

文章编号:2095-2163(2021)07-0191-06

中图分类号:TP391

文献标志码:A

基于机器学习的小样本学习综述

胡西范, 陈世平

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要:机器学习在数据密集型应用中十分广泛,但缺点是当数据集很小时往往效果欠佳。近年来,人们提出了小样本学习来解决这个问题。小样本学习指只利用少量样本来训练识别这些样本的机器学习模型。由于小样本学习的实用价值,业界提出很多针对的研究方法,但是目前国内缺少该问题的综述。本文中,对目前业界提出的小样本学习模型及算法进行了总结和探索。首先,给出了小样本学习的问题定义,并介绍了其他一些相关的机器学习问题;然后,根据先验知识,通过从3种数据增强方法和4种模型详细介绍了小样本学习方法;最后,对小样本的未来发展进行了展望。

关键词:机器学习;小样本学习;图像分类;数据增强

A survey of few-shot learning based on machine learning

HU Xifan, CHEN Shiping

(School of Optical-Electrical & Computer Engineering, University of Shanghai for Science & Technology, Shanghai 200093, China)

[Abstract] Machine learning is widely used in data-intensive applications, but the drawback is that it tends to be less effective when the data set is small. In recent years, few-shot learning has been proposed to solve this problem. Few-shot learning refers to a machine learning model that uses only a small number of samples to train the recognition of these samples. Due to the practical value of few-shot learning, the industry has put forward a lot of research methods, but there is a lack of domestic review on this issue. This paper summarizes and explores the few-shot learning models and algorithms proposed by the industry. Firstly, the paper defines the problem of few-shot learning and introduces some other related machine learning problems. Then, according to the prior knowledge, the paper introduces the few-shot learning method in detail through three data enhancement methods and four models. Finally, the future development of small sample is prospected.

[Keywords] machine learning; few-shot learning; image classification; data enhancement

0 引言

相比于机器学习,人类能够通过利用过去学到的少量样本数据来快速完成新概念的学习任务,并且做出准确的预测和评估^[1]。例如,给一个儿童看一个陌生人的几张照片,那么就可以从其他若干新的照片中找出包含该陌生人的照片。模仿人类的学习过程,缩小人工智能和人类之间的差距是机器学习非常重要的一个方向。

1 小样本学习

针对样本缺失的应用场景,国内外学者提出了一种新的机器学习概念:小样本学习(Few-Shot Learning)^[2-4]。典型的机器学习应用程序^[5],如上

面的举例,需要大量具有监督信息^[6]的数据。然而,正如引言中提到的,这可能是困难且复杂的。小样本学习是机器学习的一种特殊情况,下面给出其正式的定义。

1.1 问题定义

小样本学习是在只有目标少量训练样本的环境中,如何训练一个可以有效地识别这些目标样本的一类机器学习模型。现在的小样本问题主要是监督学习问题,例如小样本分类问题只给出每个类的几个带标签的例子来学习分类器。在工业应用领域,小样本学习因其可以大大降低数据采集和标定成本,在诸多视觉任务中已经得到研究人员的关注,其中包括:图像分类^[7]、图像检索^[8]、目标跟踪^[9]、短文本情感分析^[10]、语言模型^[11]、网络结构搜索^[12]等

基金项目:国家自然科学基金(61472256,61170277);上海市一流学科建设项目(S1201YLXK);上海理工大学科技发展基金资助项目(16KJFZ035,2017KJFZ033);沪江基金资助项目(A14006)。

作者简介:胡西范(1995-),男,硕士研究生,主要研究方向:小样本学习;陈世平(1964-),男,博士,教授,主要研究方向:云计算、信息检索及深度学习。

通讯作者:陈世平 Email:huxifan6@163.com

收稿日期:2021-03-06

问题。

1.2 相关的机器学习任务

(1) 弱监督学习 (Weakly supervised learning)^[13]: 弱监督学习是介于有监督和无监督之间的一类学习方法^[14], 从经验 E 中学习, 只包含弱监督信息 (如不完整、不准确、或嘈杂的监督信息)。与小样本学习相关的问题是在监督信息不完全的情况下, 小样本学习的经验只有少量标签信息。根据有无人为干预, 弱监督学习又可分为以下 2 类:

① 半监督学习 (Semi-supervised learning)^[15]: 半监督学习通过使用少量标记, 同时使用大量未标记数据对模型进行训练, 是监督学习和无监督学习相结合的一种机器学习方法。正例-未标注学习 (Positive and Unlabeled Learning)^[16] 是半监督学习的一个特例, 指在只有正类和未标记数据的情况下, 训练二分类器。例如, 在社交网络中推荐朋友^[17], 根据朋友列表只能了解用户当前的朋友, 而用户与其他人的关系是未知的。

② 不平衡学习 (Imbalanced learning)^[18]: 不平衡学习的根本问题是, 不平衡数据会极大地损害一般标准学习算法的性能。一般标准算法假设或期望平衡的类分布或相等的误分类代价。因此, 当面对复杂的不平衡数据集时, 这些算法不能很好地表示数据的分布特征, 这样便会使跨类数据的准确性较差。相反, 小样本学习用少量的样本训练和测试输出, 同时可能以其他输出作为学习的先验知识。

(2) 迁移学习 (Transfer learning)^[19-20]: 迁移学习将某个领域或任务上学习到的知识或模式应用到不同但相关的数据缺乏的领域或问题中。还可以用于跨域推荐、跨时间空间和移动设备的 WiFi 定位等应用。领域自适应 (Domain Adaptation)^[21] 是迁移学习中的一种代表性方法, 指的是利用信息丰富的源域 (source domain) 样本来提升目标域 (target domain) 模型的性能。源域和目标域往往属于同一类任务, 但是分布不同。例如, 在情感分析中, 源域数据包含用户对电影的评论, 而目标域数据包含客户对日用商品的评论。

(3) 元学习 (Meta-learning): 或者称为学会学习 (Learning to learn)^[22], 这一类学习随着经验和任务数量的增长, 在每个任务上的表现得到改进, 每当学会解决一个新的任务, 就越有能力解决其他新的任务。具体来说, 元学习器在多轮任务中迭代学习到一些通用性的知识 (元知识), 利用学习到的元知识可以帮助新的任务快速迭代, 提高新任务的性能。

元学习方法可以用来处理小样本学习问题。

1.3 问题分类

在所有的机器学习问题中, 通常都存在预测误差, 无法获得完美的预测。在小样本学习中, 由于样本量不够导致经验风险最小化带来的最优解和真实解之间的误差将变大^[23-25], 因此必须使用先验知识来解决以上问题。

这里将现有的小样本学习研究分为 3 类, 即: 利用先验知识增强监督信号; 利用先验知识缩小假设空间的大小; 利用先验知识更改给定假设空间中对最优假设的搜索, 也就是数据、模型、算法三类。接下来的章节会对每类研究方法进行详细介绍。

2 数据

基于数据的小样本学习方法主要是利用先验知识增强样本数据 D_{train} , 从而将监督信息进一步增强, 利用充足的数据样本来实现可靠的经验风险最小化, 如图 1 所示。

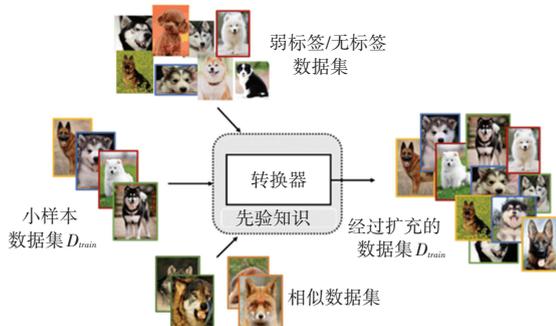


图 1 通过数据增强解决小样本学习问题

Fig. 1 Solving small sample learning problems through data enhancement

以图像数据扩充为例, 经典的图像数据扩充方法主要有: 翻转^[26]、裁剪^[27]、缩放^[28]、旋转^[29]、改变亮度、图像模糊等。然而这些人工设计的方法有很大的依赖性, 需要耗费大量的人力和专业知识。此外, 这种方法的迁移性比较差, 对一个数据集制定的数据增强方法很难适用到另一个数据集当中。因为人类不可能列举出所有可能的不变性, 因此传统的人工增强数据方法不能完全适用于解决小样本问题。

