. alleviate the problem of error propagation in traditional segmentation decoding by maximizing the estimation of expectation likelihood with respect to the pivot-to-source model [7]. Chen et al. propose a teacher-student framework, in which a source-to-target translation model, as student, is trained without any parallel corpora [8]. While another pivot-to-target translation model, as teacher, guides the student to learn on a source-pivot parallel corpus.

Zero-shot learning and dual learning are combined. Through dual learning, it gains better performance than solely using zero-shot learning [9, 10]. The basic idea of dual learning is that two agents with different models translate the original sentence to another language and then translate it back; by comparing the original and the translated back sentences both the models can improve their performance without any labels of data [11]. Gu et al. propose a new translation approach that share lexical and sentence level representations of multiple languages [10]. A universal word-level representation space is shared by several languages, along with a sentence-level expert model assuming that sentences of different languages have similar structures. A novel task of NMT presents a multi-agent setting in which agent learners engage in image description games with corporations. Through the games, agents improve their own translation ability [12]. On account of the sensitivity of the hyper-parameter setting in the training stage, Guetal.capture the spurious correlations of mutual information between languages to improve the performance of the zero-shot model [13].

1.2 Natural Language Understanding

Natural language understanding (NLU) aims to parse natural sentences into the form that a machine can use directly. For example, an NLU module transfers a sentence to a series of instructions to make some queries in a database. Leveraging the similarity between word representations of natural language and instruction representations of computer language, zero-shot learning can be easily achieved,i.e.training the word embeddings of the respective language, learning the mappings based on seed pairs, generalizing the mapping to unseen pairs is feasible [14, 15]. When the training dataset for mapping an utterance to instruction is hard to build, the zero-shot NLU comes into play. Using domain-independent questions and sequential structure, it is showed that when trained in one specific domain(e.g.insurance), the system can also work in another specific domain (e.g.sightseeing) [16–19]. Ferreira, Jabaian et al. [20] and Ferreira, Masson, et al. [21] adopt an online adaptative strategy to refine the initial model progressively. Besides, an adversarial algorithm is used to learn the policy of online adaption [21].

1.3 Constructing Knowledge Graph

To construct a knowledge graph, there are two steps. First, extract the entities from the natural sentences; second, extract relations between the entities. In both of the steps, zero-shot learning can help to extract knowledge when training data is limited. When extracting entities, there may not be enough pairs of the object and its entity label. Then the names of objects and entities can be mapped to embeddings so that the agent can connect any unseen words to embedding s of entities label [22, 23]. Some more further problems like fine-grained entity typing [24, 25], in which the several probable entities type should be recognized, and open entity classification [26, 27], in which newly defined entities can be identified flexibly, are explored. It should be emphasized that Zhang et al. adopts a memory augmented approach through which the knowledge from seen types can be transferred to the unseen ones [27]. Relation extraction methods with zero-shot learning are proposed [28].To

eliminate the restriction that the relations should all be seen before, Jia et al. propose a Parasitic Neural Network to extract the unseen relation though learning the general representations of relation types [29].

In addition, some other problems related to constructing knowledge graphs are explored with zero-shot learning, such as entity linking [30], entity property recognition [31], and so on [32, 33].

1.4 Other Sub-Fields of NLP

With regard to the classification problem, the representations learned by NMT can be used for the downstream tasks. Eriguchi et al. propose a multilingual classifier which reuse the encoder learned by multilingual NMT to construct a task-specific class i fier [34]. Other classification works refers to text classification [35–37], documents' topic classification [38],intent classification [12, 39, 40], etc. With regard to generation problem with zero-shot learning, the relative works are domain natural language generation [41], headline generation [42], abstract generation [43], question generation [44], dialog generation [45], etc. Besides, other attempts with zero-shot learning, are made, such as cross-lingual slot filling [46], cross-lingual document retrieval [47], and cross-lingual word alignment [48].

REFERENCES

[1] Wei Wang, Vincent W Zheng, Han Yu, and Chunyan Miao. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–37, 2019.

[2] Tomas Mikolov, Quoc V Le, and Ilya Sutskever. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*, 2013.

[3] Ndapandula Nakashole and Raphael Flauger. Knowledge distillation for bilingual dictionary induction. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2497–2506, 2017.

