

# 多任务学习的研究

张敏灵

(东南大学计算机科学与工程学院, 南京 210096)

- 5 **摘要:** 多任务学习起源于机器学习领域对归纳偏置问题的研究, 该学习范式通过对多个相关任务同时进行学习, 利用任务之间传递的有用信息, 以获取更优的归纳偏置用于假设空间搜索。由于多任务学习在提高学习系统泛化能力方面具有显著的能力, 该课题已成为目前国际机器学习界的一个研究热点。本文将对多任务学习的研究现状做一个简要介绍, 并对未来的研究工作进行展望。
- 10 **关键词:** 机器学习; 多任务学习; 懒惰学习; 多示例学习; 多标记学习
- 中图分类号:** TP181

## Research on Multi-Task Learning

ZHANG Minling

- 15 (School of Computer Science and Engineering, Southeast University, NanJing 210096)

**Abstract:** Multi-task learning originates from the research of the inductive bias problem in machine learning. In this learning framework, several related tasks are learned simultaneously. In this way, the system attempts to enhance its generalization ability by exchanging useful information among different tasks to achieve better inductive bias for hypothesis space exploration. Due to its significant ability in improving the learning system's generalization ability, multi-task learning has become one of the hottest research topics in machine learning community. In this paper, we will briefly review related works on multi-task learning and indicates several issues for future work.

- 20 **Keywords:** machine learning; multi-task learning; lazy learning; multi-instance learning; multi-label learning
- 25

## 0 引言

- 国际机器学习界的权威 T. M. Mitchell 认为, 机器学习就是对一类特殊计算机算法的研究, 这类算法可以从以往的经验中不断学习, 从而提升其在处理特定“任务(task)”时的性能<sup>[1]</sup>。一般情况下, 学习系统根据从导师或者环境中获取的训练示例进行学习, 以尽可能正确地
- 30 对训练集之外的示例进行预测或其它方式的处理。此时, 与任务相关的经验蕴涵在训练示例中, 而学习系统的目标就是提高其处理未知样本时的性能, 即泛化能力。传统的机器学习技术主要针对单任务学习(single-task learning)问题, 此时训练集中的所有示例均反映了单个任务的信息。在实际应用中, 由于受到时间、人力、经济等多种因素的制约, 人们往往只能获得与某个任务相关的有限甚至少量示例。因此, 在遇到训练样本不充分的情况时, 学习
- 35 系统将很难获取足够的信息进行学习以得到具有强泛化能力的模型。

- 虽然目前大多数机器学习技术主要面对的是单个任务的学习场景, 但在真实世界中多个相关的学习任务往往是同时出现的。例如, 当一名守门员在学习如何扑救时, 并不是单纯地学习如何去扑球, 而是会涉及许多相关的学习内容, 比如如何进行预判, 如何移动脚步, 如何
- 40 腾身起跳, 如何平稳落地等。实际上, 人类在学习如何完成一项任务时, 往往都需要对一些与该任务相关的任务进行学习并从中汲取经验, 从而起到举一反三与融会贯通的作用。人类的学习过程如此, 计算机在解决实际问题时的情形也十分相似。例如, 在计算机辅助医疗诊断中, 系统在根据病人的症状判断一个人是否患有某种疾病时, 如果还能利用该病人的其

基金项目: 教育部博士点基金新教师项目(200802941009)

作者简介: 张敏灵, (1979-), 男, 副教授, 主要研究方向: 机器学习、数据挖掘. E-mail: zhangml.seu@gmail.com

它一些健康方面的信息(例如相关的并发症等),将有助于学习问题的解决<sup>[2,3]</sup>;在个性化软件设计中,同时学习多个用户在进行某项操作(如过滤垃圾邮件)时各自的行为习惯,将有助于发现用户之间的共性以及个性特征<sup>[4]</sup>。

由此可见,解决真实世界的问题往往需要面对多个相关的学习任务,而如何利用多个相关任务之间蕴涵的有价值的信息来提高学习系统的性能,正是“多任务学习(multi-task learning)”<sup>[2,3]</sup>所面临的核心问题。本文接下来将对多任务学习研究的起源和研究现状进行简要介绍,然后对该研究领域进一步的研究工作进行展望。

