基于混合式协同训练的人体动作识别算法研究

景陈勇 詹永照 姜 震
(江苏大学计算机科学与通信工程学院 镇江 212013)

摘 要 人体动作识别是计算机视觉研究中备受关注的课题。现有的动作识别方法大多属于监督学习,需要大量的有标记数据来训练识别模型。然而,在现实应用中有标记的数据成本较高,而无标记数据很容易获取。提出一种基于混合式协同训练的新型人体动作识别算法——Co-KNN-SVM,该算法利用动作识别领域不同类型的方法来构建基分类器,并进行迭代的相互训练以提高泛化性能,可以降低标注成本,并实现不同识别方法的优势互补。此外,还改进了协同训练中对伪标记数据的选择方法和迭代训练策略,有效控制了伪标记数据的噪声影响,提高了协同训练的识别效果。实验结果表明,所提算法可以有效地识别视频中的人体动作。

关键词 动作识别,监督学习,混合式协同训练,噪声
中图法分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.07.049

Research on Action Recognition Algorithm Based on Hybrid Cooperative Training

JING Chen-yong ZHAN Yong-zhao JIANG Zhen
(School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract Human action recognition is an important issue in the field of computer vision. Existing action recognition methods mostly belong to supervised learning category, in which a large number of labeled data are needed to train the recognition model. However, in many real-world tasks, labeled data are often expensive to get, while unlabeled data are readily available in abundance. In this paper, a novel human action recognition algorithm, named as Co-KNN-SVM, was proposed based on hybrid collaborative training. Different types of recognition methods for action recognition field are employed in this method to build the base classifiers, which are then iteratively retrained to increase their generalization abilities. In general, our method can decrease the labeling cost and achieve complementary advantages of different recognition algorithms. In order to decrease the impact of the noise in pseudo labeled data and improved the recognition performance, we also improved the selection method for the pseudo label data and the iterative training strategy in co-training style algorithms. The experimental results show that the proposed algorithms can identify human action in the video more effectively.

Keywords Action recognition, Supervised learning, Hybrid collaborative training, Noise

1 引言

人体动作识别是计算机视觉领域研究的热点问题之一,在视频监控、智能交通等领域有着广泛的应用。然而在真实自然场景下,背景复杂、摄像机运动和物体变化等问题增加了动作识别的复杂性[1-2]。目前的动作识别方法大致分为3类。

(1)基于模板的方法。这类方法又分为模板匹配[4-6]和动态时间规整(Dynamic Time Warping)[2]。前者需要事先对某一特定动作建立特征数据样本模板库,识别时只需获取待识别动作样本同样的特征数据与模板库中的模板进行匹配,其算法简单,但很难构造出足够的模板来处理不同的行人姿态。后者针对两个具有不同时间长度的动作模板,按照一定的时间规整曲线进行调整,可以较好地解决人体动作在时间上的不确定性问题,但模型构建比较困难,并且无法真正反映动态系统在特征空间的分布属性。

(2)基于概率统计的方法。该方法从一系列训练数据中学习得到分类器[7-9],其优点在于引入了概率框架,较好地解决了同类动作模式间的不确定性问题,鲁棒性较好,但其需要较多训练数据,并且很难解决姿态遮挡的问题。

(3)基于文法的方法[3]。这类方法是将人体动作分解为一连串的符号,首先识别这些符号,然后将人体动作表示为一系列生成的符号动作流。该方法有利于对复杂结构的理解和对先验知识的有效利用,但其计算复杂度高,空间尺度鲁棒性对底层描述符号的依赖较大。

