

# Learn to Adapt for Generalized Zero-Shot Text Classification

Yiwen Zhang, Caixia Yuan\*, Xiaojie Wang, Ziwei Bai, Yongbin Liu

Beijing University of Posts and Telecommunications

{garyzhang, yuancx, xjwang, bestbzw, liuyb}@bupt.edu.cn

## Abstract

Generalized zero-shot text classification aims to classify textual instances from both previously seen classes and incrementally emerging unseen classes. Most existing methods generalize poorly since the learned parameters are only optimal for seen classes rather than for both classes, and the parameters keep stationary in predicting procedures. To address these challenges, we propose a novel *Learn to Adapt* (LTA) network using a variant meta-learning framework. Specifically, LTA trains an adaptive classifier by using both seen and virtual unseen classes to simulate a generalized zero-shot learning (GZSL) scenario in accordance with the test time, and simultaneously learns to calibrate the class prototypes and sample representations to make the learned parameters adaptive to incoming unseen classes. We claim that the proposed model is capable of representing all prototypes and samples from both classes to a more consistent distribution in a global space. Extensive experiments on five text classification datasets show that our model outperforms several competitive previous approaches by large margins. The code and the whole datasets are available at <https://github.com/Quareia/LTA>.

## 1 Introduction

Text classification plays an important role in many natural language processing (NLP) applications, such as question classification, news categorization, user intent classification and so on (Minaee et al., 2021). Although a wide variety of methods have been proved successful in supervised text classification, they often break down when applied to make predictions for incrementally emerging classes without labeled training data (Pourpanah et al., 2020). Unlike zero-shot learning (ZSL) that aims to classify unseen class instances at test time (Romera-Paredes and Torr, 2015; Wang et al., 2019), generalized zero-shot learning (GZSL), which we focus

on in this work, aims to classify text samples from both previous seen and emerging novel classes. Since there is a strong bias towards seen classes (Xian et al., 2019a), GZSL is a more challenging yet critical problem.

Previously methods mainly focus on transductive approaches for generalized zero-shot text classification. Rios and Kavuluru (2018) use a graph convolution network to enhance the unseen class label embeddings. Zhang et al. (2019) and Song et al. (2020) generate illusion feature embeddings for unseen classes based on side information, *i.e.*, class-level attributes or text description. More recently, Ye et al. (2020) use reinforced self-training methods to leverage unlabeled data during training.

With the assumption that no knowledge about unseen categories is available during the model learning phase, researchers resort to inductive approaches to handle generalized zero-shot text classification. ReCapsNet (Liu et al., 2019) uses a dimensional attention-based intent capsule network and constructs zero-shot class prototypes by similarity matrix transformation. SEG (Yan et al., 2020) exploits an outlier detection approach that can be directly applied on ReCapsNet, which discriminates the domain first, then outputs the final class label.

However, the existing methods still have two key limitations. Firstly, while the goal of these methods is to transfer beneficial knowledge for unseen classes, these models merely learn optimal parameters by minimizing the loss of instances from seen classes, regardless of explicitly calibrating the predictions on unseen classes. Therefore, domain bias problem (Xian et al., 2019a) towards seen classes is not fairly resolved. Secondly, although some of them take into account the inter-class relationship when constructing prototypes for unseen classes (Liu et al., 2019), the models keep static no matter what different new classes emerge in future applications. As a result, these models show a large quality gap between instances from seen classes

\*Corresponding author.

and from emerging unseen classes.

To address these problems, motivated by the success of meta-learning in the few-shot learning task (Vinyals et al., 2016; Snell et al., 2017; Sung et al., 2018; Finn et al., 2017), we present a novel *Learn to Adapt* (LTA) network for generalized zero-shot text classification. Concretely, the proposed LTA learns over multiple learning episodes that mimic GZSL setting explicitly during training, making the learning setting consistent with the test environment and thereby improving generalization. The model notably extends its ability from two views: prototype adaptation and sample adaptation. In each episode, the LTA adjusts the representative prototypes of both seen classes and "fake" unseen classes, with the assumption that unseen classes will help in calibrating representation of seen ones and thereby enable the model to learn the class-sensitive representations. The updating for all prototypes is then used to generate a set of calibration parameters, called semantic components, to guide the adaptation of sample embeddings, which is designed to compensate for the shrinking features (Chen et al., 2018) that are ignored during training if they are not discriminating for seen classes, but could be critical for recognizing unseen classes. The refined sample embeddings are then classified based on similarity scores with all adapted class prototypes. The same setting can be directly applied in testing, where the LTA executes class prediction and adapts the learned model rationally in an on-the-fly manner.

In summary, our contributions include: (i) We propose a novel *Learn to Adapt* (LTA) network for generalized zero-shot text classification which is capable of adapting incrementally between seen classes and emerging unseen classes at test time. (ii) We propose a methodology for calibrating both prototypes and sample embeddings to deduce a global representation space, efficiently avoiding over-fitting on seen classes. (iii) Experimental results on five generalized zero-shot text classification datasets show that our method outperforms previous methods with a large margin.

## 2 Related Work

**Generalized Zero-Shot Learning** The challenge of zero-shot learning (ZSL) has been the focus of attention in recent years, especially in the applications of image classification (Socher et al., 2013; Xian et al., 2019a; Wang et al., 2019), intent detec-

tion (Xia et al., 2018; Liu et al., 2019; Yan et al., 2020), and question classification (Fu et al., 2018). Different from ZSL, generalized zero-shot learning (GZSL) that attempts to categorize instances from both seen and unseen classes is a more realistic condition that matches with practical applications. For example, a question classifier for a question answering system has to classify not only the questions ever asked but also new questions incrementally emerging from the users.

There are two key issues that GZSL has to address: (1) how to incrementally learn beneficial knowledge for unseen classes from seen ones, and (2) how to tackle the domain bias caused by the extremely imbalanced data of seen and unseen domains.

To alleviate the first issue, some of the earliest works on ZSL attempt to learn a matching model between instance embedding and class prototype embeddings represented by extra information including class-level attribute, text description, or their combinations (Frome et al., 2013; Nam et al., 2016; Zhu et al., 2019; Xia et al., 2018). In a similar vein, other methods (Wang et al., 2018; Rios and Kavuluru, 2018; Si et al., 2021) also investigate the semantic relationship between the side information for obtaining better prototype representation. Nevertheless, these models are trained using data from seen classes and fail to incrementally adapt to emerging new classes.