## 1 研究进展

多任务学习早期的研究工作源于对机器学习中的一个重要问题,即“归纳偏置(inductive bias)”问题的研究。机器学习的过程可以看作是对与问题相关的经验数据进行分析,从中归纳出反映问题本质的模型的过程。归纳偏置的作用就是用于指导学习算法如何在模型空间中进行搜索,搜索所得模型的性能优劣将直接受到归纳偏置的影响,而任何一个缺乏归纳偏置的学习系统都不可能进行有效的学习<sup>[5]</sup>。不同的学习算法(如决策树,神经网络,支持向量机等)具有不同的归纳偏置,人们在解决实际问题时需要人工地确定采用何种学习算法,实际上也就是主观地选择了不同的归纳偏置策略。一个很直观的想法就是,是否可以将归纳偏置的确定过程也通过学习过程来自动地完成,也就是采用“学习如何去学(learning to learn)”<sup>[6]</sup>的思想。多任务学习恰恰为上述思想的实现提供了一条可行途径,即利用相关任务中所包含的有用信息,为所关注任务的学习提供更强的归纳偏置。

受上述思想的启发,R. Caruana 对多任务学习的相关问题,如什么是多任务学习,该学习范式是否可行、为什么可行、在什么情况下可行等进行了初步的分析。1997 年,其主要研究成果发表于国际机器学习界的权威刊物《Machine Learning》,标志着多任务学习这一机器学习概念的正式提出<sup>[2]</sup>。R. Caruana 解决多任务学习问题的方法其实比较简单,他首先构造一个前馈神经网络,该网络的隐层结点为所有任务共享而每个任务则对应于一个输出结点。在神经网络的训练过程完成后,输入结点到隐层结点的连接权中即包含了所有任务之间的共享信息,而各个任务自身的特定信息则位于隐层结点到各输出结点的连接权中。虽然 R. Caruana 所采用的多任务学习方法并不复杂,但是却在自动驾驶、机器人目标识别以及辅助医疗诊断等领域得到了成功应用<sup>[2,3]</sup>。类似的基于神经网络的方法还出现在其他一些研究者的工作中<sup>[7,8]</sup>。

作为一种新的机器学习范式,多任务学习引起了机器学习理论研究者的关注。J. Baxter<sup>[9]</sup>对传统的 VC 维<sup>[10]</sup>概念进行了扩展,定义了扩展 VC 维用于度量多个假设空间构成的“簇”的复杂度。基于扩展 VC 维的概念,J. Baxter 给出了一个学习系统在对  $T$  个学习任务进行学习时,所得模型在  $T$  个任务上的平均错误率上界,并且证明该上界将随着任务数目  $T$  的增加而不断减小。此外,J. Baxter<sup>[11]</sup>还从信息论和贝叶斯学习的角度出发,分析了从  $T$  个任务进行有效的学习时各个任务所需提供的信息量。同样基于扩展 VC 维的概念,S. Ben-David 与 R. Schuller<sup>[12]</sup>扩展了 J. Baxter<sup>[9]</sup>关于多个任务平均错误率上界的分析工作,得出了一个关于单个任务的更紧的错误率上界。这些研究成果的取得对多任务学习为什么有效给出了一个理论上的初步解释。

除了上述的理论分析工作之外,机器学习的研究者们还对多任务学习进行了大量的应用方法研究,主要包括基于正则化的方法以及基于层次贝叶斯模型的方法。T. Evgeniou 等人<sup>[13]</sup>基于他们早期在多任务学习方面的工作<sup>[14,15]</sup>,提出了一种多任务学习的正则化框架,并且基

于针对向量输出函数的核方法<sup>[16]</sup>将多任务学习的正则化问题转为传统的单任务学习问题求解。R. Ando 与 T. Zhang<sup>[17]</sup>考察了如何利用多个相关任务来进行结构学习,以找出多个任务的共享结构表示。与此同时他们还给出了如何利用未标记数据来生成结构学习所需的辅助任务,为半监督学习问题的解决提供了一条新途径。B. Bakker 与 T. Heskes<sup>[18]</sup>对他们的初期工作<sup>[19,20]</sup>进行了扩展,将 R. Caruana 关于多任务神经网络学习的工作与统计领域内的“多层分析”技术<sup>[21]</sup>相结合,给出了一种基于多层贝叶斯结构的多任务学习方法。该方法中各个任务不仅共享输入层与隐层之间的连接权,与各任务特定的隐层与输出层之间的连接权也通过一个共同的先验概率模型实现松散的关联。除此之外,研究者们还相继提出了多种利用随机过程来对多任务学习进行建模的方法,例如使用高斯过程<sup>[22-24]</sup>, Markov 过程<sup>[25-27]</sup>, Dirichlet 过程<sup>[28,29]</sup>, 以及  $t$  过程<sup>[30]</sup>等。

