

文章编号: 2095-2163(2019)01-0149-05

中图分类号: TP391

文献标志码: A

# 计算机视觉中相似度学习方法的研究进展

王法强<sup>1</sup>, 张宏志<sup>1</sup>, 王鹏<sup>2</sup>, 邓红<sup>2</sup>, 张大鹏<sup>1</sup>

(1 哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001; 2 东北农业大学 理学院, 哈尔滨 150030)

**摘要:** 相似度学习方法通过学习合适的相似度度量以改进模型分类或聚类效果。现有的研究表明,相似度学习方法在很多计算机视觉问题中起到重要的作用。近年来随着数据规模的急剧增大和应用领域的多样化,相似度学习问题发展了很多新的研究领域。本文介绍了近年来相似度学习问题的研究进展和发展过程,包括从传统的二元组和三元组约束发展到新型相似度约束、从欧氏距离与马氏距离发展到新型相似度度量、从图像间的相似度学习发展到图像集之间的相似度学习、从单一模态相似度学习发展到跨模态相似度学习。最后本文展望了相似度学习未来可能的发展方向。

**关键词:** 相似度学习; 距离度量学习; 深度相似度学习

## Research progress on the similarity learning methods in computer vision

WANG Faqiang<sup>1</sup>, ZHANG Hongzhi<sup>1</sup>, WANG Peng<sup>2</sup>, DENG Hong<sup>2</sup>, ZHANG Dapeng<sup>1</sup>

(1 School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China;

2 College of Science, Northeast Agricultural University, Harbin 150030, China)

**[Abstract]** The similarity learning method aims to improve the classification or clustering performance by learning the appropriate similarity metric. The existing researches demonstrate that the similarity learning methods play very important roles in a number of computer vision applications. With the rapid growing data scale and the variant application fields in recent years, the similarity learning problem develops many new research fields. This paper introduces the research development progresses of similarity learning in recent years, i.e. from the traditional doublet and triplet constraints to new similarity constraint, from Euclidean distance and Mahalanobis distance to new similarity measure, from image based similarity learning to image set based similarity learning, from single modality similarity learning to multi-modality similarity learning. Finally, this paper prospects the possible develop direction of similarity learning in the future.

**[Key words]** similarity learning; distance metric learning; deep similarity learning

## 0 引言

相似度学习是计算机视觉中的一类重要方法,该方法通过在训练集中学习相似度(或距离)度量,使得同类别样本之间的相似度提高(或距离减小),不同类别样本之间的相似度降低(或距离增大)。相似度学习可以被嵌入到很多机器学习方法中,如k近邻分类、k均值聚类等。相似度度量学习已经应用于很多计算机视觉问题中,如人脸识别<sup>[1-2]</sup>、行人识别<sup>[3-5]</sup>和图像检索<sup>[6-7]</sup>等。对于不同的应用场景和不同的数据分布,从已有的训练数据集中学习合适的相似度度量,已经成为了计算机视觉中一个活跃的研究领域。一个好的相似度度量可以很大程度上提升计算机视觉应用的实际效果。

近年来,随着应用领域新需求的产生,数据规模的增大,以及数据来源的多样化,由相似度学习问题即衍生了一系列新的研究领域。本文首先探讨了相

似度的表示方式和传统的典型相似度学习方法,然后综述了近年来相似度学习方法的研究进展和发展过程,最后展望未来相似度学习方法可能的研究领域和发展方向。

## 1 相似度的表示

在目前的相似度学习方法中,大部分方法使用基于马氏距离或欧氏距离来表示图像之间的相似度,图像之间的距离越大,相似度越小,反之亦然。

对于基于马氏距离的相似度学习方法,给定2个样本 $\mathbf{x}$ 和 $\mathbf{y}$ ,样本之间的马氏距离定义为:

$$d_M^2(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^T \mathbf{M} (\mathbf{x} - \mathbf{y}) \quad (1)$$

