Zero-shot learning (ZSL) is a problem setup in deep learning where, at test time, a learner observes samples from classes which were not observed during training , and needs to predict the class that they belong to. The name is a play on words based on the earlier concept of one-shot learning , in which classification can be learned from only one, or a few, examples.

Zero-shot methods generally work by associating observed and non-observed classes through some form of auxiliary information, which encodes observable distinguishing properties of objects. [1] For example, given a set of images of animals to be classified, along with auxiliary textual descriptions of what animals look like, an artificial intelligence model which has been trained to recognize horses, but has never been given a zebra, can still recognize a zebra when it also knows that zebras look like striped horses. This problem is widely studied in computer vision , natural language processing , and machine perception . [2]

Background and history

The first paper on zero-shot learning in natural language processing appeared in a 2008 paper by Chang, Ratinov, Roth, and Srikumar, at the AAAI'08 , but the name given to the learning paradigm there was dataless classification . [3] The first paper on zero-shot learning in computer vision appeared at the same conference, under the name zero-data learning . [4] The term zero-shot learning itself first appeared in the literature in a 2009 paper from Palatucci, Hinton, Pomerleau, and Mitchell at NIPS'09 . [5] This terminology was repeated later in another computer vision paper [6] and the term zero-shot learning caught on, as a take-off on one-shot learning that was introduced in computer vision years earlier. [7]

In computer vision, zero-shot learning models learned parameters for seen classes along with their class representations and rely on representational similarity among class labels so that, during inference, instances can be classified into new classes.

In natural language processing, the key technical direction developed builds on the ability to "understand the labels"—represent the labels in the same semantic space as that of the documents to be classified. This supports the classification of a single example without observing any annotated data, the purest form of zero-shot classification. The original paper [3] made use of the Explicit Semantic Analysis (ESA) representation but later papers made use of other representations, including dense representations. This approach was also extended to multilingual domains, [8] [9] fine entity typing [10] and other problems. Moreover, beyond relying solely on representations, the computational approach has been extended to depend on transfer from other tasks, such as textual entailment [11] and question answering. [12]

The original paper [3] also points out that, beyond the ability to classify a single example, when a collection of examples is given, with the assumption that they come from the same distribution, it is possible to bootstrap the performance in a semi-supervised like manner (or transductive learning).

Unlike standard generalization in machine learning, where classifiers are expected to correctly classify new samples to classes they have already observed during training, in ZSL, no samples from the classes have been given during training the classifier. It can therefore be viewed as an extreme case of domain adaptation .

Prerequisite information for zero-shot classes

Naturally, some form of auxiliary information has to be given about these zero-shot classes, and this type of information can be of several types.

Learning with attributes: classes are accompanied by pre-defined structured description. For example, for bird descriptions, this could include "red head", "long beak". [6] [13] These attributes are often organized in a structured compositional way, and taking that structure into account improves learning. [14] While this approach was used mostly in computer vision, there are some examples for it also in natural language processing. [15]

Learning from textual description. As pointed out above, this has been the key direction pursued in natural language processing. Here class labels are taken to have a meaning and are often augmented with definitions or free-text natural-language description. This could include for example a wikipedia description of the class. [10] [16] [17]

Class-class similarity. Here, classes are embedded in a continuous space. A zero-shot classifier can predict that a sample corresponds to some position in that space, and the nearest embedded class is used as a predicted class, even if no such samples were observed during training. [18]

Generalized zero-shot learning

The above ZSL setup assumes that at test time, only zero-shot samples are given, namely, samples from new unseen classes. In generalized zero-shot learning, samples from both new and known classes, may appear at test time. This poses new challenges for classifiers at test time, because it is very challenging to estimate if a given sample is new or known. Some approaches to handle this include:

a gating module, which is first trained to decide if a given sample comes from a new class or from an old one, and then, at inference time, outputs either a hard decision, [19] or a soft probabilistic decision [20]

a generative module, which is trained to generate feature representation of the unseen classes--a standard classifier can then be trained on samples from all classes, seen and unseen. [21]

Domains of application

Zero shot learning has been applied to the following fields:

image classification

semantic segmentation

image generation

object detection

natural language processing

computational biology [22]

See also

One-shot learning in computer vision

Transfer learning

Fast mapping

Explanation-based learning

References