上述动作识别方法大多是有监督的。鉴于动作识别的场景复杂性,为了获得更好的识别表现,需要大量的有标记样本

到稿日期:2016-06-29 返修日期:2016-09-02 本文受江苏省重点研发计划项目(1721190141),江苏省高校自然科学基金(15KJB520006)资助。
景陈勇(1992—),男,硕士生,主要研究方向为多媒体技术,E-mail:ujshenry@qq.com;詹永照(1962—),男,博士,教授,博士生导师,CCF高级会员,主要研究方向为人机交互、模式识别、多媒体技术;姜 震(1976—),男,博士,副教授,主要研究方向为机器学习。

来训练模型，代价高昂。因此，利用无标记数据来提升识别效果成为了一个非常有前景的研究方向。本文针对动作识别领域的特点提出一种新型混合式协同训练方法——Co-KNN-SVM，该方法可以有效结合无标记数据来提高动作识别的准确率；同时还改进了协同训练中对伪标记数据的选择策略。实验结果表明，所提算法具有更为优越的性能。

2 相关工作

半监督学习结合有标记数据和无标记数据来改善学习性能，近年来已涌现出许多这类算法。协同训练(Co-training)是其中非常成功的一类方法，利用基分类器之间的差异性，相互提供伪标记数据来进行迭代训练，从而提高模型的泛化性能[10-11]。基于不同的视角(特征集)或不同的训练集来构建基分类器是最常见的做法[12-13]，此外也有一些研究者基于不同的学习算法来构建基分类器[11-12]。如何降低伪标记数据的噪声是协同训练类算法的关键。

目前，将协同训练应用于动作识别领域的研究还较少。Gupta 等人[13]提出了基于 Co-training 的动作识别方法，他们选取视频集的视觉和纹理特征作为两个不同的视角，取得了较好的识别效果。随后，Yu 等人[14]提出了基于贝叶斯无向图模型的协同训练识别方法，为了同时优化两个视角，该方法为高斯过程分类器引进了一种新型的 co-training 内核，进一步提高了性能。为了满足标准协同训练中对视图独立性与冗余性的要求，Liu 等人[15]提出了 Inter-view 和 Intra-view 这两种置信度评价方法，同时它们也可以用来度量不同视图间的独立性。Tang 等人[16]提出了一种基于多学习器协同训练的动作识别方法，该方法首先利用基于 Q 统计量的选择算法来选取基分类器集；然后基于分类器成员委员会的标记近邻置信度计算公式来评估未标记数据的置信度，从中选取置信度较高的数据加入到训练集中并更新分类器；最后利用迭代训练后的基分类器集合进行决策输出。

上述算法都是利用不同的视角来构建基分类器，而且大多仅根据置信度来选择伪标记样本，未考虑伪标记数据与实际样本空间的分布差异(分布噪声)。本文提出一种基于动作识别中不同类型的算法来构建基分类器的新思路，以实现动作识别领域中不同方法的优势互补；此外，对伪标记数据的选择方法和迭代训练策略进行改进，有效控制了伪标记数据对训练样本的噪声影响，提升了协同训练的效果。第 3 节将给出该算法的详细介绍。

3 Co-KNN-SVM 算法

3.1 基分类器

本节采用一个基于模板的模型和一个基于概率统计的模型来构建协同训练的基分类器。

(1)基于模板的分类器

由于 KNN(K 最近邻法)方法主要依赖周围有限的邻近样本，因此对于类域的交叉或重叠较多的待分样本集，该方法更为合适，并且其可以较好地避免样本的不平衡问题。因此，选择 KNN 作为基于模板的分类器。

KNN 算法的思想为：如果一个样本在特征集中的 k 个最近邻的样本中的大多数同属于某一个类别，则该样本也属于该类别。对于未标记样本 x，只需要比较 x 与 $N=\sum_{i=1}^{c}N_i$ 个已知类别样本之间的距离即可决策 x 与离它最近的样本同类。通常采用余弦距离作为两个样本向量之间的距离。一个样本的最近邻就是在上述定义下与其距离最近的样本。