The key problem of the second issue is that the model is trained with data from the seen classes and the parameters are actually optimized on the seen domain, thus they are not aware of unseen classes. Assuming the extra information about unseen classes is available, another prominent approach attempts to use generative models to generate virtual samples or features for unseen domains (Xian et al., 2018; Schönfeld et al., 2019; Zhang et al., 2019; Song et al., 2020). By using synthesized samples, the generative approaches can convert GZSL problem to the conventional supervised learning problem where biases towards seen classes are largely alleviated. Additionally, studies also extend to exploit the unlabeled data for unseen classes (Xian et al., 2019b; Rahman et al., 2019; Ye et al., 2020). However, these models assume that they have access to the extra information about the unseen classes, which is not very realistic since often neither the test data nor their label descriptions is available at the training phase (as supposed in

this work). In contrast, our model can involve all classes (seen and unseen) jointly during inference, essentially it is trained towards continuous generalization for new classes, hence it is capable to adapt to incoming new classes dynamically.

**Episode-Based Training in GZSL** Our approach is primarily based on the episodic training paradigm that has been widely used in few-shot learning (FSL) (Vinyals et al., 2016; Snell et al., 2017; Sung et al., 2018). The primitive goal of episodic training is to quickly learn a meta-task from sampled classes. A particular advantage of episodic training is that, by constructing meta-tasks, the setting of training is consistent with that of testing, which is essential for classification problems.

Studies extend to exploit episodic training in the "generalized" settings. Verma et al. (2020) constructs model-agnostic meta-tasks to train generative models on GZSL. In addition, Gidaris and Komodakis (2018) utilizes weight generators to update unseen prototypes in generalized FSL (GFSL). Subsequently, to update both seen and unseen prototypes, Ye et al. (2021) exploits attention mechanism while Shi et al. (2020) takes advantage of graph neural networks in GFSL. Yu et al. (2020) use a generative network to generate unseen prototypes in GZSL. These methods only consider the prototype adaptation while the sample embeddings are still static whatever the unseen classes are. Additionally, Bao et al. (2020) uses distributional signatures to update sample embeddings in GFSL. Considering that distributional signatures can be equal for two different tasks, our method uses a novel semantic update extractor to update samples following the prototype adaptation rather than statistical information.

A compelling property of our method is that it tackles knowledge transferring and domain bias simultaneously in an episodic training framework by adapting both prototypes and sample embeddings, and draws a fast adaptation to the novel classes without the cost of dramatic damage in discriminating the seen classes.

### 3 Methodology

#### 3.1 Problem Definition

Formally, let  $\mathcal{Y}^s = \{y_1^s, \dots, y_{C^s}^s\}$  and  $\mathcal{Y}^u = \{y_1^u, \dots, y_{C^u}^u\}$  denote  $C^s$  seen classes and  $C^u$  unseen classes respectively, and  $\mathcal{Y} = \mathcal{Y}^s \cup \mathcal{Y}^u$  denote the global label space with  $\mathcal{Y}^s \cap \mathcal{Y}^u = \emptyset$ .

Suppose we have a collection of training samples  $\mathcal{D}^s = \{(x_j^s, y_j^s, a_j^s)\}_{j=1}^M$ , that consists of  $M$  samples from  $C^s$  seen classes, where  $x_j^s \in \mathcal{X}^s$  represents  $j$ -th text utterance,  $y_j^s$  and  $a_j^s$  are its one-hot class label and corresponding class-level textual description, respectively. At the test time, provided with a class description set  $\mathcal{A}^u = \{a_j^u\}_{j=1}^{C^u}$  for unseen classes, the GZSL task is to classify the test instance into either a seen or an unseen class.

#### 3.2 Overview

**Encoder** An textual input  $x$  with  $T$  words is encoded by a BERT (Devlin et al., 2019) (or any other textual encoder) into a sequence of hidden vectors  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T] \in \mathbb{R}^{T \times d_h}$ , where  $d_h$  is the dimension of the hidden vectors. The text embedding  $f(x) \in \mathbb{R}^{d_h}$  is then obtained by averaging over the  $T$  hidden vectors.

**Training** In the training stage, we apply an episodic learning paradigm, which trains the model by simulating multiple generalized zero-shot text classification tasks on seen classes. Following the principle that train and test conditions must match (Vinyals et al., 2016) and recent studies on "generalized" setting (Gidaris and Komodakis, 2018; Shi et al., 2020; Ye et al., 2021; Bao et al., 2020; Verma et al., 2020; Yu et al., 2020), the  $i$ -th episode involves an  $N^s$ -way  $K$ -shot learning task for seen classes, denoted as  $\mathcal{D}_i^s = \{(x_j^s, y_j^s, a_j^s)\}_{j=1}^{N^s \times K}$  with  $K$  labelled instances for each of the  $N^s$  classes, which are randomly sampled from the seen data  $\mathcal{D}^s$ , and a  $N^u$ -way  $K$ -shot learning task for "fake" unseen classes, denoted as  $\mathcal{D}_i^u = \{(x_j^u, y_j^u, a_j^u)\}_{j=1}^{N^u \times K}$  which is also from  $\mathcal{D}^s$ , with  $N^s + N^u \leq C^s$ . More precisely, let  $\mathcal{Y}_i^s$  and  $\mathcal{Y}_i^u$  denote the sampled seen class space and sampled "fake" unseen class space respectively, with  $\mathcal{Y}_i^s \subset \mathcal{Y}^s$ ,  $\mathcal{Y}_i^u \subset \mathcal{Y}^u$ , and  $\mathcal{Y}_i^s \cap \mathcal{Y}_i^u = \emptyset$ . For a new query instance  $x$ , the generalized zero-shot learning model performs:

$$\hat{y} = \arg \max_{y \in \{\mathcal{Y}_i^s \cup \mathcal{Y}_i^u\}} p(y|x, \mathcal{D}_i^s, \mathcal{D}_i^u) \quad (1)$$

The model has to maintain a globally consistent joint class prototype space as well as dynamic adaptation to unseen classes with zero labeled instances. In this end, we design a *Learn to Adapt* (LTA) network which first introduces a pre-trained and learnable look-up table  $\mathbf{S}$  to store embeddings of the seen prototypes, and obtain the "fake" seen classes  $\mathbf{S}_i$  from  $\mathbf{S}$ . The "fake" unseen prototypes

