// 介绍各个领域面临的问题，然后介绍ZSL是如何应用到其中的。

A Novel Model of Zero-Shot Learning on Natural Language Processing

Abstract

Introduction

1. overview of zero-shot learning

2. overview of natural language processing

3. the application of Zero-Shot learning on Natural Language Processing

Zero-Shot Learning is applied to several sub-fields of Natural Language Processing, which mainly are Neural Machine Translation, Natural Dialogue System, and Name Entity Recognition. In the next sub-sections, we will introduce the relevant works of Zero-Shot Learning on NLP.

**3.1 Neural Machine Translation**

Traditionally, neural machine translation relies of large amounts of parallel corpus. To translate a zero-shot source word to a target language, the basic assumption is that the map of representations between different linguistic words ***are linear***. Similar semantics of different linguistic words under the same context have similar representations.[1] 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 map from the source language to the target language. Using the same map, 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 in 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 the source sentence to the target one. The state-of-the-art deep-learning-based methods relies on big data. For the lack of parallel corpus of the training is 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 (Wu and Wang, 2007; Utiyama and Isahara, 2007[2]), [2] 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(Zoph and Knight, 2016). It is implicated that the zero-shot translation model can learn the common, underlying structures of multiple languages. [5] proposes a more general method to translate multiple languages. All the language-pairs are trained to learn shared parameters, which allows the models to generalize in the corpus-missing languages. [13] gives the definition of zero-shot consistency to measure the feasibility of zero-shot learning, and hereby take a different approach from a probabilistic perspective aiming to improve the training procedure of [5].

[4] alleviates the problem of error propagation in traditional segmentation decoding by maximize the estimation of expectation likelihood with respect to pivot-to-source model. [7] propose a teacher-student framework, in which a source-to-target translation model, as student, is trained without any parallel corpora. While another pivot-to-target translation model, as teacher, guide the student to learn on a source-pivot parallel corpus.

zero-shot learning and dual learning are combined. Though 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.[9.1] [10] proposes a new translation approach that share lexical and sentence level representations of multiple languages. A universal word-level representation space is shared by several languages, along with a sentence-level experts 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. [11]

On account of the sensitivity of hyper-parameter setting in the training stage, [14] captures the spurious correlations of mutual information between languages to improve the performance of zero-shot model. Deservedly, adding a trick of regularization makes the system more robust [27].

In terms of evaluation

评估：[16] [17]

[1] [2013] Exploiting similarities among languages for machine translation

[2] [2016] Zero-Resource Translation with Multi-Lingual Neural Machine Translation

[3] [2017] Knowledge distillation for bilingual dictionary induction

[4] [2017] Maximum Expected Likelihood Estimation for Zero-Resource Neural Machine Translation

[5] [2017] Google’s Multilingual Neural Machine Translation System- Enabling Zero-Shot Translation

[6] [2017] Zero-Shot Translation for Indian Languages with Sparse Data

[7] [2017] A teacher-student framework for zero-resource neural machine translation

[8] [2018] Zero-shot cross-lingual classification using multilingual neural machine translation

[9] [2018] Zero-shot dual machine translation

[9.1] [2016] Dual Learning for Machine Translation

[10] [2018] Universal neural machine translation for extremely low resource languages

[11] [2018] Zero-resource neural machine translation with multi-agent communication game

[12] [2018] Improving Zero-Shot Translation of Low-Resource Languages

[13] [2019] Consistency by agreement in zero-shot neural machine translation

[14] [2019] Improved zero-shot neural machine translation via ignoring spurious correlations

[15] [2019] Cross-lingual Pre-training Based Transfer for Zero-shot Neural Machine Translation

[16] [2019] Evaluating the Supervised and Zero-shot Performance of Multi-lingual Translation Models

[17] [2019] Evaluating the cross-lingual effectiveness of massively multilingual neural machine translation

[27] [2019] Improving Zero-shot Translation with Language-Independent Constraints

**3.2 Natural Language Understanding**

Natural language understanding (NLU) aims to parse natural sentences into the form that machine can use directly. For example, an NLU module transfer a sentence to a series of instructions in order 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 respective language, learning the mappings based on a seed pairs, generalizing the mapping to unseen pairs is feasible.[18][19] When the training dataset for mapping a 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)[20][22][24][25][28]. [21][23] adopts a online adaptative strategy to refine the initial model progressively. In addition, an adversarial bandit algorithm is used to learn the policy of online adaption[23].