(2)基于概率统计的分类器

对于基于概率统计的分类器，选取 SVM(支持向量机)作为基分类器，其具有能解决高维问题的特点。SVM 作为一种线性分类器，寻求建立一个由 $f(x)=wx+b$ 表示的决策面。当 $f(x)=0$ 时 $P(y|x;\Lambda)=1$，否则为 0。其模型参数 $\Lambda=\{w,b\}$ 可通过极大化式(1)所示的目标函数来估算。

$$\begin{aligned}&\min_{w,b,\xi}=\frac{1}{2}\|w\|^2+c\sum_{i=1}^{n}\xi_i\\&\text{s.t. } y_i((w\cdot x_i)+b)\geqslant 1-\xi_i,i=1,\cdots,n\\&\quad\xi_i\geqslant 0,i=1,\cdots,n\end{aligned}\tag{1}$$

其中，ξ_i 为标准数据上的松弛变量，c 为给定的惩罚因子。

3.2 混合协同训练算法

协同训练的一项重要工作是如何选择合适的伪标记数据，从而降低训练样本的分类噪声和分布噪声。因此，通过提高伪标记数据的准确率和控制类别比例来提高协同训练的性能。

(1)为了降低伪标记数据中的分类噪声，根据其置信度来递增地向训练样本中添加无标记数据。

首先计算基于模板的分类器对无标记数据的预测置信度。对于无标记数据 x_i，KNN 能够对每个类别 c_j 给出一个预测概率：

$$P(y=c_j|x_i)=d_m(x)/k\tag{2}$$

其中，k 为测试样本的近邻点个数，$d_m(x)$ 为近邻样本中所预测类别样本的数量。为了计算预测样本的置信度，最直接的方法是将数据预测类别的概率作为权重，选择最大的类预测概率 $P(y=c_{max\_j}|x_i)$ 作为置信度 $C_{KNN}(x_i)$，即 $C_{KNN}(x_i)=P(y=c_{max\_j}|x_i)$。但仅将类的最大预测概率作为置信度不够合理，因此采用一种新的置信度计算标准：通过类别最大的概率以及它与第二大概率的差值来衡量置信度，如式(3)所示：

$$C_{KNN}(x_i)=P_{KNN}(y=c_{max\_j}|x_i)-P_{KNN}(y=c_{sub\_max\_j}|x_i)\tag{3}$$

采用此方法计算置信度能够把可能位于类重叠区域的数据排除在 SVM 的伪标记数据集之外，有效解决了 SVM 在类重叠情况下性能下降的问题。

对于基于概率统计的分类器 SVM，由于它只是通过决策面来划分类别，并不输出预测概率，因此参考文献[17]中的方法来转化预测概率 $P(y=c_j|x_i)$，然后同样把置信度表示为：

$$C_{SVM}(x_j)=P_{SVM}(y=c_{max\_j}|x_j)-P_{SVM}(y=c_{sub\_max\_j}|x_j)\tag{4}$$

因此，SVM 选择的伪标记数据的加入有助于解决 KNN 在小样本情况下性能不足的问题，并降低训练样本中类别不平衡的不利影响。

(2)在协同训练中，由于伪标记数据不是独立随机选取的，因此其分布与真实分布之间必然存在误差。采取控制伪标记数据集中各个类别比例的方法来降低这种分布噪声，在

协同训练中按照有标记集类别的比例来动态地添加伪标记数据。

除了按照置信度选择伪标记数据外，还改变了以往在协同训练中单调添加伪标记的做法。首先构造一个伪验证集 V，其由初始训练集 L 以及 $P_1 \cup P_2$（P_1，P_2 代表伪标记数据集）中置信度最高的前 30%组成，然后通过计算分类器 h_i 在 V 上的准确率 $A_v(h_i)$ 来估计其性能；另一方面，采用 $e(h_i)$ 来估计其错误率：

$$e(h_i)=\sum_{j=1}^{k}|p_U(y=j|h_i)-p_L(y=j)| \tag{5}$$

其中，$P_U(j|h_i)$ 与 $P_L(j)$ 分别表示 h_i 对 U 预测结果中 j 所占的比例和 L 中类别 j 的比例。因此，结合 $A_v(h_i)$ 和 $e(h_i)$，把会导致分类器退化的伪标记数据及时移除。混合协同训练算法的具体描述如算法 1 所示。