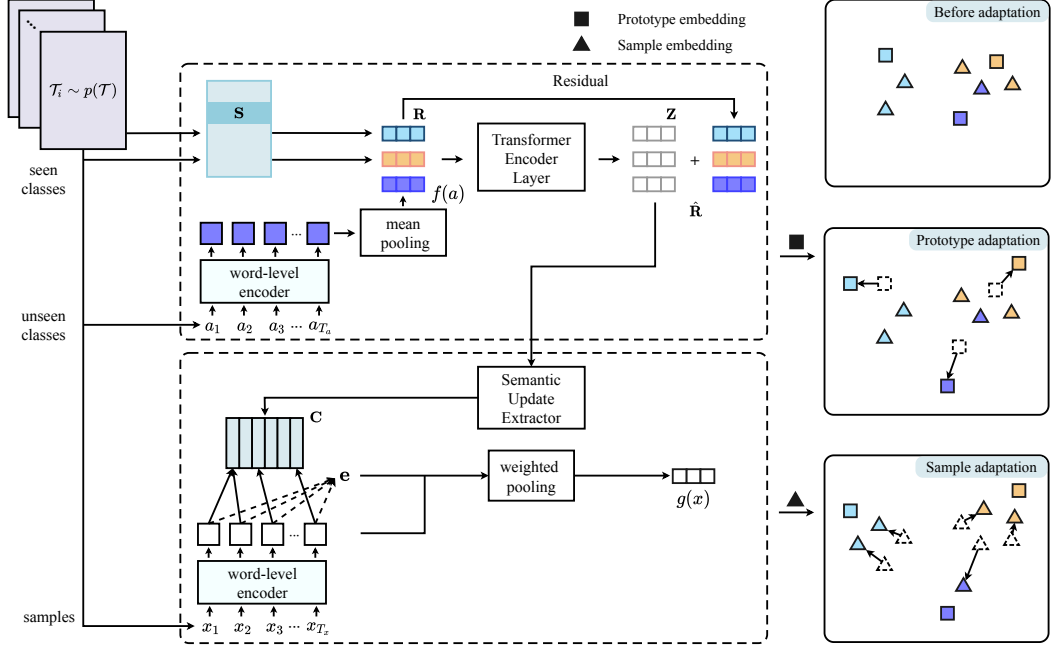


Figure 1: Illustration of the proposed LTA framework. The right part demonstrates the prototype adaptation and sample adaptation, in which  $\blacksquare$  and  $\blacktriangle$  respectively denote prototypes and samples, dotted border and solid border represent before and after adaptation, respectively.

are encoded into a matrix  $\mathbf{U}_i$  with a BERT encoder using "fake" unseen class descriptions. Then the  $\mathbf{S}_i$  and the  $\mathbf{U}_i$  are concatenated and fed into a transformer encoder layer to explicitly calibrate the seen prototype space and unseen prototype space. Meanwhile, a matrix of semantic components  $\mathbf{C}$  is generated conditioned on the updating of the prototypes. With the belief that the instance feature space should be also calibrated according to the prototypes in an on-the-fly manner,  $\mathbf{C}$  is further used for updating the feature embedding output by the same encoder.

### 3.3 Prototype adaptation

The proposed LTA network first introduces a learnable look-up table  $\mathbf{S} \in \mathbb{R}^{C^s \times d_h}$  from which to extract the "fake" seen prototypes  $\mathbf{S}_i \in \mathbb{R}^{(C^s - N^u) \times d_h}$  on demand. Following [Gidaris and Komodakis \(2018\)](#); [Ye et al. \(2021\)](#); [Shi et al. \(2020\)](#), the  $\mathbf{S}$  is firstly initialized by the prototypes trained using a supervised metric learning classifier on seen classes. The detail of the supervised metric learning classifier will be described in the experiment section. We claim that this initialization step will reduce the variance caused by the sampling episode sequences. The "fake" unseen prototypes  $\mathbf{U}_i$  is produced by the BERT encoder  $f(\cdot)$  using their corresponding class descriptions:

$$\mathbf{U}_i = [f(a_y)]_{y \in \mathcal{Y}_i^u} \in \mathbb{R}^{N^u \times d_h}.$$

Then the joint prototype matrix  $\mathbf{R}$  is obtained by concatenating  $\mathbf{S}_i$  and  $\mathbf{U}_i$ ,  $\mathbf{R} = [\mathbf{S}_i, \mathbf{U}_i] \in \mathbb{R}^{C^s \times d_h}$ , with  $\mathbf{r}_k$  as the  $k$ -th prototype. Then  $\mathbf{R}$  is fed into a single Transformer encoder layer ([Vaswani et al., 2017](#)) to explicitly model the updates for both seen prototypes and novel prototypes:

$$\begin{aligned} \mathbf{Z} &= \text{TransformerEncoder}(\mathbf{R}) \\ &= \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}_o \\ \text{where } \text{head}_i &= \text{Softmax}\left(\frac{\mathbf{R} \mathbf{W}_q \mathbf{W}_k \mathbf{R}}{d_h}\right) \mathbf{R} \mathbf{W}_v \end{aligned} \quad (2)$$

$$\hat{\mathbf{R}} = \mathbf{R} + \mathbf{Z} \quad (3)$$

where  $\mathbf{Z} \in \mathbb{R}^{C^s \times d_h}$  highlights the adjustment after mutual reflections,  $\mathbf{W}_o, \mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v \in \mathbb{R}^{d_h \times d_h}$  are trainable parameters, and the updated prototypes  $\hat{\mathbf{R}} \in \mathbb{R}^{C^s \times d_h}$  is regarded as the calibrated representative prototypes of both seen and unseen categories, with  $\hat{\mathbf{r}}_k$  as the adjusted  $k$ -th prototype. The self-attentions used in Transformer is agile to capture the inter-class relationship of seen and unseen classes and thereby it is beneficial to derive globally discriminative prototypes. The prototype adaptations simultaneously update both seen and unseen classes, which enables the model to represent and discriminate the newly incoming categories in an on-the-fly manner.



