概率隐含语义分析方法[159,160]、协同矩阵分解技术[161]、积极迁移学习框架[162].

根据源领域和目标领域是否有标注样本[99],可以把迁移学习分为归纳迁移学习、直推式迁移学习和无监督迁移学习.归纳迁移学习[157,163−166]的目标领域中有少量标注样本,根据源领域中是否有标注样本,还可以把归纳迁移学习划分成多任务学习(源领域中有标注数据)、自学习(源领域中没有标注数据).直推式迁移学习[167−170]是只有源领域中有标签样本,源领域和目标领域的数据相关但不相同,两个领域的任务相同.无监督迁移学习[156,171]处理的是源领域和目标领域中都没有标签数据的问题.

按照迁移的内容对迁移学习进行划分,主要有特征表示迁移、实例迁移、参数迁移和关联关系迁移.特征表示迁移期望联合表示的特征优于只基于目标领域中数据的特征表示[172],代表性工作有自学习聚类算法[156]、基于概率的隐含语义分析算法[152]、基于流形结构的算法[154]等.实例迁移是指从源领域训练数据中抽取适合测试数据的实例,迁移到目标领域增加训练数据数目.经典模型包括 Tradaboosting 算法[157]、基于隐含稀疏领域迁移方法[173]等.参数迁移假设源领域和目标领域的一些参数是共享的,代表性工作有基于高斯过程 Gaussian Process(GP)的模型[174,175]以及结合层次贝叶斯 hierarchical Bayesian(HB)框架的模型[176,177].关联关系迁移的方法通常数据表示为关联关系,例如社会网络数据.通过源领域和目标领域的数据,迁移它们之间的关联关系,代表性工作有利用 Markov Logic 网络进行关系的迁移[178−180].

以上介绍的迁移学习算法形式各异,例如基于矩阵分解、概率主题模型等.表 4 对迁移学习方法进行了总结.

Table 4 Description of different kinds of transfer learning strategies

表 **4** 不同迁移学习方法的描述

迁移学习		描述	代表性工作
按照源领域和目标领域的特征空间是否相同划分	同构迁移学习	源领域和目标领域的特征空间相同,可以把与目标领域相关的源领域中的数据直接应用到目标领域中	基于主题的隐含语义分析算法[152]、谱分析算法[153]、自学习聚类算法[156]、Tradaboosting算法[157]、基于深度学习的方法[181−183]等
	异构迁移学习	源领域和目标领域的特征空间不同,通常需要学习异构的源领域和目标领域之间的关系,可以直接进行特征映射,也可以映射到共同的子空间中	风险最小化框架[158]、概率隐含语义分析方法[159,160]、协同矩阵分解技术[161]、积极迁移学习框架[162]、基于深度学习的方法[100,184,185]等
按照源领域和目标领域中是否有标注数据划分	归纳迁移学习	目标领域中有少量标注数据,根据源领域中是否有标注数据,可以把归纳迁移学习划分成多任务学习(源领域中有标注数据)、自学习(源领域中没有标注数据)	Tradaboosting算法[157]、基于Procrustes的流形对齐方法[164]、基于深度学习的方法[183]等
	直推式迁移学习	只有源领域中有标注样本,源领域和目标领域的数据相关但不相同,两个领域的任务相同	Structural correspondence学习模型(SCL)[167]、Bridged refinement模型[168]、基于深度学习的方法[182,186]等
	无监督迁移学习	源领域和目标领域中都没有标签数据的问题	自学习聚类算法STC[156]、迁移判别分析TDA[172]等
按照迁移的内容划分	特征表示迁移学习	期望源领域和目标领域学到的联合表示特征优于只基于目标领域中数据的特征表示	基于主题的隐含语义分析算法[152]、基于流形结构的算法[154]、自学习聚类算法[156]等、基于深度学习的方法[184,187,188]
	实例迁移学习	从源领域训练数据中抽取一些实例,迁移到目标领域增加训练数据数目	Tradaboosting算法[157]、基于重建的隐含稀疏领域迁移方法[173]等
	参数迁移学习	挖掘源领域和目标领域的共享参数,用于目标领域中	基于高斯过程Gaussian Process(GP)的模型[174,175]以及结合层次贝叶斯hierarchical Bayesian(HB)框架的模型[176,177]等
	关联关系迁移学习	通过源领域和目标领域的数据,迁移它们之间的关联关系	利用Markov Logic网络迁移的模型[178−180]等

这些模型大都是浅层结构,随着具有多层结构的深度学习在一些领域获得了成功,研究者把深度学习引入到迁移学习模型中,通过构建具有很多隐层的结构,学习更有用的特征,最终提升迁移学习在目标领域中任务的性能[100,181−193].Bengio 等人[181]研究了无监督预训练特征的有效性,将其应用到迁移学习场景下.Glorot 等人[182]将不同领域的数据输入到叠加降噪自动编码器中,学习更加健壮的特征,对源领域和目标利用中的样本进行重

新表示.Oquab 等人[183]首先利用源领域中已标注的样本训练 CNN,然后增加适应层缩小两个领域之间的差异,并且利用目标领域中已标注样本微调训练的 CNN.Yosinski 等人[189]量化了深层神经网络中每一层特征的可迁移性.Zhuang 等人[190]结合深度自动编码器进行迁移,通过最小化源领域和目标领域的隐藏层的 KL 距离获得领域不变的特征.Long 等人[186]提出了联合自适应网络的结构,应对目标领域中已标注样本数量较少的问题.Sun 等人[187]和 Rozantsev 等人[187]提出了深度领域自适应的方法.文献[100,184,185]设计了基于深度学习的异构迁移学习算法,应用到人脸识别、图像-文本迁移中.深度学习在迁移学习上获得了较好的性能,但是在其可解释性和参数调整方面,仍然需要进一步研究.