除了上述基于正则化与层次贝叶斯模型的方法,还出现一些基于 Logistic 回归<sup>[31]</sup>、径向基函数网络<sup>[32]</sup>、支持向量机<sup>[33]</sup>以及独立成分分析<sup>[34]</sup>等技术的多任务学习方法。此外,研究者们还对多任务学习范式下的属性选择问题进行了研究,此时学习系统的目标是为多个任务寻找一个共享的输入属性空间,从而基于转化后的表示空间进行学习以得到泛化能力更强的预测模型<sup>[35-37]</sup>。目前,多任务学习技术已在模式识别<sup>[3][28][31]</sup>、辅助医疗诊断<sup>[2,3]</sup>、数据挖掘<sup>[18][20]</sup>、软件设计<sup>[23][38]</sup>、语音处理<sup>[39]</sup>等多个领域中得到了成功应用。显然,多任务学习不仅具有较高的理论研究价值,还具有广阔的应用前景。近期,国际机器学习界的权威会议 ICML'05, NIPS'05 以及 ECML'07 分别设置了主题为“Meta-Learning”, “Inductive Transfer: 10 Years Later”以及“Planning to Learn”的 Workshop,内容均涉及多任务学习。仅在 2007 年,在权威会议 ICML'07 以及 NIPS'07 上就出现了 8 篇与多任务学习相关的研究论文,占到了相当可观的比重。如此活跃的国际学术交流现状表明,多任务学习已成为目前国际机器学习界的一大研究热点,并逐渐迎来了其发展的一个高峰阶段。

## 2 进一步的工作

上一节对多任务学习的研究现状作了一个简要介绍。结合多任务学习的研究现状以及作者自身的研究基础,本节将对多任务学习有待进一步研究的问题进行讨论。

R. Caruana<sup>[2,3]</sup>在提出多任务学习范式时,主要采用了多层前馈神经网络来实现任务之间的信息传递。但他特别指出,多任务学习决不仅仅归入一种单一的算法,而应该是问题求解思想,求解技术以及求解算法的有机结合体。然而,作为一类重要的机器学习技术,目前尚未出现基于懒惰学习(lazy learning)技术<sup>[40]</sup>的多任务学习算法。与目前多任务学习方法大多采用的“积极学习(eager learning)”技术不同,懒惰学习技术不需要预先进行模型训练,而是在获得待处理样本后进行实时处理,特别适合增量学习的需要。设计出基于懒惰学习技术的多任务学习算法,在理论与应用两方面都有较高的价值。

已有研究成果表明,在许多情况下多任务学习范式的确比传统的单任务学习范式具有更强的问题求解能力。我们认为,多任务学习之所以在很多情况下有效,在真实世界对象中普遍存在的歧义性(ambiguity)<sup>[41]</sup>也许是根本原因之一。这一点从直观上是很容易理解的:由于对象具有歧义性,因此学习系统无法准确地获取对象的语义信息,这些不准确的输入信息将对基于归纳偏置的假设空间搜索过程带来很大的负面影响。然而,考虑到多任务学习在获取强归纳偏置上的独特能力<sup>[2,3]</sup>,在需要对歧义性对象进行分析与建模时,学习系统获得的强归纳偏置将在很大程度上弥补不准确的输入信息造成的不利影响。因此,从对象歧义性的角度出发对多任务学习进行研究,不仅有望得到泛化能力更强的学习算法,还可能为探寻多任

务学习有效的根本原因提供线索。目前,国际机器学习界存在两种流行的歧义性对象学习框架,分别为多示例学习(multi-instance learning)<sup>[42]</sup>与多标记学习(multi-label learning)<sup>[43]</sup>。

多示例学习框架是1997年T. G. Dietterich等人<sup>[42]</sup>基于药物活性预测问题的研究所提出的。在该学习框架下,与传统的监督学习框架相比,虽然每个对象对应的概念标记是唯一的,但其不再采用单一的示例来表示,而是采用多个示例构成的“包(bag)”的表示形式。因此,多示例学习可以看作是从输入空间,也就是对象的内容表示上来处理歧义性。已有的研究工作表明<sup>[44]</sup>,对于歧义性对象而言,采用包的表示形式比采用单个示例的表示形式要显得更加自然。因此,对于多任务学习而言,也许我们可以利用多示例学习技术来分析各个学习任务在输入空间上可能存在的歧义性,通过合适的方式转换歧义性对象的内容表示,以更好地进行问题求解。因此,设计基于多示例学习技术的多任务学习算法是一个值得进一步研究的问题。