### 3.4 Sample adaptation

As been discussed in (Chen et al., 2018), the zero-shot learning tasks are prone to produce semantics loss, where some features would be discarded during training if they are not discriminating for seen classes, but critical for recognizing unseen classes. We observe that the similar problem is exacerbated in GZSL task due to the extreme unbalance between seen and unseen classes. We tackle this problem by introducing sample adaptation following the trajectories of prototypes adaptation. In concrete, we apply a semantic update extractor via attention mechanism to capture synchronous updating of the prototypes:

$$\mathbf{F} = \mathbf{Z}\mathbf{W}_1 \quad (4)$$

$$\mathbf{A} = \text{Softmax}(\mathbf{W}_3 \text{ReLU}(\mathbf{W}_2 \mathbf{F}^T)) \quad (5)$$

$$\mathbf{C} = \mathbf{A}\mathbf{F} \quad (6)$$

where  $\mathbf{W}_1 \in \mathbb{R}^{d_h \times d_h}$ ,  $\mathbf{W}_2 \in \mathbb{R}^{d_a \times d_h}$ ,  $\mathbf{W}_3 \in \mathbb{R}^{d_r \times d_a}$  are trainable parameters,  $\mathbf{A}$  denotes the attention weight matrix and  $\mathbf{C} \in \mathbb{R}^{d_r \times d_h}$  extracts different semantic components with  $\mathbf{c}_l$  as its  $l$ -th semantic components. To offset the semantic loss mentioned above, we use these semantic components to guide the adaptation of sample embeddings. Concretely, we compare the attention score for each  $\mathbf{h}_t$  to get the most related semantic adjustment and reconstruct the contribution of each word-level feature:

$$e_t = \text{Softmax}(\alpha \max_l (\frac{\mathbf{h}_t \mathbf{c}_l}{\|\mathbf{h}_t\| \|\mathbf{c}_l\|})) \quad (7)$$

$$g(x) = \sum_{t=1}^T e_t \mathbf{h}_t \quad (8)$$

where the self-attention weight  $e_t$  is used to re-weight the  $t$ -th word of sample  $x$  to be classified, and  $\alpha$  is a learnable temperature scalar to control the differentiation of Softmax scores (Gidaris and Komodakis, 2018). In this way, the different attention weights discriminate the importance of words rather than averaging them.

One notable reason of choosing of the above feature-level calibration is that, in classification task, the encoder is trained to produce feature embedding that collapses to its ground-truth prototype, therefore the adjustment of feature embedding should cater to the adjustment of a reliable global prototype space. In addition, since this calibration is applied after the encoding, it reduces

the complicated parameter tuning for a massive encoder (e.g., BERT), which elegantly helps the GZSL task to fast adapt to the incoming test instances.

### 3.5 Loss function

With the adapted prototypes  $\hat{\mathbf{R}}$  and the adapted sample  $g(x)$ , a Softmax classifier is used:

$$p(\hat{y} = y | x) = \frac{\exp(s(g(x), \hat{\mathbf{r}}_y))}{\sum_{\hat{y}} \exp(s(g(x), \hat{\mathbf{r}}_{\hat{y}}))} \quad (9)$$

where  $s(a, b) = \frac{\gamma ab}{\|a\| \|b\|}$  is cosine similarity with a learnable temperature scalar  $\gamma$ . Finally the model is trained by minimizing the losses across  $N$  episodes:

$$\mathcal{L} = \frac{1}{N} \sum_i \mathcal{L}_i \quad (10)$$

where  $\mathcal{L}_i$  is the loss of the  $i$ -th episode:

$$\mathcal{L}_i = -\frac{1}{(N^s + N^u)K} \sum_{(x,y,a) \in \mathcal{D}_i^s \cup \mathcal{D}_i^u} \log p(\hat{y} = y | x) \quad (11)$$

The training process is summarized in Algorithm 1.

---

#### Algorithm 1: LTA training algorithm.

---

**Input:** distribution over tasks  $p(\mathcal{T})$ , class set  $\mathcal{Y}^s$

**Output:** learned model parameters

```

1 while not done do
2   Randomly sample a meta GZSL task
    $\mathcal{T}_i \sim p(\mathcal{T})$  with seen meta-test  $\mathcal{D}_i^s$  and
   unseen meta-test  $\mathcal{D}_i^u$ .
3   Get adapted prototypes  $\hat{\mathbf{R}}$  by Eq 2~3.
4   Get semantic components  $\mathbf{C}$  by Eq 4~6.
5   for all  $\mathcal{D}_i^s \cup \mathcal{D}_i^u$  do
6     Get adapted sample embeddings by
     Eq 7~8.
7   end
8   Update model by Eq 9~11.
9 end

```

---

## 4 Experiments

### 4.1 Datasets

**Intent Classification Datasets.** We collect four intent classification datasets. (1) **SNIPS-SLU** (Coucke et al., 2018), a widely used benchmark for English GZSL intent detection with 5 seen intents and 2 unseen intents. (2) **SMP-18** (Zhang et al.,

2017), a Chinese dialogue corpus for user intent detection with 24 seen intents and 6 unseen intents. (3) **ATIS** (Hemphill et al., 1990), an English airline travel domain dataset, from which we extract 17 intents with at least 5 samples, and split them into 12 seen intents and 5 unseen intents. (4) **CLINC** (Larson et al., 2019) is a recently published intent detection dataset includes 22,500 in-scope queries covering 150 intent classes from 10 domains. We randomly split them into 120 seen intents and 30 unseen intents.

**Question Classification Dataset.** In order to draw a comprehensive analysis of the proposed method, we construct a question classification task from the Quora Question Pairs dataset<sup>1</sup>, which is aimed to identify duplicate questions. We collect questions with at least 5 duplicate samples into classes. In each class, we choose the question with minimum words as the label description, called the standard question, which is widely used in real-world question-answering systems (Sakata et al., 2019). Table 1 summarizes all datasets statistics. It is worth to note that intents in ATIS are highly unbalanced with *flight* accounts for about 87% of training data.

Table 1: Dataset statistics. "FS" indicates "few-shot", "BAL" indicates "balance", "IBAL" indicates "imbalance". The "avg #samples" indicates the average number of samples per class.

Dataset	#classes		#samples		sent len	type
	seen	unseen	total	avg		
SNIPS	5	2	13802	1384	9.10	BAL
SMP	24	6	2460	60	4.83	FS
ATIS	12	5	4972	245	11.44	IBAL
Clinic	120	30	22500	105	8.23	BAL
Quora	1360	340	17394	7	10.46	FS

**Dataset Settings.** Following (Siddique et al., 2021), we randomly sample seen and unseen classes for 10 runs instead of manual selection used in (Yan et al., 2020), which leads to more fair results because every class could be unseen class. We randomly take 70% samples of each seen class as the training set and the remaining 30% as the seen test, and take all the samples of unseen classes as the unseen test. All the textual labels of the same class are regarded as the description for this class.