研究者从理论层面对迁移学习进行了研究,Ben-David 等人[194]基于 VC 维对领域适应性问题给出了推广性的界.对于有限 VC 维的情况,可用文献[195]中提出的方法,从有限个样本估计适应推广能力.Blitzer 等人[169]从源数据加权组合获得模型,并给出在特定的经验风险最小化的情形下的误差率.Ben-David 等人[196]研究了分类器能够在目标领域很好地完成分类任务的条件.Mansour 等人[197,198]针对任意目标分布给出了基于源领域和目标领域之间 Rényi 散度的领域推广误差,提出通过加权经验分布可较为准确地反映目标领域分布.Zhang 等人[199]提出一种新的框架来分析典型的领域适应学习过程的理论性质,分别开发了 Hoeffding 型、Bennett 型和 McMiarmid 型偏差不等式,提出了基于 Rademacher 复杂度的泛化边界,并分析了渐近收敛性和学习过程的收敛速度.Kumagai[200]分析了参数迁移学习的 margin 边界.然而,正如庄福振等人[101]所指出的,尽管研究者对迁移学习已经进行了一些理论尝试,但还远远不够,尤其是需要深入开展迁移学习有效性的理论研究.

通过以上分析可以发现,不同多视图学习和迁移学习的算法设计理论各不相同.然而从认知的角度考虑,多源学习方法的总体目标是为了学习隐藏于多源数据下的共同知识.已有的多源学习方法只是根据具体问题设计的具体算法,这些方法没有共性的约束,不能指导新的学习算法设计.

4 未来研究方向的思考

虽然在基础理论研究和应用领域,多源数据学习已经成为研究热门且存在一些较为成熟的技术,但是现有的多源数据学习算法对应的学习任务差别巨大,其学习算法的表示也严重碎片化,彼此形式差别极大.早在 2004 年,周志华就指出[201],机器学习“以 Tom Mitchell 的经典教科书为例,很难看到基础学科(例如数学、物理学)教科书中那种贯穿始终的体系,也许会让人感到这不过是不同方法和技术的堆砌”.

机器学习算法(包括单源和多源数据学习算法)的表示碎片化和形式差别化,对研究如何统一机器学习算法的理论带来很大的困难.如果像 Vapnik 那样将机器学习问题看作是一个基于经验值的函数估计问题[202],则会失去对学习问题的内在约束,对很多机器学习算法的设计启发性不足.更重要的是,如果将机器学习问题看作一个基于经验值的函数估计问题,则几乎完全隐蔽了对学习问题所具有的共同内蕴认知性质,即学习是为了完成一个认知任务(从数据中形成知识).多源数据是由单源数据组合而成的,多源数据学习也是一个典型的认知任务,希望像人一样进行多源学习.因此,应该将多源数据学习所具有的共同内蕴的认知性质挖掘出来.由于现今的单源数据学习理论对于这一点研究不足,而现有的多源数据学习方法往往是由单源数据学习方法扩展而来,所以面临着与单源数据学习同样的理论薄弱问题,很难解决多源数据学习算法的理论性能问题.根据上述分析,我们认为,多源数据的理论和算法的未来研究方向包括如下几个方面.

(1) 目前,人类学习的认知机理存在着两个关键问题:一是学习的认知机理不一致,不能确定是相似性还是简单性起主导作用;二是尽管多模态学习是人类学习的基本特性,但当前研究大多探讨单模态的学习机理.在人类知觉学习、类别学习和概念表征以及机器学习中,相似性都起到了重要作用.因此,未来的研究热点之一是以相似性为中心,基于单模态相似性来研究多模态相似性的整合机制,从而得到适合于机器学习使用的人类多源数据学习的认知机理,即概念表征的认知机理.

(2) 为了从认知上统一各种学习算法,研究者已经做出了很多努力,提出了许多理论——PAC 学习理论、统计学习理论、概率图理论等.现在的学习理论都只覆盖了部分学习算法.因此,两位机器学习领域的指标性人物 Jordan 和 Mitchell 在《Science》上提出了机器学习的一个重要挑战:能否建立一个能够统管所有机器、人和生

物的学习理论?机器学习的认知目的是从数据中得到知识,而知识的基本单位是概念.因此,未来的研究热点之一是如何利用概念表征的认知机理进行概念的统一表示理论,进而将学习算法进行统一表示,特别是将多源数据学习进行统一表示,在此基础之上,发现学习算法的基本认知假设,即机器学习公理化研究.

(3) 现在的多源数据学习算法大致的设计思路有 3 种:前期融合算法,强调特征融合;后期融合算法,强调结果融合;中期融合算法,强调子空间共性.这些思路都需要设计概念表示理论,即学习算法的认知基本假设.未来的研究热点之一是在机器学习公理化的研究基础上,研究多源的学习算法设计原则和学习算法评估原则,研究一系列多源数据学习算法设计.

References:

[1] Nature. Big data. 2008. http://www.nature.com/news/specials/bigdata/index.html

[2] Science. Special online collection: Dealing with data. 2011. http://www.sciencemag.org/site/special/data/

[3] Silver D, Huang A, Maddison CJ. Mastering the game of Go with deep neural networks and tree search. Nature, 2016,529(7587): 484−489. [doi: 10.1038/nature16961]

[4] http://www.aaai.org/Conferences/AAAI/2016/aaai16speakers.php#Hassabis

[5] Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor J. Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 2001,18(1):32−80. [doi: 10.1109/79.911197]

[6] Cohen I, Cozman F, Sebe N, Cirelo M, Huang T. Semisupervised learning of classifiers: Theory, algorithms and their applications to human- computer interaction. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2004,26(12):1553−1567. [doi: 10.1109/TPAMI. 2004.127]

[7] Bekkerman R, Jeon J. Multi-Modal clustering for multimedia collections. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2007. 1−8. [doi: 10.1109/CVPR.2007.383223]

[8] Wang J, Duan L, Liu Q, Lu H, Jin J. A multimodal scheme for program segmentation and representation in broadcast video streams. IEEE Trans. on Multimedia, 2008,10(3):393−408. [doi: 10.1109/TMM.2008.917362]

[9] Cristani M, Bicego M, Murino V. Audio-Visual event recognition in surveillance video sequences. IEEE Trans. on Multimedia, 2007,9(2):257−267. [doi: 10.1109/TMM.2006.886263]

[10] Liu J, Lai W, Hua XH, Huang Y, Li S. Video search re-ranking via multi-graph propagation. In: Proc. of the 15th ACM Int'l Conf. on Multimedia. 2007. 208−217. [doi: 10.1145/1291233.1291279]

[11] Valiant L. A theory of the learnable. Communications of the ACM, 1984,27(11):1134−1142. [doi: 10.1145/1968.1972]

[12] Cortes C, Vapnik VN. Support vector networks. Machine Learning, 1995,20(3):273−297. [doi: 10.1007/BF00994018]

[13] Kleinberg J. An impossibility theorem for clustering. In: Proc. of the Advances in Neural Information Processing Systems. 2003. 463−470.