其中, $\mathbf{M}$ 为距离度量矩阵,为了满足马氏距离的非负性, $\mathbf{M}$ 应当为半正定矩阵。

虽然传统的马氏距离度量学习方法可以从训练数据中学习距离度量,但是这些方法一般基于人工设计的样本特征,无法突破人工设计的特征对模型

基金项目: 国家自然科学基金(61671182,61471146)。

作者简介: 王法强(1989-),男,博士研究生,主要研究方向:计算机视觉、相似度学习。

收稿日期: 2018-05-17

效果造成的瓶颈。由于近年来基于卷积神经网络的深度学习方法在图像分类、人脸识别、目标检测和目标跟踪等领域取得了标志性重大进展,很多学者将相似度学习与卷积神经网络进行结合,研发了一系列深度相似度学习方法。这些方法通过联合学习图像的深度特征与相似度度量,取得了比基于人工设计的特征的相似度学习更好的效果。

在深度相似度学习方法中,研究通过卷积神经网络学习并提取图像  $x$  和  $y$  的特征,分别表示为  $f(x)$  和  $f(y)$ ,所以样本  $x$  和  $y$  之间的相似度可以使用图像特征  $f(x)$  和  $f(y)$  的欧氏距离表示如下:

$$d_f^2(x, y) = \|f(x) - f(y)\|_2^2 \quad (2)$$

除了公式(1)和(2)所示的相似度函数外,还有如余弦相似度、卡方距离等相似度函数均已在相似度学习方法中获得了成功使用。

## 2 典型相似度学习方法回顾

自从距离度量学习问题由 Xing 等人<sup>[8]</sup>提出以来,众多学者发展了一系列方法来学习以上形式的相似度度量。根据约束形式的不同,大多数相似度学习方法可以分为基于二元组约束的相似度学习方法以及基于三元组约束(又称相对距离约束或相对相似度约束)的相似度学习方法。本节将从约束形式的角度对现有的典型相似度学习方法进行总结和回顾。对此可做阐释分述如下。

### 2.1 基于二元组约束的相似度学习方法

给定训练图像集,可以建立一系列二元组  $(x, y)$ ,根据二元组中的2个样本是否属于同一类,就可以将二元组分为同类二元组集合  $S$  和异类二元组集合  $D$ 。对于集合  $S$  和  $D$  中的二元组,分析指出研究中需分别满足下列约束:

$$\begin{aligned} s(x, y) &\leq b + 1 - \xi_{(x,y)} \quad \forall (x, y) \in D \\ s(x, y) &\geq b - 1 + \xi_{(x,y)} \quad \forall (x, y) \in S \end{aligned} \quad (3)$$

其中,  $s(x, y)$  为样本  $x$  和  $y$  的相似度,  $\xi_{(x,y)}$  为二元组  $(x, y)$  对应的非负松弛变量。

二元组约束的优点在于可以将其应用于弱监督学习问题,即每个样本的类别标记未知,仅能获得每个二元组为同类二元组或异类二元组的标记信息。

对于马氏距离度量学习方法, Xing 等人<sup>[8]</sup>通过最小化同类二元组的距离,并且约束异类二元组的距离大于给定阈值,从而将距离度量学习构造为一个凸优化问题。Davis 等人<sup>[9]</sup>提出了一种基于信息论的度量学习方法,该方法在满足二元组约束的同时,通过引入度量矩阵的先验矩阵,由此求得了最小

化度量矩阵与先验矩阵的 KL 散度,从而使得度量矩阵和先验矩阵尽可能接近。

对于深度相似度学习方法, Sun 等人<sup>[10]</sup>将二元组约束构造为对比损失 (*Contrastive Loss*) 函数,并将其与 Softmax 损失结合,提出了一种深度相似度学习方法,再将其应用于人脸验证问题,在 LFW 数据集上的准确率达到了 99.15%。Lin 等人<sup>[11]</sup>提出了一种泛化的相似度函数,并且同样使用二元组损失函数对相似度和深度网络进行学习,而后则将该方法应用于跨域匹配问题。