<sup>1</sup> [www.kaggle.com/c/quora-question-pairs](http://www.kaggle.com/c/quora-question-pairs)

## 4.2 Baseline Methods

To validate the benefits of the proposed LTA, we compare against with other approaches in three aspects:

**Supervised Learning Methods.** To show the performances on seen classes with supervised learning instead of GZSL setting, we use (1) **BiLSTM** (Schuster and Paliwal, 1997) and (2) **BERT** (Devlin et al., 2019) as the encoder with a linear Softmax classifier, which only requires samples and one-hot label.

**Metric Learning Methods.** Metric-based embedding methods are commonly used as baselines for GZSL. Thus we introduce three different metric learning methods: (1) **EucSoftmax**: We adapt (Snell et al., 2017) that uses squared Euclidean distance as the metric and Softmax classifies; (2) **Zero-shot DNN**: We adapt (Kumar et al., 2017) that uses squared Euclidean distance and triplet loss to maintain a margin for different classes. We choose the label embedding (prototype) as the anchor and the closest sample as negative sample in each triplet tuple; (3) **CoST**: We adapt (Gidaris and Komodakis, 2018) which uses cosine distance as the metric with a learnable temperature scalar.

**SOTA Methods.** We also compare our model with two recent state-of-the-art (SOTA) methods: (1) **ReCapsNet** (Liu et al., 2019) uses a dimensional attention-based intent capsule network and a matrix transformation method for GZSL. (2) **SEG** (Yan et al., 2020) is an outlier detection approach that can be directly applied on ReCapsNet. SEG acts as a domain discriminator which first determines whether a test sample belongs to seen classes or unseen classes and then classifies in their own domain. **RIDE** (Siddique et al., 2021) is not considered because they use outer knowledge that is not available in our settings, and they limit the intent labels to only two components "Action" and "Object".

## 4.3 Experimental Setup

**Evaluation Metrics.** We basically use accuracy (Acc) to estimate the performances on seen and unseen test sets. Besides, we adopt Macro-F1 (F1) rather than Micro-F1 to better evaluate the performances on imbalanced and few-shot datasets, because Macro-F1 gives the average weight of F1 scores for each class. For overall assessments, we adopt the widely used Harmonic Mean (HM) of

Table 2: Results (in %) on four intent benchmarks. The Top1 results of GZSL methods are highlighted in bold and underline for Top2 results, the same below.

Model	SNIPS-NLU						SMP-18					
	Seen		Unseen		HM		Seen		Unseen		HM	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Bi-LSTM	98.23	98.23	0.00	0.00	0.00	0.00	93.65	93.43	0.00	0.00	0.00	0.00
BERT	98.91	98.91	0.00	0.00	0.00	0.00	95.28	94.87	0.00	0.00	0.00	0.00
EucSoftmax	81.09	65.50	45.89	58.21	58.61	61.64	89.84	87.85	76.65	<u>77.51</u>	82.72	<u>82.36</u>
Zero-shot DNN	81.09	65.28	45.91	58.53	58.63	61.72	<b>90.97</b>	87.67	75.38	<u>77.32</u>	82.44	82.17
CosT	<b>91.68</b>	<b>75.76</b>	47.73	62.84	62.77	68.70	<u>90.65</u>	<u>88.41</u>	72.59	73.89	80.62	80.50
ReCapsNet + SEG	<b>96.26</b>	67.70	11.57	18.45	20.66	29.00	76.32	74.92	20.56	15.09	32.39	25.10
	92.11	73.08	50.29	62.33	65.06	67.28	67.10	67.39	36.65	32.84	47.70	44.16
LTA (Ours)	74.05	74.11	<b>90.09</b>	<b>84.22</b>	<b>81.28</b>	<b>78.84</b>	89.84	<b>90.79</b>	79.19	75.20	84.18	82.26
w / o Init	<u>82.57</u>	<u>75.22</u>	64.36	71.63	72.34	73.87	89.03	87.23	<b>80.71</b>	<b>81.74</b>	<b>84.67</b>	<b>84.40</b>
w / o SA	67.31	70.56	84.70	77.51	75.01	73.87	84.52	81.40	75.89	74.40	79.97	77.75
w / o A	75.26	71.82	83.85	<u>80.77</u>	<u>79.33</u>	<u>76.03</u>	84.35	86.93	76.90	73.54	80.72	80.50
Model	ATIS						CLINC					
	Seen		Unseen		HM		Seen		Unseen		HM	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Bi-LSTM	93.24	79.51	0.00	0.00	0.00	0.00	92.07	92.06	0.00	0.00	0.00	0.00
BERT	97.18	93.71	0.00	0.00	0.00	0.00	97.37	97.37	0.00	0.00	0.00	0.00
EucSoftmax	67.67	16.11	7.78	5.50	13.96	8.20	<u>96.02</u>	87.07	58.02	66.00	72.33	75.08
Zero-shot DNN	63.56	23.12	8.05	12.02	14.29	15.82	95.31	86.65	58.49	65.89	72.49	74.68
CosT	<b>98.02</b>	59.55	46.04	45.21	62.66	51.40	<b>96.31</b>	87.33	62.73	70.28	75.98	77.89
ReCapsNet + SEG	86.19	23.88	12.80	4.89	22.32	8.12	88.53	69.83	4.24	3.33	8.10	6.36
	<b>93.75</b>	40.90	14.78	6.36	25.53	11.01	81.04	78.89	9.07	5.44	16.31	10.18
LTA (Ours)	<u>96.28</u>	<u>63.13</u>	66.09	<b>55.02</b>	<u>78.38</u>	<b>58.80</b>	92.22	87.57	<u>73.18</u>	<u>75.74</u>	<u>81.60</u>	<u>81.23</u>
w / o Init	89.96	47.48	<b>69.79</b>	<u>52.14</u>	<b>78.60</b>	49.70	93.07	<b>88.19</b>	<b>73.80</b>	<b>77.54</b>	<b>82.32</b>	<b>82.52</b>
w / o SA	90.20	51.74	66.23	47.24	76.38	49.38	92.46	87.30	69.27	73.26	79.20	79.67
w / o A	94.94	<b>63.25</b>	57.52	49.19	71.64	<u>55.34</u>	93.81	<u>88.12</u>	70.11	74.58	80.25	80.79

Table 3: Results (in %) on Quora question classification dataset.