除了人工设计的样本数据扩充方法, 根据增强数据的来源将小样本数据扩充方法分为 3 类: 从训练集转换样本、从弱标签或无标签数据集转化样本、从相似数据集转换样本。

2.1 从训练集转换样本

这种策略通过将每个属于 D_{train} 的样本转化为

几个不同的样本来增强 D_{train} 。转换过程作为先验知识包含在经验 E 中,以便生成额外的样本。一篇早期的小样本学习论文^[30]通过迭代地将每个样本与其他样本对齐,从类似的类中学习一组几何变换。将学习过的转换过程应用到更大的数据集中,然后通过标准的机器学习方法来学习。同样,一组自编码器,每个从类似的类中学习,代表一个类的可能性,通过向训练样本添加新属性获得新的生成样本。通过假设所有类别在样本之间共享一些可转换的变异性,则可以学习到一个转换函数^[31],再将学习到的转换函数应用到输入样本。

2.2 从弱标签或无标签数据集转化样本

这种策略通过从弱标签或无标签的大数据集中选择带有目标标签的样本来增强 D_{train} 。例如,在监控拍摄到的照片中,有人、汽车、绿化和道路,但都不包含标签。再例如一段长时间的演讲视频,演讲者在视频中有一系列手势,所有的手势都没有明确的注释。作为这样的数据集通常包含大量不同的样本,将其增强到 D_{train} 有助于增强模型的泛化性。收集这样的数据集很容易,因为不需要人来做标记。然而,虽然收集成本很低,但一个主要问题是如何选择带有目标标签的样本来增强 D_{train} 。在文献^[32]中,研究为 D_{train} 中的每个目标标签学习了一个 SVM,用于预测来自弱标签数据集的样本的标签,然后将具有目标标签的样本添加到 D_{train} 。

2.3 从相似数据集转换样本

这种策略通过从相似、但更大的数据集聚合和调整输入-输出对来增强训练集。聚合权重通常基于样本之间的一些相似性度量。在 Tsai 等人^[33-34]的研究中,可从辅助的文本语料库中提取了聚合的权重。由于这些样本可能不来自目标小样本类,直接将聚合样本增强到 D_{train} 可能产生更大的误差。生成对抗网络(GAN)^[34]可以生成与目标样本相似、甚至和目标样本无法区分的样本数据。具体地,是由一个生成器和一个判别器构成,通过对抗学习的方式来训练^[35]。

2.4 结论

使用哪一种数据增强策略取决于实际的应用程序。通过利用目标任务的先验信息来增强数据是简单易懂的,另一方面,这种方法的缺陷是不能很好地迁移到其他数据集(特别是来自其他领域的数据集)上使用。除此之外,现有的方法主要是为图像问题设计的,因为生成的图像可以很容易地被人类视觉评估。相反,文本和音频涉及语法和结构的领

域更难生成。在文献^[36]的研究中进行了使用数据增强文本的尝试。

3 模型

若只给定少数样本的训练集 D_{train} ,同时仅使用简单模型(例如线性分类器)就可以选择较小的假设空间^[37-38]。但是,现实世界中的问题通常更为复杂,并且不能由一个小型的假设空间很好地表示^[39]。根据所使用的先验知识,可以将属于本类别的方法进一步分为多任务学习、嵌入式学习、借助外部存储学习和生成模型四种类型。下面,将分别介绍这四种类型。

3.1 多任务学习

在存在多个相关任务的情况下,多任务学习^[40]通过利用任务通用信息和特定于任务的信息同时学习这些任务。因此,可以将多任务学习用于小样本问题中。假设有一个机器学习任务 T ,该任务有相关的若干个子任务 T_1, T_2, \dots, T_n ,其中一些任务的样本很少,另一些的样本数很多。每个任务 T_i 都有一个数据集 $D_i = \{D_{train}^i, D_{test}^i\}$,其中前者是训练集,后者是测试集。在这些 T_i 任务中,将小样本任务作为目标任务,其余作为源任务。多任务学习从 D_{train}^i 学习 T_i 的参数。

由于这些任务是联合学习的,因此为任务 T_i 学习的参数受其他任务的约束。根据任务参数的约束方式,多任务学习方法又可以划分为:参数共享^[28,41-42],参数绑定^[39,43]。