算法 1　Co-KNN-SVM 算法

输入：数据集 L，U 以及伪标记数据集 $P_1=P_2=\emptyset$

输出：经过迭代训练后得到的更优的基分类器 KNN 和 SVM

初始化 $j=N=0$；$P_1'=P_2'=\emptyset$；$V=L$，最大迭代次数为 Max_N。

1. 分别在 L 上训练 KNN 和 SVM，然后对 U 中的数据进行分类，并分别估计其 $e(h_1^{(0)})$，$A_L(h_1^{(0)})$ 和 $e(h_2^{(0)})$，$A_L(h_2^{(0)})$。
2. 重复步骤 3－步骤 4，直到 N>Max_N。
3. SVM 按类别比例从 $U-P_1$ 中选择 m 个置信度最大的并且与 KNN 预测一致的数据加入 p_1' 中，并令 $p_1=p_1\cup p_1'$；KNN 按类别比例从 $U-P_2$ 中选择 m 个置信度最大的并且与 NB 预测一致的数据加入 p_2' 中，并令 $p_2=p_2\cup p_2'$。
4. 分别在 $L\cup P_1$，$L\cup P_2$ 上重新训练 KNN 和 SVM，然后用它们估计其 $e(h_n^{j+1})$ 和 $A_v(h_n^{j+1})$，其中 n=1，2。

   If $A_v(h_n^{j+1})<A_v(h_n^j)$ or $e(h_n^{j+1})>e(h_n^j)$

   $P_1=P_1-P_1'$；$h_n^{(j+1)}-h_n^j$

   Set $P_n'=\emptyset$

设置 $N=N+m$；$j=j+1$。

3.3　结合 KNN 和 SVM

通过 3.2 节中的训练算法，可以得到两个性能良好的基分类器。为了进一步提高准确率，将 KNN 和 SVM 结合起来进行最终的预测。虽然能把 SVM 的预测结果转化为概率形式，但是 KNN 和 SVM 的预测概率可能并不在同一个尺度上，因此直接把二者的预测概率结合起来并不能得到满意的结果。本节将 KNN 和 SVM 对预测的置信度做了归一化处理，然后按照下式给出最终的分类结果。

$$P(y_i|x_i)=\begin{cases}P(y_i|x_i), & \text{if } (y_i|x_i)\neq P(y_i|x_i) \text{ and } \dfrac{C_{KNN}(x_i)}{\sum_{x_i\in U}C_{KNN}(x_j)}>\dfrac{C_{SVM}(x_i)}{\sum_{x_j\in U}C_{SVM}(x_j)}\\ P(y_i|x_i), & \text{otherwise}\end{cases}$$

4　实验结果与分析

4.1　实验数据库

首先采用动作识别数据库 UCF YouTube Action Data Set[18]（见图 1），该数据库包含 11 类体育运动：basketball shooting，biking，diving，golf swinging，horse back riding，soccer juggling，swinging，tennis swinging，trampoline jumping，volleyball spiking，walking with a dog。其中每一类由 25 个人做动作，每个人做 4～7 组，共有 1590 个视频。在此视频库的数据集上，使用一种比较简单的关键帧获取方法，即选取每个视频的首帧、中间帧、尾帧作为关键帧，然后提取 4 种特征作为表征动作，分别为颜色特征、纹理特征、径向 Tchebichef 矩特征以及多尺度 LBP 特征，最后把这些特征融合起来作为最终的特征，即 $A=(t,W,Hc,LBP)$。其中，A 表示融合后的特征，t 表示径向 Tchebichef 矩特征，W 表示纹理特征，Hc 表示颜色特征，LBP 表示多尺度 LBP 特征。