### 2.2 基于三元组约束的相似度学习方法

基于三元组约束的相似度学习方法对训练集建立一系列的三元组约束,对于三元组  $(x, y, z)$ ,其中  $x$  和  $y$  类别相同,  $x$  与  $z$  类别不同,则该三元组应当建立以下约束:

$$d(x, z) - d(x, y) \geq 1 - \xi_{(x,y,z)} \quad (4)$$

该类约束使得每个样本在与同类样本和异类样本的距离对比中,其与同类样本的距离更近。其优点在于可以应用于排序 (Ranking) 问题(如图像检索)。在该问题中,研究中只有参考图像与查询图像的相对相似度排序信息,而没有每个图像的类别信息,在这种情况下三元组约束就可以得到有效应用。

在马氏距离度量学习方法中,典型的基于三元组约束的方法、如大间隔最近邻方法则是由 Weinberger 等人于 2009 年提出<sup>[12]</sup>,设计上即通过最大化每个样本与其同类样本和不同类样本的距离间隔来学习距离度量。Parameswaran 等人<sup>[13]</sup>将该方法扩展到多任务学习问题上。Kedem 等人<sup>[14]</sup>还将该方法延拓至卡方距离等非线性距离度量的学习研究中。

在深度相似度学习方法中, Schroff 等人<sup>[2]</sup>提出了 FaceNet 模型。该模型将三元组约束构造为损失函数的形式,再将其与深度卷积网络结合,旨在进行深度相似度学习,最终该模型在人脸验证中取得了突出的效果。Ding 等人<sup>[15]</sup>同样使用三元组约束学习深度相似度,也将其应用于行人识别问题。Zhang 等人<sup>[6]</sup>使用海明贴近度 (Hamming affinity) 作为相似度函数,又通过三元组约束构造铰链损失,随后就可展开图像的哈希编码学习。

## 3 相似度学习方法的近期发展

近年来,随着计算机视觉的应用需求以及视觉数据来源的多样化,对相似度学习模型的要求也在逐步提高。针对这种现状,研究人员相继推出了一

系列新的相似度学习方法,这些方法的创新点包括新型距离(相似度)表示、新型相似度约束、图像集之间的相似度学习,以及跨模态的相似度学习等。基于此,本节将阐述近年来相似度学习方法的发展过程和研究进展。研究内容详见如下。

### 3.1 从欧氏距离和马氏距离到新型距离(相似度)表示

在传统的距离度量学习与深度相似度学习中,一般采用公式(1)和(2)所示的马氏距离或深度特征的欧氏距离来表示样本间的相似度。除了这2种方式以外,研究人员还根据不同的应用场景或需求构造了其它类型的相似度函数。Chen等人<sup>[16]</sup>提出了一种联合贝叶斯方法,该方法使用2个样本属于不同类和同类的概率比值作为2个样本是否同类决策函数,并且分析该函数相当于马氏距离与2个样本交叉内积项的组合。Li等人<sup>[17]</sup>在马氏距离中引入了二阶函数表示的自适应阈值,从而将马氏距离扩展为预测2个样本是否同类的二阶决策函数。Li等人<sup>[18]</sup>将马氏距离与余弦相似度进行结合,构造了一种新的相似度函数,并且基于三元组约束设计了联合优化距离与相似度度量的方法。Lin等人<sup>[11]</sup>通过将传统马氏距离中的线性投影扩展到仿射变换,并且将马氏距离与余弦相似度结合,定义了一种通用相似度度量,该项工作还将该相似度与深度网络结合,提出了一种深度相似度学习方法。Wang等人<sup>[19]</sup>将单图像特征的欧氏距离和图像对特征的二值分类输出作为判断图像对是否同类的决策函数,并将两者进行结合构造了一种包含单图像特征和图像对特征的新型相似度函数。

### 3.2 从传统约束到新型约束

传统的相似度学习方法一般通过在优化问题中引入二元组约束或三元组约束来学习相似度度量。近年来,研究人员针对相似度学习发表了一系列新型相似度约束的成果。Chen等人<sup>[20]</sup>将三元组约束进行扩展,提出了四元组相似度损失函数,并将其与深度网络进行结合,即使得四元组中2个同类样本的深度相似度高于2个异类样本的深度相似度。Wang等人<sup>[21]</sup>对三元组损失函数加以改进,通过约束三元组中负样本点与其它2个样本构成的角度小于给定阈值构造了角度损失(Angular loss)。Song等人<sup>[22]</sup>提出了一种新的度量学习框架,该框架通过优化聚类质量度量NMI构造损失函数。