3.2 嵌入学习

嵌入学习^[44-45]将每个样本 $x_i \in X \subseteq R^d$ 嵌入到一个低维的 $z_i \in Z \subseteq R^m$,这样相似的样本距离会缩小,而差异较大的样本则更容易分离。然后,利用这个较低维的 Z 中构造一个更小的假设空间 H ,此时只需要少量的训练样本。嵌入学习具有以下关键组成部分:

- (1) 将测试样本 $x_{test} \in D_{test}$ 嵌入 Z 的函数 f 。
- (2) 将训练样本 $x_i \in D_{train}$ 嵌入 Z 的函数 g 。
- (3) 相似性函数 $s(\cdot, \cdot)$ 来测量 $f(x_{test})$ 和 Z 中的 $g(x_i)$ 之间的相似度。

根据该类的嵌入 $g(x_i)$ 与 Z 中的 $f(x_{test})$ 最相似,将测试样本 x_{test} 分配给 x_i 类。尽管可以为 x_i 和 x_{test} 使用通用的嵌入函数,但是使用 2 个单独的嵌入函数可以获得更好的准确性^[46-47]。根据嵌入函数 f 和 g 的参数是否随任务而变化,将这些小样本学习

方法归为3类,即:特定于任务的嵌入模型;不变任务(即一般)嵌入模型;混合嵌入模型,可同时编码特定于任务的信息和不变于任务的信息。

3.3 使用外部存储器学习

使用外部存储器学习^[48-49]从 D_{train} 中提取知识,并将提取的信息存储在特定的外部存储器中。然后,将每个新样本 x_{test} 由从内存中提取的内容用加权平均值表示。这限制了 x_{test} 由内存中的内容表示,因此实质上减小了假设空间 H 的大小。小样本学习中通常使用键值存储器^[50],根据存储器的功能此类方法又可以分为准确表达^[51-52]和参数完善^[53-54]两类。

3.4 生成模型

生成模型方法借助先验知识(如图2所示)从观测到的 x_i 估计概率分布 $p(x)$ 。 $p(x)$ 的估计通常涉及 $p(x|y)$ 和 $p(y)$ 的估计。生成模型方法可以用于处理很多任务,例如生成^[55-58]、识别^[57-58]、重构^[57]和图像翻转^[56]。

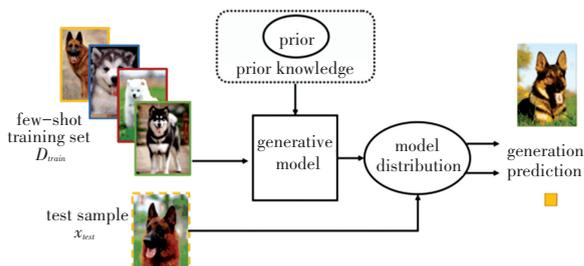


图2 通过生成模型解决小样本问题

Fig. 2 Solving small sample problems by generating models

3.5 模型方法总结

当存在相似的任务或辅助任务时,可以使用多任务学习来约束小样本任务的假设空间 H 。但是此方法需要共同训练所有任务。因此,当添加一个新的任务时,整个多任务模型必须再次共同训练,这将导致训练的成本增加。当存在一个由各类样本数据组成的大规模数据集时,可以考虑使用嵌入学习方法。此方法的优点是可以将样本映射到良好的嵌入空间,并易于分离来自不同类别的样本,因此需要较小的 \tilde{H} 。但是,当小样本任务与其他任务没有密切关系时,实验效果将会不佳。当有可用的内存网络时,可以通过在内存的基础上训练一个简单的模型(例如分类器),将其简单地用于小样本学习任务。通过使用专门设计的更新规则,可以有选择地保护内存插槽。但是本方法的缺点是会导致额外的存储空间和计算成本,随着内存大小的增加缺点越明显,而使用的外部存储器大小通常也会受到限制。

最后,当除了小样本任务之外还想要执行诸如生成和重构之类的任务时,可以使用生成模型。训练好的生成模型也可以用于生成样本对数据进行扩充。

4 总结与展望

由于获取大量标注样本会消耗不少的时间和人力,小样本学习在深度学习领域逐渐受到人们重视。目前小样本学习研究主要集中在图像分类、字符识别等方向。在其他监督学习场景中同样可以发挥小样本学习理论的作用,包括目标检测、语义分割、图像检索、手势识别、视频目标检测等问题。除了视觉任务,在自然语言处理中领域的文本和计算机视觉的视频问题同样是小样本学习值得探索的研究方向。