[4] Orhan Firat, Baskaran Sankaran, Yaser Al-Onaizan, Fatos T Yarman Vural, and Kyunghyun Cho. Zero-resource translation with multi-lingual neural machine translation. *arXiv preprint arXiv:1606.04164*, 2016.

[5] Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351, 2017.

[6] Maruan Al-Shedivat and Ankur P Parikh. Consistency by agreement in zero-shot neural machine translation. *arXiv preprint arXiv:1904.02338*, 2019.

[7] Hao Zheng, Yong Cheng, and Yang Liu. Maximum expected likelihood estimation for zero-resource neural machine translation. In *IJCAI*, pages 4251–4257, 2017.

[8] Yun Chen, Yang Liu, Yong Cheng, and Victor OK Li. A teacher-student framework for zero-resource neural machine translation. *arXiv preprint arXiv:1705.00753*, 2017.

[9] Lierni Sestorain, Massimiliano Ciaramita, Christian Buck, and Thomas Hofmann. Zero-shot dual machine translation. *arXiv preprint arXiv:1805.10338*, 2018.

[10] Jiatao Gu, Hany Hassan, Jacob Devlin, and Victor OK Li. Universal neural machine translation for extremely low resource languages. *arXiv preprint arXiv:1802.05368*, 2018.

[11] Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, and Wei-Ying Ma. Dual learning for machine translation. In *Advances in neural information processing systems*, pages 820–828, 2016.

[12] Yun Chen, Yang Liu, and Victor OK Li. Zero-resource neural machine translation with multi-agent communication game. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[13] Jiatao Gu, Yong Wang, Kyunghyun Cho, and Victor OK Li. Improved zero-shot neural machine translation via ignoring spurious correlations. *arXiv preprint arXiv:1906.01181*, 2019.

[14] Emmanuel Ferreira, Bassam Jabaian, and Fabrice Lefèvre. Zero-shot semantic parser for spoken language understanding. In *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[15] Majid Yazdani and James Henderson. A model of zero-shot learning of spoken language understanding. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 244–249, 2015.

[16] Kugatsu Sadamitsu, Yukinori Homma, Ryuichiro Higashinaka, and Yoshihiro Matsuo. Zero-shot learning for natural language understanding using domain-independent sequential structure and question types. In *INTERSPEECH*, pages

3306–3310, 2017.

[17] Ankur Bapna, Gokhan Tur, Dilek Hakkani-Tur, and Larry Heck. Towards zero-shot frame semantic parsing for domain scaling. *arXiv preprint arXiv:1707.02363*, 2017.

[18] Jonathan Herzig and Jonathan Berant. Decoupling structure and lexicon for zero-shot semantic parsing. *arXiv preprint arXiv:1804.07918*, 2018.

[19] Sungjin Lee and Rahul Jha. Zero-shot adaptive transfer for conversational language understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6642–6649, 2019.

[20] Emmanuel Ferreira, Bassam Jabaian, and Fabrice Lefevre. Online adaptive zero-shot learning spoken language understanding using word-embedding. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5321–5325. IEEE, 2015.

[21] Emmanuel Ferreira, Alexandre Reiffers Masson, Bassam Jabaian, and Fabrice Lefèvre. Adversarial bandit for online interactive active learning of zero-shot spoken language understanding. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6155–6159. IEEE, 2016.

[22] Panupong Pasupat and Percy Liang. Zero-shot entity extraction from web pages. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 391–401, 2014.

[23] Marco Guerini, Simone Magnolini, Vevake Balaraman, and Bernardo Magnini. Toward zero-shot entity recognition in task-oriented conversational agents. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 317–326, 2018.

[24] Yukun Ma, Erik Cambria, and Sa Gao. Label embedding for zero-shot fine-grained named entity typing. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 171–180, 2016.

[25] Rasha Obeidat, Xiaoli Fern, Hamed Shahbazi, and Prasad Tadepalli. Description-based zero-shot fine-grained entity typing. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 807–814, 2019.

[26] Ben Zhou, Daniel Khashabi, Chen-Tse Tsai, and Dan Roth. Zero-shot open entity typing as type-compatible grounding. *arXiv preprint arXiv:1907.03228*, 2019.