[14] Valiant L. Probably approximately correct: Nature's Algorithms for Leaning and Prospering in a Complex World. New York: Basic Books, 2013.

[15] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo: Morgan Kaufmann Publishers, 1988.

[16] Pavlas J. Searching for the algorithms underlying life. Quanta Magazine, 2016. https://www.quantamagazine.org/the-hidden-algorithms-underlying-life-20160128/

[17] Snedeker J. Word learning. In: Binder MD, Hirokawa N, Windhorst U, eds. Encyclopedia of Neuroscience. Springer-Verlag, 2009. 503−508.

[18] Urcuioli PJ, Wasserman EA, Zentall TR. Associative concept learning in animals: Issues and opportunities. Journal of the Experimental Analysis of Behavior, 2014,101(1):165−170. [doi: 10.1002/jeab.62]

[19] Daniel TA, Cook RG, Katz JS. Temporal dynamics of task switching and abstract-concept learning in pigeons. Front Psychol, 2015,6:1−8. [doi: 10.3389/fpsyg.2015.01334]

[20] Gingras G, Rowland BA, Stein BE. The differing impact of multisensory and unisensory integration on behavior. Journal of Neuroscience, 2009,29(15):4897−4902. [doi: 10.1523/JNEUROSCI.4120-08.2009]

[21] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science, 2015,349(6245):255−260. [doi: 10.1126/science.aaa8415]

[22] Stein BE, Stanford TR. Multisensory integration: Current issues from the perspective of the single neuron. Nature Reviews Neuroscience, 2008,9(4):255−266. [doi:10.1038/nrn2331]

[23] Fu X, Cai L, Liu Y, Jia J, Chen W, Yi Z, Zhao G, Liu Y, Wu C. A computational cognition model of perception, memory, and judgment. Science China (Information Sciences), 2014,57(3):1−15. [doi: 10.1007/s11432-013-4911-9]

[24] Seger CA. How do the basal ganglia contribute to categorization? Their role in generalization, response selection, and learning via feedback. Neuroscience and Biobehavioral Reviews, 2008,32(2):265−278. [doi: 10.1016/j.neubiorev.2007.07.010]

[25] Seger CA, Miller EK. Category learning in the brain. Annual Review of Neuroscience, 2010,33:203−219. [doi: 10.1146/annurev.neuro.051508.135546]

[26] Fugate JM. Categorical perception for emotional faces. Emotion Review, 2013,5(1):84–89. [doi: 10.1177/1754073912451350]

[27] Young AW, Rowland D, Calder AJ, Etcoff NL, Seth A, Perrett DI. Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. Cognition, 1997,63:271−313. [doi: 10.1016/S0010-0277(97)00003-6]

[28] Ashby FG, Alfonso-Reese LA, Turken AU, Waldron EM. A neuropsychological theory of multiple systems in category learning. Psychological Review, 1998,105:442−481. [doi: 10.1037/0033-295X.105.3.442]

[29] Flavell JH. Cognitive Development. Englewood Cliffs: Prentice-Hall, 1985.

[30] Ashby FG, OBrien JB. Category learning and multiple memory systems. Trends in Cognitive Sciences, 2005,9:83−89. [doi: 10.1016/j.tics.2004.12.003]

[31] Ernst MO, Bülthoff HH. Merging the senses into a robust percept. Trends in Cognitive Sciences, 2004,8(4):162−169. [doi: 10.1016/j.tics.2004.02.002]

[32] Alais D, Newell FN, Mamassian P. Multisensory processing in review: From physiology to behaviour. Seeing and Perceiving, 2010,23(1):3−38. [doi: 10.1163/187847510X488603]

[33] Wassenhove van V, Grantand KW, Poeppel D. Temporal window of integration in auditory-visual speech perception. Neuropsychologia, 2007,45(3):598−607. [doi: 10.1016/j.neuropsychologia.2006.01.001]

[34] Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron, 2008,57(1):11−23. [doi: 10.1016/j.neuron.2007.12.013]

[35] Ursino M, Cuppini C, Magosso E. A neural network for learning the meaning of objects and words from a featural representation. Neural Networks, 2015,63:234−253. [doi: 10.1016/j.neunet.2014.11.009]

[36] Shams L, Seitz AR. Benefits of multisensory learning. Trends in Cognitive Sciences, 2008,12(11):411−417. [doi: 10.1016/j.tics.2008.07.006]

[37] Hahn U, Chater N, Riehardson L. Similarity as transformation. Cognition, 2003,87(1):1−32. [doi: 10.1016/S0010-0277(02)00184-1]

[38] Lin N, Guo QH, Han ZZ, Bi Y. Motor knowledge is one dimension for concept organization: Further evidence from a Chinese semantic dementia case. Brain and Language, 2011,119(2):110−118. [doi: 10.1016/j.bandl.2010.07.001]

[39] Bub DN, Masson ME, Cree GS. Evocation of functional and volumetric gestural knowledge by objects and words. Cognition, 2008,106(1):27−58. [doi: 10.1016/j.cognition.2006.12.010]

[40] Vingerhoets G. Knowing about tools: Neural correlates of tool familiarity and experience. Neuroimage, 2008,40(3):1380−1391. [doi: 10.1016/j.neuroimage.2007.12.058]

[41] Buxbaum LJ, Kalénine S. Action knowledge, visuomotor activation, and embodiment in the two action systems. Annals of the New York Academy of Sciences, 2010,1191(1):201−218. [doi: 10.1111/j.1749-6632.2010.05447.x]

[42] Barsalou LW. Grounded cognition. Annual Review of Psychology, 2008,59:617−645. [doi: 10.1146/annurev.psych.59.103006.093639]

[43] Ell SW, Ing AD, Maddox WT. Critrial noise effects on rule-based category learning: The impact of delayed feedback. Attention, Perception, and Psychophysics, 2009,71(6):1263−1275. [doi: 10.3758/APP.71.6.1263]

