Transfer Learning

Subjects: Computer Science, Artificial Intelligence Contributor: Juan Miguel Valverde

Transfer learning refers to machine learning techniques that focus on acquiring knowledge from related tasks/domains to improve generalization in the tasks/domains of interest.

transfer learning machine learning

1.Introduction

Transfer learning (or knowledge transfer) is a strategy to address the variation in the data distributions within heterogeneous datasets by reutilizing knowledge from source problems to solve target tasks. This strategy, inspired by psychology^[1], aims to exploit common features between related tasks and domains. For instance, an expert in magnetic resonance imaging (MRI) can specialize in computed tomography (CT) imaging faster than someone with no knowledge in either MRI or CT.

2. Domain in Transfer Learning

According to Pan and Yang^[2], a domain in transfer learning can be defined as $\mathcal{D} = \{\mathcal{X}, P(X)\}$ where \mathcal{X} is the feature space, and P(X) with $X = \{x_1, \ldots, x_n\} \subset \mathcal{X}$ is a marginal probability distribution. For example, in the context of MRI, \mathcal{X} could include all possible images derived from a particular MRI protocol, acquisition parameters, and scanner hardware, and P(X) depend on, for instance, subject groups, such as adolescents or elderly people. Tasks comprise a label space \mathcal{Y} and a decision function f, i.e., $\mathcal{T} = \{\mathcal{Y}, f\}$. The decision function is to be learned from the training data (X, Y). Tasks in MR brain imaging can be, for instance, survival rate prediction of cancer patients, where f is the function that predicts the survival rate, and \mathcal{Y} is the set of all possible outcomes. Given a source domain \mathcal{D}_S and task \mathcal{T}_S , and a target domain \mathcal{D}_T and task \mathcal{T}_T , transfer learning reutilizes the knowledge acquired in \mathcal{D}_S and \mathcal{T}_S to improve the generalization of f_T in \mathcal{D}_T [2]. Importantly, \mathcal{D}_S must be related to \mathcal{D}_T , and \mathcal{T}_S must be related to \mathcal{T}_T [3]; otherwise, transfer learning can worsen the accuracy on the target domain. This phenomenon, called negative transfer, has been recently formalized in Wang et al.[4].

Transfer learning approaches can be categorized based on the availability of labels in source and/or target domains during the optimization^[2]: unsupervised (unlabeled data), transductive (labels available only in the source domain), and inductive (labels available in the target domains and, optionally, in the source domains).

References

- 1. R. S. Woodworth; E. L. Thorndike; The influence of improvement in one mental function upon the efficiency of other functions. (I).. *Psychological Review* **1901**, *8*, 247-261, 10.1037/h0074898.
- 2. Sinno Jialin Pan; Qiang Yang; A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* **2009**, *22*, 1345-1359, 10.1109/tkde.2009.191.
- 3. Liang Ge; Jing Gao; Hung Ngo; Kang Li; Aidong Zhang; On handling negative transfer and imbalanced distributions in multiple source transfer learning. *Statistical Analysis and Data Mining: The ASA Data Science Journal* **2014**, 7, 254-271, 10.1002/sam.11217.
- 4. Zirui Wang; Zihang Dai; Barnabas Poczos; Jaime Carbonell; Characterizing and Avoiding Negative Transfer. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* **2019**, 11285-11294, 10.1109/cvpr.2019.01155.

Retrieved from https://encyclopedia.pub/entry/history/show/21867