多标记学习的研究起源于文档分类研究中遇到的歧义性问题<sup>[43]</sup>。在该学习框架下,与传统的监督学习框架相比,虽然每个对象仍然采用单一的示例进行表示,但其不再对应于一个概念标记,而是同时对应于多个概念标记。因此,多标记学习可以认为是从输出空间,也就是概念标记上来处理对象的歧义性。如果我们将学习系统对每个标记进行预测作为一个单独的任务,那么就可以在某种程度上把多标记学习看作一类特殊的多任务学习问题,此时所有的任务共享相同的示例表示。因此,多标记学习领域内已有的研究成果对于多任务学习的研究者来说是十分具有借鉴意义的。因此,研制出结合了多标记学习和多任务学习技术的学习算法,将有助于我们增加对上述两类技术本质的了解,从而进一步促进这两类学习技术的发展。

## [参考文献] (References)

- [1] Mitchell T M. Machine learning. New York: McGraw-Hill, 1997.
- [2] Caruana R. Multitask learning. Machine Learning, 1997, 28(1): 41-75.
- [3] Caruana R. Multitask learning. PhD thesis, School of Computer Science, Carnegie Mellon University, 1997.
- [4] Zhang J. A probabilistic framework for multi-task learning. PhD thesis, School of Computer Science, Carnegie Mellon University, 2006.
- [5] Mitchell T M. The need for biases in learning generalizations. In: Shavlik J W, Dietterich T G, eds. Readings in Machine Learning, San Mateo, CA: Morgan Kaufmann, 1990, 184-191.
- [6] Thrun S, Pratt L. Learning to learn. Norwell, MA: Kluwer Academic Publishers, 1998.
- [7] Thrun S. Is learning the n-th thing any easier than learning the first? In: Touretzky D S, Mozer M C, Hasselmo M E, eds. Advances in Neural Information Processing Systems 8, Cambridge, MA: MIT Press, 1996, 640-646.
- [8] Silver D, Mercer R. Selective functional transfer: inductive bias from related tasks. In: Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing, Cancun, Mexico, 2001, 182-189.
- [9] Baxter J. A model of inductive bias learning. Journal of Artificial Intelligence Research, 2000, 12: 149-198.
- [10] Vapnik V N. Statistical learning theory. New York: Wiley, 1998.
- [11] Baxter J. A bayesian/information theoretic model of learning to learn via multiple task sampling. Machine Learning, 1997, 28(1): 7-39.
- [12] Ben-David S, Schuller R. Exploiting task relatedness for multiple task learning. In: Proceedings of the 16th Annual Conference on Learning Theory, Washington, DC, 2003, 567-580.
- [13] Evgeniou T, Micchelli C A, Pontil M. Learning multiple tasks with kernel methods. Journal of Machine Learning Research, 2005, 6(Apr): 615-637.
- [14] Evgeniou T, Pontil M. Regularized multi-task learning. In: Proceedings of the 10th Conference on Knowledge Discovery and Data Mining, Seattle, WA, 2004, 109-117.
- [15] Micchelli C A, Pontil M. Kernels for multi-task learning. In: Saul L K, Weiss Y, Bottou L, eds. Advances in Neural Information Processing Systems 17, Cambridge, MA: MIT Press, 2005, 921-928.
- [16] Micchelli C A, Pontil M. On learning vector-valued functions. Neural Computation, 2005, 17(1): 177-204.
- [17] Ando R, Zhang T. A framework for learning predictive structures from multiple tasks and unlabeled data. Journal of Machine Learning Research, 2005, 6(Nov): 1817-1853.
- [18] Bakker B, Heskes T. Task clustering and gating for bayesian multitask learning. Journal of Machine