[44] Ashby FG, Maddox WT. Human category learning 2.0. The Year in Cognitive Neuroscience, 2010,1224(1):147−161. [doi: 10.1111/j.1749-6632.2010.05874.x]

[45] Hahn U. Similarity. Wiley Interdisciplinary Reviews: Cognitive Science, 2014,5(3):271−280. [doi: 10.1002/wcs.1282]
[46] Kloos H, Sloutsky VM. What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. Journal of Experimental Psychology: General, 2008,137(1):52−72. [doi: 10.1037/0096-3445.137.1.52]
[47] Lelkes AD. Algorithms and complexity results for learning and big data [Ph.D. Thesis]. University of Illinois at Chicago, 2017.
[48] Ehrenfeucht A, Haussler D, Kearns M, Valiant L. A general lower bound on the number of examples needed for learning. Information and Computation, 1989,82(3):247−261. [doi: 10.1016/0890-5401(89)90002-3]
[49] Blumer A, Ehrenfeucht A, Haussler D, Warmuth MK. Learnability and the Vapnik-Chervonenkis dimension. Journal of the Association on Computing Machinery, 1989,36(4):929−965. [doi: 10.1145/76359.76371]
[50] Auer P, Ortner R. A new PAC bound for intersection-closed concept classes. Machine Learning, 2007,66(2-3):151−163. [doi: 10.1007/s10994-006-8638-3]
[51] Warmuth M. The optimal PAC algorithm. In: Proc. of the 17th Annual Conf. on Learning Theory. 2004. 641−642. [doi: 10.1007/978-3-540-27819-1_45]
[52] Haussler D, Littlestone N, Warmuth MK. Predicting {0,1} functions on randomly drawn points. Information and Computation, 1994,115(2):284−293.
[53] Hanneke S. Theoretical foundations of active learning [Ph.D. Thesis]. Carnegie Mellon University, 2009.
[54] Darnstadt M. The optimal PAC bound for intersection-closed concept classes. Information Processing Letters, 2014,115(4):458−461. [doi: 10.1016/j.ipl.2014.12.001]
[55] Simon HU. An almost optimal PAC algorithm. In: Proc. of the Int'l Conf. on Machine Learning Workshop. 2015. 1552−1563.
[56] Hanneke S. The optimal sample complexity of PAC learning. Journal of Machine Learning Research, 2016,17(38):1−15.
[57] Vapnik VN, Chervonenkis AY. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications, 1971,16(2):264−280. [doi: 10.1137/1116025]
[58] Vapnik VN. Statistical Learning Theory. Wiley-Interscience, 1989.
[59] Schapire R, Freund Y, Bartlett P, Lee WS. Boosting the margin a new explanation for the effectiveness of voting methods. Annals of Statistics, 1998,26:1651−1686. [doi: 10.1214/aos/1024691352]
[60] Breiman L. Prediction games and arcing algorithms. Neural Computation, 1999,11(7):1493−1517. [doi: 10.1162/089976699300016106]
[61] Wang L, Sugiyama M, Jing Z, Yang C, Zhou ZH, Feng J. A refined margin analysis for boosting algorithms via equilibrium margin. Journal of Machine Learning Research, 2011,12:1835−1863.
[62] Gao W, Zhou ZH. On the doubt about margin explanation of boosting. Artificial Intelligence, 2013,203:1−18. [doi: 10.1016/j.artint.2013.07.002]
[63] Zhang T, Zhou ZH. Large margin distribution machine. In: Proc. of the 20th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. 2014. 313−322. [doi: 10.1145/2623330.2623710]
[64] Hanneke S. Refined error bounds for several learning algorithms. Journal of Machine Learning Research, 2016,17(135):1−55.
[65] Wu XX, Zhang JP. Researches on Rademacher complexities in statistical learning theory: A survey. Acta Automatica Sinica, 2017,43(1):20−39 (in Chinese with English abstract).
[66] Shawe-Taylor J, Bartlett PL, Williamson RC, Anthony M. Structural risk minimization over data-dependent hierarchies. IEEE Trans. on Information Theory, 1998,44(5):1926−1940. [doi: 10.1109/18.705570]
[67] Bartlett PL, Mendelson SR. Rademacher and Gaussian complexities: Risk bounds and structural results. Journal of Machine Learning Research, 2003,3:463−482.
[68] Gnecco G, Sanguineti M. Approximation error bounds via Rademacher's complexity. Applied Mathematical Sciences, 2008,4(2):153−176.
[69] Koltchinskii V, Panchenko D. Rademacher processes and bounding the risk of function learning. In: Proc. of the High Dimensional Probability II. Birkhäuser Boston: Cambridge, 2000. 443−457. [doi: 10.1007/978-1-4612-1358-1_29]
[70] Bartlett PL, Bousquet O, Mendelson S. Local Rademachercomplexities. The Annals of Statistics, 2005,33(4):1497−1537. [doi: 10.1214/009053605000000282]

[71] Koltchinskii V. Local Rademacher complexities and oracle inequalities in risk minimization. Annals of Statistics, 2006,34(6): 2593−2656. [doi: 10.1214/009053606000001019]

[72] Zhivotovskiy N, Hanneke S. Localization of VC classes: Beyond local Rademacher complexities. In: Proc. of the Int’l Conf. on Algorithmic Learning Theory. 2016. 18−33. [doi: 10.1007/978-3-319-46379-7_2]

[73] Xu C, Liu T, Tao D, Xu C. Local Rademacher complexity for multi-label learning. IEEE Trans. on Image Processing, 2016,25(3): 1495−1507. [doi: 10.1109/TIP.2016.2524207]

[74] Lei YW, Ying YM. Generalization analysis of multi-modal metric learning. Analysis and Applications, 2016,14(4):503−521. [doi: 10.1142/S0219530515500104]

[75] Gao W, Zhou ZH. Dropout Rademacher complexity of deep neural networks. Science China (Information Sciences), 2016,59(7): 1−12. [doi: 10.1007/s11432-015-5470-z]

[76] Wainwright MJ, Jordan MI. Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning, 2008,1(1-2):1−305. [doi: 10.1561/2200000001]

[77] Koller D, Friedman N. Probabilistic Graphical Models. MIT Press, 2009.