### 3.3 从图像之间的相似度学习到图像集之间的相似度学习

传统的相似度学习方法一般对2张图像之间的

相似度进行学习。然而在一些计算机视觉问题中,研究常常是将一组同一类别的图像看作一个图像集,并希望准确度量图像和图像集之间的相似度,以及2个图像集之间的相似度,如图像-视频匹配、视频-视频匹配等。因而需要继续加强与图像集相关的相似度学习方法的深入探究与开发。研究可知,其关键问题在于如何表示图像集,以及如何表示图像样本点到图像集之间以及2个图像集之间的相似度。

Zhu等人<sup>[23]</sup>提出了一种样本点-集合和集合-集合距离度量学习方法,该方法使用仿射包表示图像集,并基于此将图像样本点到图像集之间的距离定义为样本点到图像集仿射包中最近点的距离,将2个图像集之间的距离定义为2个相应仿射包的最近点的距离。Wang等人<sup>[24]</sup>用流形表示图像集,并且将流形表示为一组局部线性子空间,并基于此构建了样本点、子空间和流形之间的距离度量学习框架。Lu等人<sup>[25]</sup>将图像集同样表示为流形,并通过深度网络学习非线性投影,再将每个流形投影到公共子空间中,而后计算求出其距离。Huang等人<sup>[26]</sup>针对样本点到集合的匹配问题,提出了欧几里得-黎曼距离度量,该方法将图像集合表示为黎曼流形,然后将其嵌入到高维希尔伯特空间中,最后将样本集所在的希尔伯特空间和样本点所在的欧几里得空间投影至相同的欧几里得子空间中,并计算投影后样本点和样本集之间的距离。

### 3.4 从单一模态相似度学习到跨模态相似度学习

传统的相似度学习方法一般只针对单一模态,例如自然图像之间的相似度学习、素描图像之间的相似度学习、文本之间的相似度学习等,然而传统方法不能处理一些跨模态匹配的应用场景,如自然图像与素描图像匹配、图像与文本匹配等。

针对这些新型应用场景,研究人员通过将单一模态的相似度学习扩展到跨模态相似度学习,提出了一系列解决方法。这些方法可以分为2类,对其中的每一部分可给出研究表述如下。

一类是基于跨模态样本生成的方法,即通过将其中一个模态的样本转换为另一模态,使得2个样本模态相同,然后对2个样本进行匹配,如Zhang等人<sup>[27]</sup>针对人脸-素描图像匹配问题,通过全卷积网络将人脸图像生成对应的伪素描图像,再将其与素描图像进行匹配。

另一类是基于特征学习的方法,这种方法一般寻找一个公共子空间,在此基础上学习不同模态的

样本到该子空间的投影,并在公共子空间中学习不同模态样本之间的距离。如 Liong 等人<sup>[28]</sup>提出了深度耦合度量学习 (Deep Coupled Metric Learning, DCML)方法,通过学习 2 个深度网络,将 2 个模态的样本非线性变换到公共空间,并最大化类间散度、最小化类内散度。Bronstein 等人<sup>[29]</sup>提出了一种跨模态相似度学习框架,将不同模态的样本投影到公共的海明 (Hamming)空间中,通过 Boosting 算法学习这种投影,再通过这种投影参数化相似度度量。

## 4 结束语

相似度学习是计算机视觉中的一类重要方法。本文从相似度的表示和相似度的约束两方面回顾了现有的典型相似度学习方法,并且整体分析了近年来相似度学习的发展过程与研究进展。对于相似度学习方法的未来走向,可以从应用领域和训练方法等方面对现有的相似度学习方法进行扩展,或提出新的相似度学习方法。例如,在有些计算机视觉应用中,2 个自然图像之间往往存在非常复杂的对应关系,如何设计更符合样本图像以及网络结构特点的新型相似度 (或距离) 函数对图像的对应关系建模即已成为目前亟待解决的热点问题。此外,由于深度网络的高度非凸性,使得二元组或三元组的选取变得更加重要。所以设计一种能够随着训练的逐层推进,从训练集中自适应地选取二元组 (或三元组) 的方式,也可以在识别效果和训练效率方面起到较大的作用。