随着深度学习的不断发展,小样本学习应用的领域将会不断扩大,小样本学习研究的前景将会十分广阔。

5 结束语

本文首先介绍了小样本学习的研究背景和问题定义,接着从3种数据增强方法和4种模型详细介绍了小样本学习方法,最后对小样本学习研究进行了前景展望。

参考文献

- [1] SEBASTIAN T, PRATT L. Learning to learn[M]. New York: Springer Science & Business Media, 2012.
- [2] GARCIA V, BRUNA J. Few-shot learning with graph neural networks[J]. arXiv preprint arXiv:1711.04043, 2017.
- [3] DUAN Y, ANDRYCHOWICZ M, STADIE B, et al. One-shot imitation learning [C]//Advances in Neural Information Processing Systems. Long Beach, CA: Microsoft, 2017: 1087-1098.
- [4] ORESHKIN B, LÓPEZ P R, LACOSTE A. Tadam: Task dependent adaptive metric for improved few-shot learning[C]//Advances in Neural Information Processing Systems. Montreal, Canada: NIPS, 2018: 721-731.
- [5] REN M, TRIANTAFILLOU E, RAVI S, et al. Metalearning for semi-supervised few-shot classification[J]. arXiv preprint arXiv:1803.00676, 2018.
- [6] MITCHELL M T. Machine learning[M]. New York: McGraw-Hill, 1997.
- [7] MOHRI M, ROSTAMIZADEH A, TALWALKAR A. Foundations of machine learning[M]. Cambridge: MIT Press, 2018.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. Nevada, USA: Winton, 2012: 1097-1105.
- [9] SMEULDERS A W M, WORRING M, SANTINI S, et al.