[78] Jordan MI. Graphical models. Statistical Science, 2004,19(1):140−155. [doi: 10.1214/088342304000000026]

[79] Pearl J. Asymptotic properties of minimax trees and game-searching procedures. Artificial Intelligence, 1980,14(2):113−138.

[80] Jensen FV. An Introduction to Bayesian Networks. Springer-Verlag, 1996.

[81] Jordan MI, ed. Learning in Graphical Models. Kluwer: The Netherlands, 1998.

[82] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 1977,39(1):1−38.

[83] Pearl J. Reverend Bayes on inference engines: A distributed hierarchical approach. In: Proc. of the 2nd National Conf. on Artificial Intelligence. 1982. 133−136.

[84] Murphy KP, Weiss Y, Jordan MI. Loopy belief propagation for approximate inference: An empirical study. In: Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence. 1999. 467−475.

[85] Mooij J, Kappen H. Sufficient conditions for convergence of the sum-product algorithm. IEEE Trans. on Information Theory, 2007, 53(12):4422−4437. [doi: 10.1109/TIT.2007.909166]

[86] Wainwright MJ, Jaakkola TS, Willsky AS. Tree-Reweighted belief propagation and approximate ML estimation by pseudo-moment matching. In: Proc. of the Int’l Conf. on Artificial Intelligence and Statistics. 2003. 1−8.

[87] Minka TP. Expectation propagation for approximate Bayesian inference. In: Proc. of the Conf. on Uncertainty in Artificial Intelligence. 2001. 362−369.

[88] Johnson J, Malioutov D, Willsky A. Walk-Sum interpretation and analysis of Gaussian belief propagation. In: Proc. of the Advances in Neural Information Processing Systems. 2006. 1−8.

[89] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 1970,57(1):97−109. [doi: 10.1093/biomet/57.1.97]

[90] McAllester DA. Some PAC Bayesian theorems. In: Proc. of the Annual Conf. on Computational Learning Theory. 1998,51: 230−234.

[91] Seldin Y, Tishby N. PAC-Bayesian analysis of co-clustering and beyond. Journal of Machine Learning Research, 2010,11: 3595−3646.

[92] Sun S, Shawe-Taylor J, Mao L. PAC-Bayes analysis of multi-view learning. Information Fusion, 2017,35:117−131. [doi: 10.1016/j.inffus.2016.09.008]

[93] Seldin Y, Tishby N. A PAC-Bayesian approach to unsupervised learning with application to co-clustering analysis. Journal of Machine Learning Research, 2010,3:1−46. [doi: 10.1561/2200000016]

[94] Higgs M, Shawe-Taylor J. A PAC-Bayes bound for tailored density estimation. In: Proc. of the Algorithmic Learning Theory. 2010. 148−162. [doi: 10.1007/978-3-642-16108-7_15]

[95] Yarowsky D. Unsupervised word sense disambiguation rivaling supervised methods. In: Proc. of the Association for Computational Linguistics. 1995. 189−196. [doi: 10.3115/981658.981684]

[96] Xu C, Tao DC, Xu C. A survey on multi-view learning. arxiv: 1304.5634, 2013.

[97] Zhou ZH. Ensemble Methods: Foundations and Algorithms. Boca Raton: Chapman and Hall/CRC, 2012.

[98] Zhao J, Xie X, Xu X, Sun S. Multi-View learning overview: Recent progress and new challenges. Information Fusion, 2017,38: 43−54. [doi: 10.1016/j.inffus.2017.02.007]

[99] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans. on Knowledge and Data Engineering, 2010,22(10):1345−1359. [doi: 10.1109/TKDE.2009.191]

[100] Saxena S, Verbeek J. Heterogeneous face recognition with cnns. In: Proc. of the European Conf. on Computer Vision Workshop on Transferring and Adapting Source Knowledge in Computer Vision. 2016. 483−491. [doi: 10.1007/978-3-319-49409-8_40]

[101] Zhuang FZ, He Q, Shi ZZ. Survey on transfer learning research. Ruan Jian Xue Bao/Journal of Software, 2015,26(1):26−39 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4631.htm [doi: 10.13328/j.cnki.jos.004631]

[102] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: Proc. of the Workshop on Computational Learning Theory. 1998. 92−100. [doi: 10.1145/279943.279962]

[103] Nigam K, Ghani R. Analyzing the effectiveness and applicability of co-training. In: Proc. of the 9th Int'l Conf. on Information and Knowledge Management. 2000. 86−93. [doi: 10.1145/354756.354805]

[104] Yu S, Krishnapuram B, Rosales R, Rao RB. Bayesian co-training. Journal of Machine Learning Research, 2011,12:2649−2680.

[105] Kumar A, Daume H. A co-training approach for multi-view spectral clustering. In: Proc. of the Int'l Conf. on Machine Learning. 2011. 393−400.

[106] Zhou ZH, Li M. Tri-Training: Exploiting unlabeled data using three classifiers. IEEE Trans. on Knowledge and Data Engineering, 2005,17(11):1529−1541. [doi: 10.1109/TKDE.2005.186]

[107] Li M, Zhou ZH. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans, 2007,37(6):1088−1098. [doi: 10.1109/TSMCA.2007.904745]

[108] Sindhwani V, Rosenberg DS. An RKHS for multi-view learning and manifold co-regularization. In: Proc. of the 25th Int'l Conf. on Machine Learning. 2008. 976−983. [doi: 10.1145/1390156.1390279]

[109] Li G, Chang K, Hoi S. Multiview semi-supervised learning with consensus. IEEE Trans. on Knowledge and Data Engineering, 2012,24(11):2040−2051. [doi: 10.1109/TKDE.2011.160]

[110] Xu X, Li W, Xu D, Tsang IW. Co-Labeling for multi-view weakly labeled learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016,38(6):1113−1125. [doi: 10.1109/TPAMI.2015.2476813]

[111] Gönen M, Alpaydın E. Multiple kernel learning algorithms. Journal of Machine Learning Research, 2011,12:2211−2268.

[112] Lanckriet GR, Cristianini N, Bartlett P. Learning the kernel matrix with semidefinite programming. Journal of Machine Learning Research, 2004,5:27−72.