## 参考文献

- [1] GUILLAUMIN M, VERBEEK J, SCHMID C. Is that you? Metric learning approaches for face identification [C]// The 12<sup>th</sup> IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009: 498-505.
- [2] SCHROFF F, KALENICHENKOD, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [J]. arXiv preprint arXiv:1503.03832, 2015.
- [3] ZHENG Weishi, GONG Shaogang, XIANG Tao. Person re-identification by probabilistic relative distance comparison [C]// The 24<sup>th</sup> IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2011. Colorado, USA: IEEE, 2011: 649-656.
- [4] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning [C]//Proc. IEEE International Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 2197-2206.
- [5] LIAO Shengcai, LI S Z. Efficient PSD constrained asymmetric metric learning for person re-identification [C]//2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 3685-3693.
- [6] ZHANG Ruimao, LIN Liang, ZHANG Rui, et al. Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification [J]. IEEE Transactions on Image Processing, 2015, 24(12): 4766-4779.
- [7] BHATTARAI B, SHARMA G, JURIE F. CP-mtML: Coupled projection multi-task metric learning for large scale face retrieval [J]. arXiv preprint arXiv:1604.02975, 2016.
- [8] XING E P, NG A Y, JORDAN M I, et al. Distance metric learning with application to clustering with side-information [C]// Advances in Neural Information Processing Systems 15 (NIPS). Vancouver, British Columbia: dblp, 2002: 505-512.
- [9] DAVIS J V, KULIS B, JAIN P, et al. Information-theoretic metric learning [C]// Proceedings of the 24<sup>th</sup> International Conference on Machine Learning. Corvallis, OR, USA: ACM, 2007: 209-216.
- [10] SUN Yi, CHEN Yuheng, WANG Xiaogang, et al. Deep learning face representation by joint identification-verification [C]// NIPS'14 Proceedings of the 27<sup>th</sup> International Conference on Neural Information Processing Systems - Volume 2. Montreal, Canada: ACM, 2014, 27:1988-1996.
- [11] LIN L, WANG G, ZUO W, et al. Cross-domain visual matching via generalized similarity measure and feature learning [J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, 39(6): 1089-1102.
- [12] WEINBERGER K Q, SAUL L K. Distance metric learning for large margin nearest neighbor classification [J]. Journal of Machine Learning Research, 2009, 10: 207-244.
- [13] PARAMESWARAN S, WEINBERGER K Q. Large margin multi-task metric learning [C]// Advances in Neural Information Processing Systems. British Columbia, Canada: dblp, 2010: 1867-1875.
- [14] KEDEM D, TYREE S, WEINBERGER K Q, et al. Non-linear metric learning [J]. Advances in Neural Information Processing Systems, 2012, 4: 2582-2590.
- [15] DING Shengyong, LIN Liang, WANG Guangrun, et al. Deep feature learning with relative distance comparison for person re-identification [J]. Pattern Recognition, 2015, 48(10): 2993-3003.
- [16] CHEN Dong, CAO Xudong, WANG Liwei, et al. Bayesian face revisited: A joint formulation [M]// FITZGIBBON A, LAZEBNIK S, PERONA P, et al. Computer Vision-ECCV 2012. ECCV 2012. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, 2012, 7574: 566-579.
- [17] LI Zhen, CHANG Shiyu, LIANG Feng, et al. Learning locally-adaptive decision functions for person Verification [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, Oregon, USA: IEEE, 2013: 3610-3617.
- [18] LI Mu, WANG Qilong, ZHANG D, et al. Joint distance and similarity measure learning based on triplet-based constraints [J]. Information Sciences, 2017, 406-407: 119-132.
- [19] WANG Faqiang, ZUO Wangmeng, LIN Liang, et al. Joint learning of single-image and cross-image representations for person re-identification [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016). Las Vegas: IEEE, 2016: 1288-1296.