Model	Seen		Unseen		HM	
	Acc	F1	Acc	F1	Acc	F1
BiLSTM	71.70	69.04	0.00	0.00	0.00	0.00
EucSoftmax	79.88	74.42	56.85	62.39	66.43	67.88
Zero-shot DNN	72.52	67.42	48.68	53.27	58.26	59.52
CosT	<b>88.50</b>	81.39	62.21	73.55	73.06	77.27
LTA (Ours)	84.69	<b>83.56</b>	<u>74.83</u>	<b>76.93</b>	<b>79.45</b>	<b>80.11</b>
w / o Init	82.11	81.99	<b>75.49</b>	76.53	78.66	79.17
w / o SA	84.95	<u>82.79</u>	73.56	76.67	78.84	<u>79.62</u>
w / o A	84.21	82.40	72.50	75.23	77.92	78.65

Acc and F1 on seen and unseen test sets rather than the overall metrics on the whole test set, because the overall metrics are disturbed by the ratio of seen and unseen test set sizes.

**Implementation Details.** We use the pretrained BERT-base encoder with  $d_h = 768$  on intent classification datasets and BiLSTM with  $d_h = 128$  hidden vector size each direction on Quora dataset. The scalars of our model is set to be  $\alpha = 10.0$ ,  $\tau = 10.0$ ,  $d_a = d_h$ , which is trained via Adam (Kingma and Ba, 2015) optimizer, with learning rates  $10^{-5}$  for BERT encoder,  $10^{-4}$  for BiLSTM encoder and  $10^{-3}$  for the other parameters. We use  $h = 4$  heads

Transformer encoder layer in prototype adaptation. During training, in order to treat seen classes and unseen classes as equal, we set  $N^{s_i} = N^{u_i}$  in every meta-test set, and we set  $K = 5$  and  $N^{s_i} = N^{u_i} = [2, 2, 2, 10, 20]$ ,  $d_r = [4, 16, 32, 64, 64]$  for SNIPS-NLU, SMP-18, ATIS, CLINC and Quora datasets, respectively.

We also conduct an ablation study to investigate the effectiveness of each proposed component. As depicted in Table 2 and Table 3, "w / o Init" refers to the model that randomly initializes  $\mathbf{R}$  rather than pretrained prototypes. "w / o SA" refers to the model that only uses prototype adaptation without "sample adaptation". "w / o A" means none of the adaptation steps is applied.

#### 4.4 Results

The results on four intent datasets and Quora dataset are given in Table 2 and Table 3, respectively. It is observed that our proposed methods achieve the overall best performances compared to baselines.

Detailed and interesting observations can also be derived from the results: (1) The metric-learning methods as the basic baselines, achieve compa-

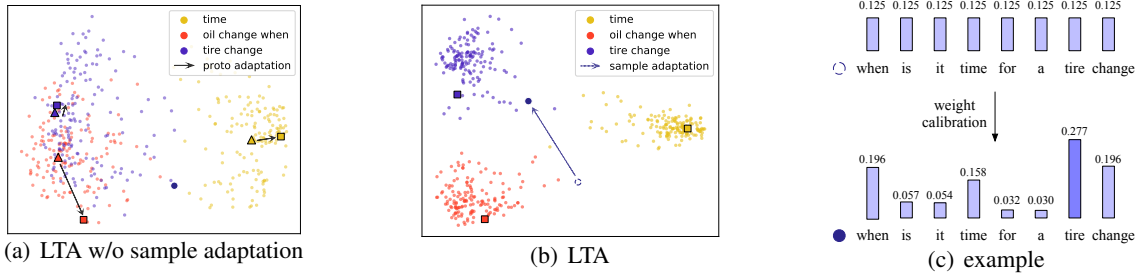


Figure 2: PCA plots of encoded unseen sample representations(·) and prototype representations(■) from (a) LTA w/o sample adaptation model and (b) full LTA model with sample adaptation (c) is an unseen example with sample-level raw attention and adapted attention. ▲ denotes the raw prototype before adaptation. ○ and ● respectively denote the example representations before and after sample adaptation.

rable results on *Seen Test* for all datasets. However, they all suffer from the domain bias problem and the performance drops with a large margin on *Unseen Test*, where the prediction is complicated due to zero-shot scenarios. (2) The performances on SNIPS-NLU and SMP-18 of ReCapsNet and SEG are worse than those in their original paper although we use the open-source codes in our experiments, that is because we randomly split the test unseen classes thus making it more challenging. Besides, these methods fail to recognize unseen samples well on datasets with a large scale of categories, yielding worse 0% Acc and F1 on Quora. The most likely reason is that ReCapsNet uses label embedding similarities to construct unseen prototypes in capsule network, which imposes a non-trivial computational and memory burden. (3) Our method shows its privilege for all datasets. In particular, with the help of continuous adapting ability, it observes a smaller gap between seen and unseen domains, which proves the adaptation on the testing phase effectively works. Although the performance on seen domain drops slightly, LTA outperforms the competitive metric-learning baselines by 9.54% HM Acc and 12.90% HM F1 averagely on the whole datasets, indicating that our model fairly balances the seen and unseen classes.

**Ablation Study.** To better understand the contribution of each component of our method, we explore three variants of LTA. We can observe that LTA with both prototype adaptation and sample adaptation outperforms those without any adaptation step in all cases. The "LTA w / o Init" has relatively stable performances. Note that "LTA w / o SA" with only prototype adaptation achieves worse performance compared to "LTA w / o A" on SNIPS-NLU, SMP-18 and CLINC. It indicates that

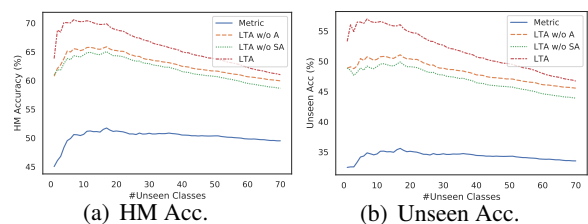


Figure 3: The performance with increasing numbers of test unseen classes on CLINC dataset.

the single prototype adaptation step may cause negative transfer, further illustrating the importance of the sample adaptation.