[113] Bach FR, Lanckriet GR, Jordan MI. Multiple kernel learning, conic duality, and the SMO algorithm. In: Proc. of the 21st Int'l Conf. on Machine Learning. 2004. 1−8. [doi: 10.1145/1015330.1015424]

[114] Sonnenburg S, Rätsch G, Schäfer C. Large scale multiple kernel learning. Journal of Machine Learning Research, 2006,7: 1531−1565.

[115] Rakotomamonjy A, Bach F, Canu S, Grandvalet Y. SimpleMKL. Journal of Machine Learning Research, 2008,9:2491−2521.

[116] Subrahmanya N, Shin YC. Sparse multiple kernel learning for signal processing applications. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010,32(5):788−798. [doi: 10.1109/TPAMI.2009.98]

[117] Hotelling H. Relations between two sets of varieties. Biometrika, 1936,28(3-4):321−377. [doi: 10.1093/biomet/28.3-4.321]

[118] Akaho S. A kernel method for canonical correlation analysis. arXiv preprint cs/0609071, 2006.

[119] Chaudhuri K, Kakade SM, Livescu K. Multi-View clustering via canonical correlation analysis. In: Proc. of the 26th Annual Int'l Conf. on Machine Learning. ACM Press, 2009. 129−136. [doi: 10.1145/1553374.1553391]

[120] Kakade SM, Foster DP. Multi-View regression via canonical correlation analysis. In: Proc. of the Learning Theory. 2007. 82−96. [doi: 10.1007/978-3-540-72927-3_8]

[121] Hardoon DR, Shawe-Taylor J. Convergence analysis of kernel canonical correlation analysis: Theory and practice. Machine Learning, 2009,74(1):23−38. [doi: 10.1007/s10994-008-5085-3]

[122] Cai J, Sun HW. Convergence rate of kernel canonical correlation analysis. In: Proc. of the Science China Mathematics. 2011. 1−10. [doi: 10.1007/s11425-011-4245-2]

[123] Blei DM, Jordan MI. Modeling annotated data. In: Proc. of the 26th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. 2003. 127−134. [doi: 10.1145/860435.860460]

[124] Diethe T, Hardoon DR, Shawe-Taylor J. Multiview fisher discriminant analysis. In: Proc. of the Advances in Neural Information Processing Systems Workshop on Learning from Multiple Sources. 2008. 1−8.

[125] Li SY, Jiang Y, Zhou ZH. Partial multi-view clustering. In: Proc. of the Association for the Advancement of Artificial Intelligence. 2014. 1968−1974.

[126] Liu J, Jiang Y, Li Z, Zhou ZH, Lu H. Partially shared latent factor learning with multiview data. IEEE Trans. on Neural Networks and Learning Systems, 2014,26(6):1233−1246. [doi: 10.1109/TNNLS.2014.2335234]

[127] Farquhar J. Two view learning: SVM-2K, theory and practice. In: Proc. of the Advances in Neural Information Processing Systems. 2005. 355−362.

[128] Yin Q, Wu S, Wang L. Unified subspace learning for incomplete and unlabeled multi-view data. Pattern Recognition, 2017,67: 313−327. [doi: 10.1016/j.patcog.2017.01.035]

[129] Xu C, Tao D, Xu C. Large-Margin multi-view information bottleneck. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2014,36(8):1558−1572. [doi: 10.1109/TPAMI.2013.2296528]

[130] Xu C, Tao D, Xu C. Multi-View intact space learning. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(12): 2531−2544. [doi: 10.1109/TPAMI.2015.2417578]

[131] Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY. Multimodal deep learning. In: Proc. of the 28th Int'l Conf. on Machine Learning. 2011. 689−696.

[132] Srivastava N, Salakhutdinov R. Multimodal learning with deep Boltzmann machines. In: Proc. of the Advances in Neural Information Processing Systems. 2012. 2222−2230.

[133] Srivastava N, Salakhutdinov R. Learning representations for multimodal data with deep belief nets. In: Proc. of the Int'l Conf. on Machine Learning Workshop on Representation Learning. 2012. 1−8.

[134] Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-View convolutional neural networks for 3D shape recognition. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 945−953. [doi: 10.1109/ICCV.2015.114]

[135] Elhoseiny M, El-Gaaly T, Bakry A, Elgammal A. A comparative analysis and study of multiview CNN models for joint object categorization and pose estimation. In: Proc. of the 33rd Int'l Conf. on Machine Learning. 2016. 888−897.

[136] Andrew G, Arora R, Bilmes J, Livescu K. Deep canonical correlation analysis. In: Proc. of the Int'l Conf. on Machine Learning. 2013. 1247−1255.

[137] Benton A, Khayrallah H, Gujral B, Reisinger D, Zhang S, Arora R. Deep generalized canonical correlation analysis. arXiv preprint arXiv:1702.02519, 2017.

[138] Wang W, Arora R, Livescu K, Bilmes J. On deep multi-view representation learning: Objectives and optimization. arXiv preprint arXiv: 1602.01024, 2016.

[139] Yang X, Ramesh P, Chitta R, Madhvanath S, Bernal E A, Luo J. Deep multimodal representation learning from temporal data. arXiv preprint arXiv:1704.03152, 2017.

[140] Du J, Ling CX, Zhou ZH. When does co-training work in real data? IEEE Trans. on Knowledge and Data Engineering, 2011,23(5): 788−799. [doi: 10.1109/TKDE.2010.158]

[141] Dasgupta S, Littman M, McAllester D. PAC generalization bounds for co-training. In: Proc. of the Neural Information Processing Systems. 2001. 375−382.

[142] Abney S. Bootstrapping. In: Proc. of the 40th Annual Meeting on Association for Computational Linguistics. 2002. 360−367.

[143] Balcan M, Blum A, Yang K. Co-Training and expansion: Towards bridging theory and practice. In: Proc. of the Advances in Neural Information Processing Systems. 2004. 89−96.

[144] Wang W, Zhou ZH. A new analysis of co-training. In: Proc. of the Int'l Conf. on Machine Learning. 2010. 1135−1142.

[145] Sindhwani V, Niyogi P, Belkin M. A co-regularization approach to semi-supervised learning with multiple views. In: Proc. of the Int'l Conf. on Machine Learning Workshop on Learning with Multiple Views. 2005. 74−79.