图 1　UCF YouTube Action Data Set

4.2　实验设计与分析

为了验证本文方法的有效性，选择 Co-training 和 MCM 算法[13,16]与其进行对比。Co-training 是一种标准的协同训练算法，采用两种不同的分类器，分别在视觉和纹理两种不同的视角上训练数据。MCM 采用文献[10]中多分类器协同训练的标准设置，首先采用基于 Q 统计量的分类器选择算法选择一组基分类器，然后对基分类器进行协同训练。

表 1 列出了不同有标记样本数下的识别准确度。从表中可以看到，协同训练中由于使用了无标记样本来提升分类器的性能，因此相比于 SVM 和 KNN 等有监督学习分类器，分类器精度有了明显的提升；同时，本文所提方法在相同的特征数据集上的识别效果都优于另外两种方法。

表 1　本文方法与其他方法的比较

有标记样本数	SVM	KNN	Co-raining	MCM	本文方法
220	0.47	0.52	0.61	0.76	0.78
440	0.65	0.67	0.70	0.78	0.81
660	0.74	0.75	0.78	0.78	0.80
平均准确率	0.62	0.67	0.70	0.77	0.80

(1)与 Co-training 的比较

由于 Co-training 算法需要两个不同的充分冗余的视图，而我们的视频特征不能充分满足这一点，因此识别准确率低于所提方法。

(2)与 MCM 的比较

MCM 通过一组 Q 统计量来选取多个基分类器进行协同训练。采取多个基分类器来选择置信度较高的伪标记样本，虽然能够较好地控制伪标记的分类噪声，但未能有效地解决分布噪声问题，因此分类结果略差于所提方法。

图 2 和图 3 分别给出了在相同特征数据集下，无标记训练样本个数与有标记训练样本个数对分类器精度的影响。从图 2 可以看到，在初始有标记样本为 220，440 和 660 的情况下，随着无标记样本的增加，分类器精度均得到了提升；并且当有标记样本为 220 时，分类器的精度提升的程度最大。因

此，对于协同训练来说，如何控制有标记和无标记样本的比例很重要。当有标签样本较少时，选择置信度较高的伪标记数据，通过协同训练能充分利用这些数据。随着有标记训练样本的增加，预测的准确率也增加，但当有标记训练样本中的数据增多时，利用已存在的有标记数据可以训练一个性能优良的分类器，因此即使继续添加无标记数据，也不能继续提高分类器的性能。

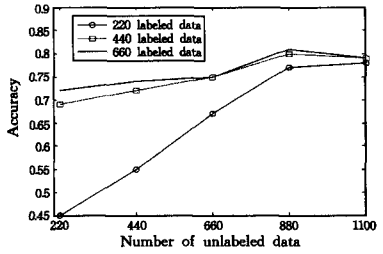

图 2 不同的无标记样本数对 Co-KNN-SVM 的影响

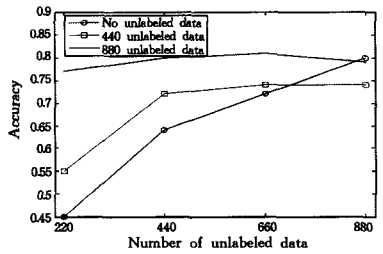

图 3 不同的有标记样本数对 Co-KNN-SVM 的影响

结束语 本文针对视频中的动作识别问题，提出了一种基于模板和基于概率统计的混合式协同训练方法，实现了现有动作识别方法的优势互补。此外，还关注了协同训练类方法中的分布噪声问题，并提出了相应的改进措施。实验结果表明，相比于其他同类动作识别方法，所提方法有效地提高了动作识别的准确率，特别是在有标记样本不足的情况下。在未来的工作中，将进一步考虑利用动作识别领域天然的多视角，并与集成学习相结合。

参 考 文 献

[1] LI R F, WANG L L. A Survey of Human Body Action Recognition[J]. Pattern Recognition and Artificial Intelligence, 2014, 27(1): 35-48. (in Chinese)
李瑞峰，王亮亮. 人体动作行为识别研究综述[J]. 模式识别与人工智能，2014，27(1)：35-48.

