SEMI-SUPERVISED LEARNING

Mohammadreza Ahmadi Teshnizi
Sajad Sabzi

SELF-TRAINING (NO SPECIFIC RELEASE YEAR)

Self-training is a simple semi-supervised learning approach where a model is trained on the initially labeled data. Then, the model is used to make predictions on the unlabeled data, and the high-confidence predictions are added to the labeled dataset. The model is iteratively retrained on this expanded dataset.

CO-TRAINING (1998)

Co-training is a technique used when multiple views or modalities of data are available. It involves training multiple models independently on different subsets of features or modalities. The models then exchange their predictions on the unlabeled data to improve each other's performance.

SEMI-SUPERVISED VARIATIONAL AUTOENCODERS (VAES) (2013)

VAEs are generative models that can be used in semi-supervised learning. By incorporating a labeled dataset into the VAE framework, you can train the model to generate data while simultaneously encouraging the latent space to be semantically meaningful.

GRAPH-BASED METHODS (NO SPECIFIC RELEASE YEAR)

Graph-based semi-supervised learning methods build a graph representation of the data, where nodes represent data points, and edges represent similarities or connections between data points. These methods propagate label information through the graph to predict labels for unlabeled nodes.

LABEL PROPAGATION (2000S)

Label propagation methods involve propagating labels from labeled data points to their neighboring unlabeled data points in a graph or similarity matrix. This technique is especially useful when data points that are close in feature space tend to have similar labels.

CONSISTENCY REGULARIZATION (NO SPECIFIC RELEASE YEAR)

Consistency regularization is a technique that encourages a model to produce consistent predictions for slightly perturbed versions of the same input data. It helps improve the model's robustness and generalization by reducing sensitivity to small variations in the input.

PSEUDO-LABELING (2010S)

Pseudo-labeling is a semi-supervised learning method where a model trained on labeled data is used to predict labels for unlabeled data. These predicted labels are then treated as if they were ground truth labels and incorporated into the training process.

MIXMATCH (2019)

MixMatch is a semi-supervised learning algorithm that combines supervised and unsupervised loss terms. It leverages data augmentation and enforces consistency between labeled and unlabeled data to improve model performance.

TEMPORAL OR SPATIAL CONSISTENCY (NO SPECIFIC RELEASE YEAR)

In domains like video or sequence data, temporal or spatial consistency can be exploited. Models are trained to make predictions that are consistent over time or space, even when the data is partially labeled.

TRANSFER LEARNING (NO SPECIFIC RELEASE YEAR)

Pretrained models on large labeled datasets can be fine-tuned using smaller amounts of labeled data for specific tasks. This is a form of semi-supervised learning when a small labeled dataset is available for a particular task.

ACTIVE LEARNING (NO SPECIFIC RELEASE YEAR)

Active learning is a semi-supervised technique where the model selects the most informative data points from the unlabeled dataset to be labeled by an oracle (human annotator). This process aims to maximize the model's learning from a limited budget of labeled data. is especially useful when data points that are close in feature space tend to have similar labels.