[146] Hussain Z, Shawe-Taylor J. Improved loss bounds for multiple kernel learning. In: Proc. of the Int'l Conf. on Artificial Intelligence and Statistics. 2011. 370−377.

[147] Ying Y, Campbell C. Generalization bounds for learning the kernel. In: Proc. of the 22nd Annual Conf. on Learning Theory. 2009. 1−9.

[148] Kloft M, Blanchard G. The local rademacher complexity of l_p-norm multiple kernel learning. Arxiv preprint arXiv: 1103.0790, 2011.

[149] Woodworth RS, Thorndike EL. The influence of improvement in one mental function upon the efficiency of other functions. Psychological Review, 1901,8(3):247−261.

[150] Long MS. Transferlearing: Problems and methods [Ph.D. Thesis]. Beijing: Tsinghua University, 2014 (in Chinese).

[151] http://socrates.acadiau.ca/courses/comp/dsilver/NIPS95_LTL/transfer.workshop.1995.html

[152] Xue GR, Dai WY, Yang Q, Yu Y. Topic-Bridged PLSA for cross-domain text classification. In: Proc. of the 31st Int’l ACM SIGIR Conf. on Research and Development on Information Retrieval. 2008. 627−634. [doi: 10.1145/1390334.1390441]

[153] Ling X, Dai WY, Xue GR, Yang Q, Yu Y. Spectral domain-transfer learning. In: Proc. of the 14th ACM SIGKDD Int’l Conf. on Knowledge Discovery and Data Mining. 2008. 488−496. [doi: 10.1145/1401890.1401951]

[154] Pan SJ, Ni X, Sun J, Yang Q, Chen Z. Cross-Domain sentiment classification via spectral feature alignment. In: Proc. of the 19th Int’l Conf. on World Wide Web. 2010. 751−760. [doi: 10.1145/1772690.1772767]

[155] Gao XB, Wang XM, Li XL, Tao D. Transfer latent variable model based on divergence analysis. Pattern Recognition, 2011, 44(10-11):2358−2366. [doi: 10.1016/j.patcog.2010.06.013]

[156] Dai WY, Yang Q, Xue GR, Yu Y. Self-Taught clustering. In: Proc. of the Int’l Conf. on Machine Learning. 2008. 200−207. [doi: 10.1145/1390156.1390182]

[157] Dai WY, Yang Q, Xue GR, Yu Y. Boosting for transfer learning. In: Proc. of the Int’l Conf. on Machine Learning. 2007. 193−200. [doi: 10.1145/1273496.1273521]

[158] Dai WY, Chen Y, Xue GR, Yang Q, Yu Y. Translated learning: Transfer learning across different feature spaces. In: Proc. of the Advances in Neural Information Processing System. 2008. 353−360.

[159] Yang Q, Chen Y, Xue GR, Dai W, Yu Y. Heterogeneous transfer learning for image clustering via the social Web. In: Proc. of the Joint Conf. of the Annual Meeting of the ACL and the Int’l Joint Conf. on Natural Language Processing. 2009. 1−9.

[160] Zhuang ZF, Luo P, Shen ZY, He Q, Xiong Y, Shi Z, Xiong H. Collaborative dual-PLSA: Mining distinction and commonality across multiple domains for Classification. In: Proc. of the Int’l Conf. on Information and Knowledge Management. 2010. 359−368.

[161] Yang L, Jing LP, Ng KM. Robust and non-negative collective matrix factorization for text-to-image transfer learning. IEEE Trans. on Image Processing, 2015,24(12):4701−4714. [doi: 10.1109/TIP.2015.2465157]

[162] Yang L, Hanneke S, Carbonell J. A theory of transfer learning with applications to active learning. Machine Learning, 2013,90(2): 161−189. [doi: 10.1007/s10994-012-5310-y]

[163] Jiang J, Zhai CX. Instance weighting for domain adaptation in NLP. In: Proc. of the 45th Annual Meeting of the Association for Computational Linguistics. 2007. 264−271.

[164] Wang C, Mahadevan S. Manifold alignment using procrustes analysis. In: Proc. of the Int’l Conf. on Machine Learning. 2008. 1120−1127. [doi: 10.1145/1390156.1390297]

[165] Chang WC, Wu Y, Liu H, Yang Y. Cross-Domain kernel induction for transfer learning. In: Proc. of the Association for the Advancement of Artificial Intelligence. 2017. 1−7.

[166] Segev N, Harel M, Mannor S, Crammer K, El-Yaniv R. Learn on source, refine on target: A model transfer learning framework with random forests. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016,39(9):1811−1824. [doi: 10.1109/TPAMI. 2016.2618118]

[167] Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning. In: Proc. of the Int’l Conf. on Empirical Methods in Natural Language Processing. 2006. 120−128.

[168] Xing DK, Dai WY, Xue GR, Yu Y. Bridged refinement for transfer learning. In: Proc. of the 11th European Conf. on Practice of Knowledge Discovery in Databases. 2007. 324−335. [doi: 10.1007/978-3-540-74976-9_31]

[169] Blitzer J, Crammer K, Kulesza A, Pereira F, Wortman J. Learning bounds for domain adaptation. In: Proc. of the Advances in Neural Information Processing Systems. 2008. 129−136.

[170] Zen G, Porzi L, Sangineto E, Ricci E, Sebe N. Learning personalized models for facial expression analysis and gesture recognition. IEEE Trans. on Multimedia, 2016,18(4):775−788. [doi: 10.1109/TMM.2016.2523421]

[171] Wang Z, Song YQ, Zhang CS. Transferred dimensionality reduction. In: Proc. of the European Conf. on Machine Learning and Knowledge Discovery in Databases. 2008. 550−565. [doi: 10.1007/978-3-540-87481-2_36]

[172] Jiang W, Chung F. Transfer spectral clustering. In: Proc. of the 2012 European Conf. on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. 2012. 789−803. [doi: 10.1007/978-3-642-33486-3_50]

[173] Zhang L, Zuo ZW, Zhang D. LSDT: Latent sparse domain transfer learning for visual adaptation. IEEE Trans. on Image Processing, 2016,25(3):1177−1191. [doi: 10.1109/TIP.2016.2516952]

[174] Lawrence ND, Platt JC. Learning to learn with the informative vector machine. In: Proc. of the 21st Int'l Conf. on Machine Learning. 2004. 1−8. [doi: 10.1145/1015330.1015382]

[175] Bonilla E, Chai KM, Williams C. Multi-Task Gaussian process prediction. In: Proc. of the 20th Advances in Neural Information Processing Systems. 2008. 153−160.