[27] Tao Zhang, Congying Xia, Chun-Ta Lu, and Philip Yu. Mzet: Memory augmented zero-shot fine-grained named entity typing. *arXiv preprint arXiv:2004.01267*, 2020.

[28] Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. Zero-shot relation extraction via reading comprehension. *arXiv preprint arXiv:1706.04115*, 2017.

[29] Shengbin Jia, Yang Xiang, et al. Parasitic neural network for zero-shot relation extraction. *arXiv preprint arXiv:2004.00499*, 2020.

[30] Shruti Rijhwani, Jiateng Xie, Graham Neubig, and Jaime Carbonell. Zero-shot neural transfer for cross-lingual entity linking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6924–6931, 2019.

[31] Wiradee Imrattanatrai, Makoto P Kato, and Masatoshi Yoshikawa. Identifying entity properties from text with zero-shot learning. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 195–204, 2019.

[32] Marek Rei and Anders Søgaard. Zero-shot sequence labeling: Transferring knowledge from sentences to tokens. *arXiv preprint arXiv:1805.02214*, 2018.

[33] Ziad Al-Halah, Makarand Tapaswi, and Rainer Stiefelhagen. Recovering the missing link: Predicting class-attribute associations for unsupervised zero-shot learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5975–5984, 2016.

[34] Akiko Eriguchi, Melvin Johnson, Orhan Firat, Hideto Kazawa, and Wolfgang Macherey. Zero-shot cross-lingual classification using multilingual neural machine translation. *arXiv preprint arXiv:1809.04686*, 2018.

[35] Yann N Dauphin, Gokhan Tur, Dilek Hakkani-Tur, and Larry Heck. Zero-shot learning for semantic utterance classification. *arXiv preprint arXiv:1401.0509*, 2013.

[36] Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. Integrating semantic knowledge to tackle zero-shot text classification. *arXiv preprint arXiv:1903.12626*, 2019.

[37] Yaoyao Dai and Benjamin Radford. Multilingual word embedding for zero-shot text classification. In *Annual Meeting of the American Political Science Association, Boston*, 2018.

[38] Yangqiu Song, Shyam Upadhyay, Haoruo Peng, Stephen Mayhew, and Dan Roth. Toward any-language zero-shot topic classification of textual documents. *Artificial Intelligence*, 274:133–150, 2019.

[39] Congying Xia, Chenwei Zhang, Xiaohui Yan, Yi Chang, and Philip S Yu. Zero-shot user intent detection via capsule neural networks. *arXiv preprint arXiv:1809.00385*, 2018.

[40] Kyle Williams. Zero shot intent classification using long-short term memory networks. *Proc. Interspeech 2019*, pages 844–848, 2019.

[41] Nina Dethlefs. Domain transfer for deep natural language generation from abstract meaning representations. *IEEE computational intelligence magazine*, 12(3):18–28, 2017.

[42] Shi-qi Shen, Yun Chen, Cheng Yang, Zhi-yuan Liu, Mao-song Sun, et al. Zero-shot cross-lingual neural headline generation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(12):2319–2327, 2018.

[43] Xiangyu Duan, Mingming Yin, Min Zhang, Boxing Chen, and Weihua Luo. Zero-shot cross-lingual abstractive sentence summarization through teaching generation and attention. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3162–3172, 2019.

[44] Hady Elsahar, Christophe Gravier, and Frederique Laforest. Zero-shot question generation from knowledge graphs for unseen predicates and entity types. *arXiv preprint arXiv:1802.06842*, 2018.

[45] Tiancheng Zhao and Maxine Eskenazi. Zero-shot dialog generation with cross-domain latent actions. *arXiv preprint arXiv:1805.04803*, 2018.

[46] Darsh J Shah, Raghav Gupta, Amir A Fayazi, and Dilek Hakkani-Tur. Robust zero-shot cross-domain slot filling with example values. *arXiv preprint arXiv:1906.06870*, 2019.

[47] Ruka Funaki and Hideki Nakayama. Image-mediated learning for zero-shot cross-lingual document retrieval. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 585–590, 2015.

[48] Tal Schuster, Ori Ram, Regina Barzilay, and Amir Globerson. Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing. *arXiv preprint arXiv:1902.09492*, 2019.