[2] HU Q, QIN L, HUANG Q M. A Survey on Visual Human Action Recognition[J]. Chinese Journal of Computers, 2013, 36(12): 2512-2524. (in Chinese)
胡琼，秦磊，黄庆明. 基于视觉的人体动作识别综述[J]. 计算机学报，2013，36(12)：2512-2524.

[3] XIE L D. Research on complex human behavior recognition based on hierarchical method[D]. Xiamen: Xiamen University, 2014. (in Chinese)
谢立东. 基于分层方法的复杂人体行为识别研究[D]. 厦门：厦门大学，2014.

[4] LIN Z, JIANG Z, DAVIS L S. Recognizing Actions by Shape-Motion Prototype Trees[C]// IEEE International Conference on Computer Vision. 2009: 444-451.

[5] JIANG Z, LIN Z, DAVIS L S. A Tree-Based Approach to Integrated Action Localization, Recognition and Segmentation[M]// Trends and Topics in Computer Vision. Springer Berlin Heidelberg, 2010: 114-127.

[6] LIU J G, ALI S, SHAH M. Recognizing human actions using multiple features [C]// Computer Vision and Pattern Recognition, 2008. IEEE Conference on IEEE, 2008: 1-8.

[7] MATIKAINEN P, HEBERT M, SUKTHANKAR R. Trajectons: Action recognition through the motion analysis of tracked features[C] // International Conference on Computer Vision Workshops. IEEE, 2009: 514-521.

[8] MATIKAINEN P, HEBERT M, SUKTHANKAR R. Representing Pairwise Spatial and Temporal Relations for Action Recognition[C] // European Conference on Computer Vision. 2010: 508-521.

[9] NATARAJAN P, NEVATIA R. Online, Real-time Tracking and Recognition of Human Actions[C]// IEEE Workshop on Motion and video Computing, 2008(WMVC 2008). 2008: 1-8.

[10] GUAN D, YUAN W, LEE Y K, et al. Activity Recognition Based on Semi-supervised Learning[C] // IEEE International Conference on Embedded and Real-Time Computing Systems and Applications. IEEE Computer Society, 2007: 469-475.

[11] LIU W, LI Y, TAO D, et al. A general framework for co-training and its applications[J]. Neurocomputing, 2015, 167(c): 112-121.

[12] JIANG Z, ZHANG S, ZENG J. A hybrid generative/discriminative method for semi-supervised classification[J]. Knowledge-Based Systems, 2013, 37(2): 137-145.

[13] GUPTA S, KIM J, GRAUMAN K, et al. Watch, Listen & Learn: Co-training on Captioned Images and Videos[M]// Machine Learning and Knowledge Discovery in Databases. DBLP, 2008: 457-472.

[14] YU S, KRISHNAPURAM B, ROSALES R, et al. Bayesian Co-Training[J]. Journal of Machine Learning Research, 2007, 12(3): 2649-2680.

[15] LIU C, YUEN P C. A Boosted Co-Training Algorithm for Human Action Recognition[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2011, 21(9): 1203-1213.

[16] TANG C, WANG W J, LI W, et al. Multi-Learner co-training model for human action recognition[J]. Journal of Software, 2015, 26(11): 2939-2950.

[17] BREFELD U, SCHEFFER T. Co-EM support vector learning [C] // International Conference on Machine Learning. ACM, 2004: 121-128.

[18] LIU J, LUO J, SHAH M. Recognizing realistic actions from videos "in the Wild"[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2009: 1996-2003.