[176] Schwaighofer A, Tresp V, Yu K. Learning Gaussian process kernels via hierarchical Bayes. In: Proc. of the 17th Advances in Neural Information Processing Systems. 2005. 1209−1216.

[177] Evgeniou T, Pontil M. Regularized multi-task learning. In: Proc. of the 10th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. 2004. 109−117. [doi: 10.1145/1014052.1014067]

[178] Mihalkova L, Huynh T, Mooney RJ. Mapping and revising Markov logic networks for transfer learning. In: Proc. of the 22nd Association for the Advancement of Artificial Intelligence. 2007. 608−614.

[179] Mihalkova L, Mooney RJ. Transfer learning by mapping with minimal target data. In: Proc. of the Association for the Advancement of Artificial Intelligence Workshop Transfer Learning for Complex Tasks. 2008. 31−36.

[180] Davis J, Domingos P. Deep transfer via second-order Markov logic. In: Proc. of the 26th Int'l Conf. on Machine Learning. 2009. 217−224. [doi: 10.1145/1553374.1553402]

[181] Bengio Y. Deep learning of representations for unsupervised and transfer learning. In: Proc. of the Int'l Conf. on Machine Learning Workshop on Unsupervised and Transfer Learning. 2012. 17−36.

[182] Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: A deep learning approach. In: Proc. of the 28th Int'l Conf. on Machine Learning. 2011. 513−520.

[183] Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 1717−1724. [doi: 10.1109/CVPR.2014.222]

[184] Chen WY, Hsu TM, Tsai YH. Transfer neural trees for heterogeneous domain adaptation. In: Proc. of the European Conf. on Computer Vision. 2016. 399−414. [doi: 10.1007/978-3-319-46454-1_25]

[185] Wang L, Li Y, Lazebnik S. Learning deep structure-preserving image-text embeddings. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 5005−5013. [doi: 10.1109/CVPR.2016.541]

[186] Long MS, Wang JM, Jordan MI. Deep transfer learning with joint adaptation networks. CoRR, vol. arXiv:1605.06636, 2016.

[187] Sun B, Saenko K. Deep coral: Correlation alignment for deep domain adaptation. In: Proc. of the European Conf. on Computer Vision Workshop on Transferring and Adapting Source Knowledge in Computer Vision. 2016. 443−450. [doi: 10.1007/978-3-319-49409-8_35]

[188] Rozantsev A, Salzmann M, Fua P. Beyond sharing weights for deep domain adaptation. CoRR, vol. arXiv:1603.06432, 2016.

[189] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Proc. of the Advances in Neural Information Processing Systems. 2014. 3320−3328.

[190] Zhuang F, Cheng X, Luo P, Pan SJ, He Q. Supervised representation learning: Transfer learning with deep autoencoders. In: Proc. of the 24th Int'l Conf. on Artificial Intelligence. 2015. 4119−4125.

[191] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. Journal of Big Data, 2016,3(1):1−40. [doi: 10.1186/s40537-016-0043-6]

[192] Ganin Y, Ustinova E, Ajakan H, Germain P, Larochelle H, Laviolette F, MarchandM, Lempitsky V. Domain adversarial training of neural networks. Journal of Machine Learning Research, 2016,17(59):1−35.

[193] Liu MY, Tuzel O. Coupled generative adversarial networks. In: Proc. of the Advances in Neural Information Processing Systems. 2016. 469–477.

[194] Ben-David S, Blitzer J, Crammer K, Pereira F. Analysis of representations for domain adaptation. In: Proc. of the Advances in Neural Information Processing Systems. 2007. 137–144.

[195] Kifer D, Ben-David S, Gehrke J. Detecting change in data streams. In: Proc. of the 30th Int'l Conf. on Very Large Data Bases. 2004. 180–191.

[196] Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Vaughan JW. A theory of learning from different domains. Machine Learning, 2010,79(1-2):151–175. [doi: 10.1007/s10994-009-5152-4]

[197] Mansour Y, Mohri M, Rostamizadeh A. Multiple source adaptation and the Rényi divergence. In: Proc. of the 25th Conf. on Uncertainty in Artificial Intelligence. 2009. 367–374.

[198] Mansour Y, Mohri M, Rostamizadeh A. Domain adaptation: Learning bounds and algorithms. In: Proc. of the 22nd Annual Conf. on Learning Theory. 2009.

[199] Zhang C, Zhang L, Fan W, Ye J. Generalization bounds for representative domain adaptation. arXiv preprint arXiv: 1401.0376, 2014.

[200] Kumagai W. Learning bound for parameter transfer learning. In: Proc. of Advances in Neural Information Processing Systems. 2016. 2721–2729.

[201] Zhou ZH. Pervasive machine learning. In: Proc. of Department of Information Science of National Natural Science Foundation of China on Major Issues in Basic Theory of Intelligent Science and Technology. 2004 (in Chinese with English abstract).

[202] Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 2000. [doi: 10.1007/978-1-4757-3264-1]

附中文参考文献:

[65] 吴新星,张军平.Rademacher 复杂度在统计学习理论中的研究:综述.自动化学报,2017,43(1):20–39.

[101] 庄福振,何清,史忠植.迁移学习研究进展.软件学报,2015,26(1):26–39. http://www.jos.org.cn/1000-9825/4631.htm [doi: 10.13328/j.cnki.jos.004631]

[150] 龙明盛.迁移学习问题与方法研究[博士学位论文].北京:清华大学,2014.

[201] 周志华.普适机器学习.见:国家自然科学基金委员会信息科学部"智能科学技术基础理论重大问题"高层研讨会.2004.

杨柳(1980－),女,河北保定人,博士,副教授,主要研究领域为机器学习,数据挖掘.

刘烨(1979－),女,博士,副研究员,CCF 专业会员,主要研究领域为认知心理学,情感计算.

于剑(1969－),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为机器学习,数据挖掘.

詹德川(1982－),男,博士,副教授,CCF 专业会员,主要研究领域为机器学习,数据挖掘.