- Learning Research, 2003, 4(May): 83-99.
- [19] Heskes T. Solving a huge number of similar tasks: a combination of multi-task learning and hierarchical bayesian modeling. In: Proceedings of the 15th International Conference on Machine Learning, Madison, WI, 1998, 233-241.
- 180 [20] Heskes T. Empirical bayes for learning to learn. In: Proceedings of the 17th International Conference on Machine Learning, Palo Alto, CA, 2000, 367-374.
- [21] Bryk S, Raudenbush W. Hierarchical linear models: applications and data analysis methods. Newbury Park: Sage Publications, 1992.
- 185 [22] Yu K, Tresp V, Schwaighofer A. Learning Gaussian processes from multiple tasks. In: Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 2005, 1012-1019.
- [23] Bonilla E V, Agakov F V, Williams C K I. Kernel multi-task learning using task-specific features. In: Proceedings of the 11th International Conference on Artificial Intelligence and Statistics, San Juan, Puerto Rico, 2007.
- 190 [24] Bonilla E V, Chai K M, Willimans C K I. Multi-task Gaussian process prediction. In: Platt J C, Koller D, Singer Y, Roweis S, eds. Advances in Neural Information Processing Systems 20, Cambridge, MA: MIT Press, 2008.
- [25] Ni K, Carin L, Dunson D. Multi-task learning for sequential data via iHMMs and the nested dirichlet process. In: Proceedings of the 24th International Conference on Machine Learning, Corvalis, Oregon, 2007, 689-696.
- 195 [26] Wilson A, Fern A, Ray S, Tadepalli P. Multi-task reinforcement learning: a hierarchical bayesian approach. In: Proceedings of the 24th International Conference on Machine Learning, Corvalis, Oregon, 2007, 1015-1022.
- [27] Liu Q, Liao X, Carin L. Semi-supervised multitask learning. In: Platt J C, Koller D, Singer Y, Roweis S, eds. Advances in Neural Information Processing Systems 20, Cambridge, MA: MIT Press, 2008.
- 200 [28] Xue Y, Dunson D, Carin L. The matrix stick-breaking process for flexible multi-task learning. In: Proceedings of the 24th International Conference on Machine Learning, Corvalis, Oregon, 2007, 1063-1070.
- [29] Xue Y, Liao X, Carin L, Krishnapuram B. Multi-task learning for classification with dirichlet process priors. Journal of Machine Learning Research, 2007, 8(Jan): 35-63.
- 205 [30] Yu S, Tresp V, Yu K. Robust multi-task learning with t-processes. In: Proceedings of the 24th International Conference on Machine Learning, Corvalis, Oregon, 2007, 1103-1110.
- [31] Lapedriza À, Masip D, Vitrià J. A hierarchical approach for multi-task logistic regression. In: Martí J, Benedí J M, Mendonça A M, Serrat J, eds. Lecture Notes in Computer Science 4478, Berlin: Springer, 2007, 258-265.
- 210 [32] Liao X, Carin L. Radial basis function network for multi-task learning. In: Weiss Y, Schölkopf B, Platt J, eds. Advances in Neural Information Processing Systems 18, Cambridge, MA: MIT Press, 2006, 795-802.
- [33] Kato T, Kashima H, Sugiyama M, Asai K. Multi-task learning via conic programming. In: Platt J C, Koller D, Singer Y, Roweis S, eds. Advances in Neural Information Processing Systems 20, Cambridge, MA: MIT Press, 2008.
- 215 [34] Zhang J, Ghahramani Z, Yang Y. Learning multiple related tasks using latent independent component analysis. In: Weiss Y, Schölkopf B, Platt J, eds. Advances in Neural Information Processing Systems 18, Cambridge, MA: MIT Press, 2006, 1585-1592.
- [35] Jebara T. Multi-task feature and kernel selection for svms. In: Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004, 55-62.
- 220 [36] Argyriou A, Evgeniou T, Pontil M. Multi-task feature learning. In: Schölkopf B, Platt J, Hoffman T, eds. Advances in Neural Information Processing Systems 19, Cambridge, MA: MIT Press, 2007, 41-48.
- [37] Argyriou A, Micchelli C A, Pontil M, Ying Y. A spectral regularization framework for multi-task structure learning. In: Platt J C, Koller D, Singer Y, Roweis S, eds. Advances in Neural Information Processing Systems 20, Cambridge, MA: MIT Press, 2008.
- 225 [38] Hockema S A. Multi-task learning for software agents. Technical Report, School of Electrical and Computer Engineering, Purdue University, IN, 1998.
- [39] Richmond K. A multitask learning perspective on acoustic-articulatory inversion. In: Proceedings of the 8th Conference of INTERSPEECH, Antwerp, Belgium, 2007.
- 230 [40] Aha D W. Special AI review issue on lazy learning. Artificial Intelligence Review, 1997, 11.
- [41] Maron O. Learning from ambiguity. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1998.
- [42] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple-instance problem with axis-parallel rectangles. Artificial Intelligence, 1997, 89(1-2): 31-71.
- [43] Schapire R E, Singer Y. BoosTexter: a boosting-based system for text categorization. Machine Learning, 2000, 39(2-3): 135-168.
- 235 [44] Zhang M-L, Zhou Z-H. Multi-instance clustering with applications to multi-instance prediction. Applied Intelligence, 2009, 31(1): 47-68.