[18] [2015] Zero-shot semantic parser for spoken language understanding

[19] [2015] A model of zero-shot learning of spoken language understanding

[20] [2017] Zero-shot Learning for Natural Language Understanding using Domain-Independent Sequential Structure and Question Types

[21] [2015] Online adaptative zero-shot learning spoken language understanding using word-embedding

[22] [2017] Towards zero-shot frame semantic parsing for domain scaling

[23] [2017] Adversarial bandit for online interactive active learning of zero-shot spoken language understanding

[24] [2018] Decoupling structure and lexicon for zero-shot semantic parsing

[25] [2019] Zero-shot adaptive transfer for conversational language understanding

[28] [2019] Zero-Shot Semantic Parsing for Instructions

**3.3 Constructing and Exploiting Knowledge Graph**

To construct a knowledge graph, there are two steps. Frist, 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 no be enough pairs of the object and its entity label. Then the names of objects and entities can be mapping to an embedding, so that the agent can connect any unseen words to embeddings of entities label [29][31]. Some more further problems like fine-grained entity typing [30][33], in which the several probable entities type should be recognized, and open entity classification [32][35], in which newly defined entities can be identified flexibly, are explored. It should be emphasized that [35] adopts a memory augmented approach through which the knowledge from seen types can be transferred to the unseen ones. Relation extraction methods with zero-shot learning are proposed [36]. To eliminate the restriction that the relations should all be seen before, [38] proposes a Parasitic Neural Network to extract the unseen relation though learning the general representations of relation types.

In addition, some other problems related to constructing knowledge graph are explored with zero-shot learning, such as entity linking [39], entity property recognition [34], and so on [37][42].

实体抽取

[29] [2014] Zero-shot entity extraction from web pages

[30] [2016] Label embedding for zero-shot fine-grained named entity typing

[31] [2018] Toward zero-shot entity recognition in task-oriented conversational agents

[32] [2019] Zero-shot open entity typing as type-compatible grounding

[33] [2019] Description-Based Zero-shot Fine-Grained Entity Typing

[35] [2020] MZET: Memory Augmented Zero-Shot Fine-grained Named Entity Typing

实体属性抽取

[34] [2019] Identifying Entity Properties from Text with Zero-shot Learning

关系抽取

[36] [2017] Zero-shot relation extraction via reading comprehension

[38] [2020] Parasitic Neural Network for Zero-Shot Relation Extraction

实体连接

[39] [2019] Zero-shot neural transfer for cross-lingual entity linking

类别属性关联

[42] [2016] Recovering the missing link: Predicting class-attribute associations for unsupervised zero-shot learning

[37] [2018] Zero-shot sequence labeling: Transferring knowledge from sentences to tokens

**3.4 Other sub-fields of NLP**

With regard to the classification problem, the representations learned by NMT can be used for the downstream tasks. [8] propose a multilingual classifier which reuse the encoder learned by multilingual NMT to construct a task-specific classifier. Other classification works refers to text classification [52][53][54], documents’ topic classification [48], intent classification [49][50][51], etc. With regard to generation problem with zero-shot learning, the relative works are domain natural language generation [43], headline generation [44], abstract generation [45], question generation [46], dialog generation [47], etc. Besides, other attempts with zero-shot learning, are made, such as cross-lingual slot filling [56], cross-lingual document retrieval [57], and cross-lingual word alignment [58].

With regard to

自然语言生成

[43] [2017] Domain Transfer for Deep Natural Language Generation from Abstract Meaning Representations

标题生成

[44] [2018] Zero-shot Cross-Lingual Neural Headline Generation

摘要生成

[45] [2019] Zero-Shot Cross-Lingual Abstractive Sentence Summarization through Teaching Generation and Attention

问题生成

[46] [2018] Zero-shot question generation from knowledge graphs for unseen predicates and entity types

对话生成

[47] [2018] Zero-shot dialog generation with cross-domain latent actions

文档话题分类

[48] [2019] Toward any-language zero-shot topic classification of textual documents

意图分类/理解

[49] [2018] Zero-shot User Intent Detection via Capsule Neural Networks

[50] [2019] Zero Shot Intent Classification Using Long-Short Term Memory Networks

[51] [2016] Zero-shot learning of intent embeddings for expansion by convolutional deep structured semantic models

[????] Zero-shot learning of user intent understanding by Convolutional Neural Networks

分类

语义分类

[52] [2013] Zero-shot learning for semantic utterance classification

[53] [2019] Integrating semantic knowledge to tackle zero-shot text classification

[54] [2018] Multilingual word embedding for zero-shot text classification

关系分类

// [55] [2018] Zero-shot Relation Classification as Textual Entailment

槽填充

[56] [2019] Robust Zero-Shot Cross-Domain Slot Filling with Example Values

文档检索

[57] [2015] Image-mediated learning for zero-shot cross-lingual document retrieval

跨语言对齐

[26] [2019] Cross-lingual alignment of contextual word embeddings, with applications to zero-shot dependency parsing

其他

[58] [2017] Zero-shot activity recognition with verb attribute induction

[59] [2017] WOAH: Preliminaries to Zero-shot Ontology Learning for Conversational Agents

[60] [2019] Zero-Shot Cross-Lingual Opinion Target Extraction

Related Work