#### 4.5 Results on Emerging Unseen Classes

As the partition of seen and unseen classes is fixed in previous experiments, to study the robustness of the proposed adaptation method, we conduct the experiment across unseen class sets of different scales on CLINC dataset. Specifically, we select 70 classes as seen classes and 10 classes as validating unseen classes. The number of testing unseen classes is varied from 1 to 70, which are randomly sampled from the remaining 70 classes. Each experiment is repeated 50 times with different sampling sets for a more stable result. Figure 3 (a) shows the HM accuracy on all classes as the number of the unseen classes increases. We can see that our LTA model outperforms the metric learning baseline and ablation models in all cases, where the improved performance is mainly attributed to the improvements on unseen classes as shown in Figure 3 (b). These results suggest that our adaptation method is robust and effective for adapting to increasing new classes as well as improving the overall performance of all classes.



## 4.6 Visualization

To demonstrate how our adaptation method works, we further visualize the encoded representation via PCA in Figure 2. When there is no unseen class, seen classes (yellow and red) is discriminative enough. But when the new class "tire change" (purple) comes, it is ambiguous with class "oil change when" (red). We observe that the seen and unseen class prototypes are updated to be far away from each other after prototype adaptation as shown in (a), which eases the domain bias problem. However, the performance is unsatisfactory since the sample representations are still not discriminative no matter how the prototype updates. As we can see, with the sample adaptation as shown in (b), the sample representations are independently clustered by the adapted prototypes and easy to be distinguished.

To further study how the sample adaptation works, we select a representative case "when is it time for a tire change" and show its attention weights used as calibration parameters in (c). The case is still misclassified after the prototype adaptation due that the common word "time" and "change" also appear in seen classes. After the sample adaptation, however, it can be seen that the word "tire" which is a keyword for classifying, gets the highest attention while the other confusing words do not. This result suggests that calibrating using attention weights helps acquire a prototype-aware representation that guides the sample adaptation.

## 5 Conclusion

This paper proposed a novel adaptive meta-learning network for generalized zero-shot text classification. The model was trained under a consistent setting with testing. In particular, it efficiently alleviated the bias towards seen classes by utilizing both prototype adaptation and sample adaptation. Experiments on five text classification datasets validated that our model achieved compelling results on both seen classes and unseen classes meanwhile was capable of fast adapting to new classes.

## Acknowledgements

We thank the anonymous reviewers for their insightful comments. The research is supported in part by the Major Research Plan of National Natural Science Foundation of China (Grant No.92067202).

## References

- Yujia Bao, Menghua Wu, Shiyu Chang, and Regina Barzilay. 2020. [Few-shot text classification with distributional signatures](#). In *8th International Conference on Learning Representations, Addis Ababa, Ethiopia*. OpenReview.net.
- Long Chen, Hanwang Zhang, Jun Xiao, Wei Liu, and Shih-Fu Chang. 2018. [Zero-shot visual recognition using semantics-preserving adversarial embedding networks](#). In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1043–1052, Salt Lake City, USA. Computer Vision Foundation / IEEE Computer Society.
- Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, Maël Primet, and Joseph Dureau. 2018. [Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces](#). *CoRR*, abs/1805.10190.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: pre-training of deep bidirectional transformers for language understanding](#). 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 4171–4186, Minneapolis, USA. Association for Computational Linguistics.
- Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. [Model-agnostic meta-learning for fast adaptation of deep networks](#). In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 1126–1135, Sydney, Australia. PMLR.
- Andrea Frome, Gregory S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc’Aurelio Ranzato, and Tomas Mikolov. 2013. [Devise: A deep visual-semantic embedding model](#). In *Advances in Neural Information Processing Systems 26: 27th Annual Conference*, pages 2121–2129, Nevada, USA.
- Hao Fu, Caixia Yuan, Xiaojie Wang, Zhijie Sang, Shuo Hu, and Yuanyuan Shi. 2018. [Zero-shot question classification using synthetic samples](#). In *5th IEEE International Conference on Cloud Computing and Intelligence Systems*, pages 714–718, Nanjing, China. IEEE.
- Spyros Gidaris and Nikos Komodakis. 2018. [Dynamic few-shot visual learning without forgetting](#). In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 4367–4375, Salt Lake City, USA. Computer Vision Foundation / IEEE Computer Society.
- Charles T. Hemphill, John J. Godfrey, and George R. Doddington. 1990. [The ATIS spoken language systems pilot corpus](#). In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley*,