- Content-based image retrieval at the end of the early years [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2000 (12): 1349-1380.
- [10] BLACKMAN S S. Multiple-target tracking with radar applications [M]. Norwood: Artech House, 1986.
- [11] YU Mo, GUO Xiaoxiao, YI Jinfeng, et al. Diverse few-shot text classification with multiple metrics [C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. New Orleans, Louisiana; Bloomberg, 2018; 1206-1215.
- [12] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model [J]. *Journal of machine learning research*, 2003, 3(2): 1137-1155.
- [13] ZOPH B, LE Q V. Neural architecture search with reinforcement learning [J]. *arXiv preprint arXiv:1611.01578*, 2016.
- [14] DIBA A, SHARMA V, PAZANDEH A, et al. Weakly supervised cascaded convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA; IEEE, 2017: 914-922.
- [15] ZHOU Zhihua. A brief introduction to weakly supervised learning [J]. *National Science Review*, 2018(1): 44-53.
- [16] ZHU X J. Semi-supervised learning literature survey [R]. Madison; University of Wisconsin-Madison, 2005.
- [17] LI Xiaoli, YU P S, LIU Bing, et al. Positive unlabeled learning for data stream classification [C]//SIAM International Conference on Data Mining. Sparks, Nevada, USA; dblp, 2009; 259-270.
- [18] SETTLES B. Active learning literature survey [R]. Madison; University of Wisconsin-Madison, 2010.
- [19] HE H, GARCIA E A. Learning from imbalanced data [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21 (9): 1263-1284.
- [20] PANS J, YANG Q. A survey on transfer learning [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22 (10): 1345-1359.
- [21] FU Y, HOSPEDALES T M, XIANG T, et al. Transductive multi-view zero-shot learning [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015, 37(11): 2332-2345.
- [22] BEN-DAVID S, BLITZER J, CRAMMER K, et al. Analysis of representations for domain adaptation [C]//Advances in Neural Information Processing Systems. Vancouver, BC, Canada; NIPS Foundation, 2007; 137-144.
- [23] HOCHREITERS, YOUNGER A S, CONWELL P R. Learning to learn using gradient descent [C]//International Conference on Artificial Neural Networks. Vienna, Austria; Springer, 2001: 87-94.
- [24] CHEN Zitian, FU Yanwei, ZHANG Yinda, et al. Multi-level semantic feature augmentation for one-shot learning [J]. *IEEE Transactions on Image Processing*, 2019, 28(9): 4594-4605.
- [25] ANTONIOU A, STORKEY A, EDWARDS H. Data augmentation generative adversarial networks [J]. *arXiv preprint arXiv:1711.04340*, 2017.
- [26] CHEN Z, FU Y, CHEN K, et al. Image block augmentation for one-shot learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Hawaii, USA; AAAI, 2019, 33: 3379-3386.
- [27] QI H, BROWN M, LOWE D G. Low-shot learning with imprinted weights [C]//Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA; IEEE, 2018; 5822-5830.
- [28] SHYAM P, GUPTA S, DUKKIPATI A. Attentive recurrent comparators [C]//International Conference on Machine Learning. Sydney, NSW, Australia; IMLS, 2017; 3173-3181.
- [29] ZHANG Yabin, TANG Hui, JIA Kui. Fine-grained visual categorization using meta-learning optimization with sample selection of auxiliary data [M]//FERRARI V, HEBERT M, SMINCHISESCU C, et al. *Computer Vision - ECCV 2018*. Lecture Notes in Computer Science. Cham; Springer, 2018, 11212: 241-256.
- [30] VINYALS O, BLUNDELL C, LILICRAP T, et al. Matching networks for one shot learning [C]//Advances in Neural Information Processing Systems. Barcelona; MIT Press, 2016; 3630-3638.
- [31] MILLER E G, MATSAKIS N E, VIOLA P A. Learning from one example through shared densities on transforms [C]//Conference on Computer Vision and Pattern Recognition. Hilton Head, SC, USA; IEEE, 2000, 1: 464-471.
- [32] HARIHARAN B, GIRSHICK R. Low-shot visual recognition by shrinking and hallucinating features [C]//International Conference on Computer Vision. Venice, Italy; IEEE, 2017; 3037-3046.
- [33] PFISTER T, CHARLES J, ZISSERMAN A. Domain-adaptive discriminative one-shot learning of gestures [C]//European Conference on Computer Vision. Zurich, Switzerland; Springer International Publishing, 2014; 814-829.
- [34] TSAI Y H, SALAKHUTDINOV R. Improving one-shot learning through fusing side information [J]. *arXiv preprint arXiv:1710.08347*, 2017.
- [35] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]//Advances in Neural Information Processing Systems. Montréal CANADA; Google, 2014; 2672-2680.
- [36] 王坤峰, 苟超, 段艳杰, 等. 生成式对抗网络 GAN 的研究进展与展望 [J]. *自动化学报*, 2017, 43(3): 321-332.
- [37] WEI J, ZOU K. EDA: Easy data augmentation techniques for boosting performance on text classification tasks [C]//Conference on Empirical Methods in Natural Language Processing and International Joint Conference on Natural Language Processing. Hong Kong, China; ACL, 2019; 6383-6389.
- [38] MITCHELL M T. *Machine Learning* [M]. New York; McGraw-Hill, 1997.
- [39] MOHRIM, ROSTAMIZADEH A, TALWALKAR A. *Foundations of machine learning* [M]. Cambridge; MIT Press, 2018.
- [40] GOODFELLOW I, BENGIO Y, COURVILLE A. *Deep learning* [M]. Cambridge; MIT Press, 2016.
- [41] ZHANG Yu, YANG Qiang. A survey on multi-task learning [J]. *arXiv preprint arXiv:1707.08114*, 2017.
- [42] HU Zikun, LI Xiang, TU Cunchao, et al. Few-shot charge prediction with discriminative legal attributes [C]//International Conference on Computational Linguistics. New Mexico, USA; ACL, 2018; 487-498.
- [43] BENAÏM S, WOLF L. One-shot unsupervised cross domain translation [C]//Advances in Neural Information Processing Systems. Montreal, Canada; NIPS, 2018; 2104-2114.
- [44] LUO Z, ZOU Y, HOFFMAN J, et al. Label efficient learning of transferable representations across domains and tasks [C]//Advances in Neural Information Processing Systems. Long Beach; Microsoft, 2017; 165-177.