- Pennsylvania, USA, June 24-27, 1990. Morgan Kaufmann.
- Diederik P. Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *3rd International Conference on Learning Representations, San Diego, USA*.
- Anjishnu Kumar, Pavankumar Reddy Muddireddy, Markus Dreyer, and Björn Hoffmeister. 2017. [Zero-shot learning across heterogeneous overlapping domains](#). In *18th Annual Conference of the International Speech Communication Association*, pages 2914–2918, Stockholm, Sweden. ISCA.
- Stefan Larson, Anish Mahendran, Joseph J. Peper, Christopher Clarke, Andrew Lee, Parker Hill, Jonathan K. Kummerfeld, Kevin Leach, Michael A. Laurenzano, Lingjia Tang, and Jason Mars. 2019. [An evaluation dataset for intent classification and out-of-scope prediction](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 1311–1316, Hong Kong, China. Association for Computational Linguistics.
- Han Liu, Xiaotong Zhang, Lu Fan, Xuandi Fu, Qimai Li, Xiao-Ming Wu, and Albert Y. S. Lam. 2019. [Reconstructing capsule networks for zero-shot intent classification](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 4798–4808, Hong Kong, China. Association for Computational Linguistics.
- Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2021. [Deep learning-based text classification: A comprehensive review](#). *ACM Comput. Surv.*, 54(3):62:1–62:40.
- Jinseok Nam, Eneldo Loza Mencía, and Johannes Fürnkranz. 2016. [All-in text: Learning document, label, and word representations jointly](#). In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 1948–1954, Phoenix, USA. AAAI Press.
- Farhad Pourpanah, Moloud Abdar, Yuxuan Luo, Xinlei Zhou, Ran Wang, Chee Peng Lim, and Xi-Zhao Wang. 2020. [A review of generalized zero-shot learning methods](#). *CoRR*, abs/2011.08641.
- Shafin Rahman, Salman H. Khan, and Nick Barnes. 2019. [Transductive learning for zero-shot object detection](#). In *2019 IEEE/CVF International Conference on Computer Vision*, pages 6081–6090, Seoul, Korea (South). IEEE.
- Anthony Rios and Ramakanth Kavuluru. 2018. [Few-shot and zero-shot multi-label learning for structured label spaces](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3132–3142, Brussels, Belgium. Association for Computational Linguistics.
- Bernardino Romera-Paredes and Philip Torr. 2015. [An embarrassingly simple approach to zero-shot learning](#). In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2152–2161, Lille, France. PMLR.
- Wataru Sakata, Tomohide Shibata, Ribeka Tanaka, and Sadao Kurohashi. 2019. [FAQ retrieval using query-question similarity and bert-based query-answer relevance](#). In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1113–1116, Paris, France. ACM.
- Edgar Schönfeld, Sayna Ebrahimi, Samarth Sinha, Trevor Darrell, and Zeynep Akata. 2019. [Generalized zero- and few-shot learning via aligned variational autoencoders](#). In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8247–8255, Long Beach, USA. Computer Vision Foundation / IEEE.
- Mike Schuster and Kuldip K. Paliwal. 1997. [Bidirectional recurrent neural networks](#). *IEEE Trans. Signal Process.*, 45(11):2673–2681.
- Xiahan Shi, Leonard Salewski, Martin Schiegg, and Max Welling. 2020. [Relational generalized few-shot learning](#). In *31st British Machine Vision Conference 2020*, , Virtual Event, UK. BMVA Press.
- Qingyi Si, Yuanxin Liu, Peng Fu, Zheng Lin, Jiangnan Li, and Weiping Wang. 2021. [Learning class-transductive intent representations for zero-shot intent detection](#). In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pages 3922–3928, Virtual Event / Montreal, Canada. ijcai.org.
- A. B. Siddique, Fuad T. Jamour, Luxun Xu, and Vagelis Hristidis. 2021. [Generalized zero-shot intent detection via commonsense knowledge](#). In *The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1925–1929, Virtual Event, Canada. ACM.
- Jake Snell, Kevin Swersky, and Richard S. Zemel. 2017. [Prototypical networks for few-shot learning](#). In *Advances in Neural Information Processing Systems 30: Annual Conference*, pages 4077–4087, Long Beach, USA.
- Richard Socher, Milind Ganjoo, Christopher D. Manning, and Andrew Y. Ng. 2013. [Zero-shot learning through cross-modal transfer](#). In *Advances in Neural Information Processing Systems 26: 27th Annual Conference*, pages 935–943, Nevada, USA.
- Congzheng Song, Shanghang Zhang, Najmeh Sadoughi, Pengtao Xie, and Eric P. Xing. 2020. [Generalized zero-shot text classification for ICD coding](#). In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pages 4018–4024. ijcai.org.

- Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H. S. Torr, and Timothy M. Hospedales. 2018. [Learning to compare: Relation network for few-shot learning](#). In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1199–1208, Salt Lake City, USA. Computer Vision Foundation / IEEE Computer Society.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pages 5998–6008, Long Beach, USA.
- Vinay Kumar Verma, Dhanajit Brahma, and Piyush Rai. 2020. [Meta-learning for generalized zero-shot learning](#). In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 6062–6069, New York, USA. AAAI Press.
- Oriol Vinyals, Charles Blundell, Tim Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. 2016. [Matching networks for one shot learning](#). In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems*, pages 3630–3638, Barcelona, Spain.
- Wei Wang, Vincent W. Zheng, Han Yu, and Chunyan Miao. 2019. [A survey of zero-shot learning: Settings, methods, and applications](#). *ACM Trans. Intell. Syst. Technol.*, 10(2):13:1–13:37.
- Xiaolong Wang, Yufei Ye, and Abhinav Gupta. 2018. [Zero-shot recognition via semantic embeddings and knowledge graphs](#). In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, pages 6857–6866, Salt Lake City, USA. Computer Vision Foundation / IEEE Computer Society.
- Congying Xia, Chenwei Zhang, Xiaohui Yan, Yi Chang, and Philip S. Yu. 2018. [Zero-shot user intent detection via capsule neural networks](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3090–3099, Brussels, Belgium. Association for Computational Linguistics.
- Yongqin Xian, Christoph H. Lampert, Bernt Schiele, and Zeynep Akata. 2019a. [Zero-shot learning - A comprehensive evaluation of the good, the bad and the ugly](#). *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(9):2251–2265.
- Yongqin Xian, Tobias Lorenz, Bernt Schiele, and Zeynep Akata. 2018. [Feature generating networks for zero-shot learning](#). In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 5542–5551. Computer Vision Foundation / IEEE Computer Society.
- Yongqin Xian, Saurabh Sharma, Bernt Schiele, and Zeynep Akata. 2019b. [F-VAEGAN-D2: A feature generating framework for any-shot learning](#). In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 10275–10284, Long Beach, USA. Computer Vision Foundation / IEEE.
- Guangfeng Yan, Lu Fan, Qimai Li, Han Liu, Xiaotong Zhang, Xiao-Ming Wu, and Albert Y.S. Lam. 2020. [Unknown intent detection using Gaussian mixture model with an application to zero-shot intent classification](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1050–1060, Online. Association for Computational Linguistics.
- Han-Jia Ye, Hexiang Hu, and De-Chuan Zhan. 2021. [Learning adaptive classifiers synthesis for generalized few-shot learning](#). *International Journal of Computer Vision*, 129(6):1930–1953.
- Zhiqian Ye, Yuxia Geng, Jiaoyan Chen, Jingmin Chen, Xiaoxiao Xu, SuHang Zheng, Feng Wang, Jun Zhang, and Huajun Chen. 2020. [Zero-shot text classification via reinforced self-training](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3014–3024, Online. Association for Computational Linguistics.
- Yunlong Yu, Zhong Ji, Jungong Han, and Zhongfei Zhang. 2020. [Episode-Based Prototype Generating Network for Zero-Shot Learning](#). In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14032–14041, Seattle, WA, USA. IEEE.
- Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. 2019. [Integrating semantic knowledge to tackle zero-shot text classification](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long Papers)*, Minneapolis, USA. Association for Computational Linguistics.
- Weinan Zhang, Zhigang Chen, Wanxiang Che, Guoping Hu, and Ting Liu. 2017. [The first evaluation of chinese human-computer dialogue technology](#). *CoRR*, abs/1709.10217.
- Pengkai Zhu, Hanxiao Wang, and Venkatesh Saligrama. 2019. [Generalized zero-shot recognition based on visually semantic embedding](#). In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2995–3003, Long Beach, USA. Computer Vision Foundation / IEEE.