DOI：10.7652/xjtuxb201307018

融合异构信息的网络视频在线半监督分类方法

杜友田[1]，辛刚[2]，郑庆华[1,2]
（1.西安交通大学智能网络与网络安全教育部重点实验室，710049，西安；
2.西安交通大学陕西省天地网重点实验室，710049，西安）

摘要：针对大规模网络视频数据的学习需要考虑无标签数据和异构信息的问题，提出了一种基于视觉和文本异构信息的网络视频在线半监督学习方法。该方法将文本和视觉看作2个视图，采用图作为基分类器对每个视图进行建模，并利用线性邻域的传播算法来预测样本类别。在不同视图之间采用多图上的协同训练，利用未标记样本增量地更新基分类器，并根据类别相关的融合方法确定最终结果，从而提高了分类准确率。实验结果表明，该方法的结果优于支持向量机方法约8.3%，在线增量更新后，学习器的性能提高了约3%，因此比较适合于大规模视频数据的在线半监督学习。

关键词：网络视频；异构信息；半监督分类
中图分类号：TP391.4 **文献标志码**：A **文章编号**：0253-987X(2013)07-0096-06

Online Semi-Supervised Web Video Classification via Heterogeneous Attribute Fusion

DU Youtian[1], XIN Gang[2], ZHENG Qinghua[1,2]
(1. MOE Key Laboratory for Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China;
2. Shaanxi Province Key Laboratory for Satellite and Terrestrial Networks, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Learning large scale of web video data requires considering unlabeled data and heterogeneous information. A novel online semi-supervised learning method is proposed for the web video classification, which adopts graphs as base classifiers on each view of texts and videos, and propagates labels by linear neighborhood propagation algorithm. The unlabeled data are chosen online with co-training strategy on multiple graphs and base classifiers are incrementally updated. The proposed method increases the classification accuracy and is suitable for online semi-supervised learning of large scale of web video data. Experimental results show that the average accuracy of this method is approximately 8.3% higher than support vector machines, and the accuracy of learners after the online incremental learning increases by approximately 3%.

Keywords: web video; heterogeneous attribute; semi-supervised classification

网络视频是互联网上一类重要的数据，具有数据规模大、异构信息并存、多为无标记数据等特点。因此，综合利用视觉、文本等异构信息对网络视频进行在线半监督分类[1]非常必要。近年来，文献[2-7]基于多模态的信息融合对网络视频分类进行了研究。Yang等人利用底层特征、语义特征、音频特征及附加文本等信息进行视频分类[2]，结果表明，多模态分类结果优于单模态，并且支持向量机(SVM)的效果最好。Cui等人抽取训练集上的视觉特征用来辅助文本信息中词语相似度的计算[3]，避免了分类

收稿日期：2012-11-15。 作者简介：杜友田(1980—)，男，讲师。 基金项目：国家杰出青年基金资助项目(6970025)；国家自然科学基金资助项目(60905018)；国家"十二五"科技支撑计划重点课题(2011BAK08B02)。
网络出版时间：2013-04-22 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20130422.0925.003.html

时抽取视觉特征，提高了分类速度和性能。Zhang 等人使用语义概念模型进行特征定义，利用文本蕴含的语义信息提高了分类性能，并使用增量式 SVM 进行分类[4]，但计算过程较为复杂。为了更好地利用数据之间的关系，Wu 等人综合使用网络视频的标题与标签、相关视频信息、视频上传者个人兴趣等信息来提高分类性能[5]。Chen 等人利用搜索引擎返回的结果来辅助网络视频分类[6]。以上工作均侧重于监督分类，不能有效地进行在线模型更新，而半监督学习能够有效改善已有方法存在的问题[8-11]。

本文提出了一种基于多图协同训练的网络视频在线半监督分类方法，先提取了文本和视觉 2 个视图上的特征，对每个视图上的特征向量，采用图作为基分类器进行建模，并采用线性邻域传播算法[8]对每个视图进行类别标签的传播，得到了该视图上的类别预测结果。在不同视图之间，采用协同训练策略[9]在线选取未标记样本，来扩充训练集并增量地更新基分类器。对于不同模态预测结果的融合，提出了一种针对类别相关的融合方法。该方法采用了局部学习方法对数据进行建模和学习，能够得到比全局学习更好的分类准确性[12]。

1　网络视频在线半监督分类框架

如图 1 所示，该框架包括多模态特征的抽取和在线半监督分类。多模态特征提取主要包括对文本特征和视觉特征进行提取，在线半监督分类中，每种模态看作一个视图并采用图作为基分类器，然后通过协同训练策略标记、选取未标记样本来在线更新学习器。具体步骤如下：①在初始训练集上采用线性邻域传播算法（LNP）进行基分类器的学习；②在线学习过程中，按照协同训练策略，用文本（视觉）视图对未标记样本进行预测，选择预测置信度高且具有代表性的样本作为好的样本提供给视觉（文本）视图；③在扩充的训练集上对分类器进行增量学习并提升性能；④利用学习到的分类器对测试样本进行分类。在实际应用中，在线学习和分类这 2 个过程可同时进行。

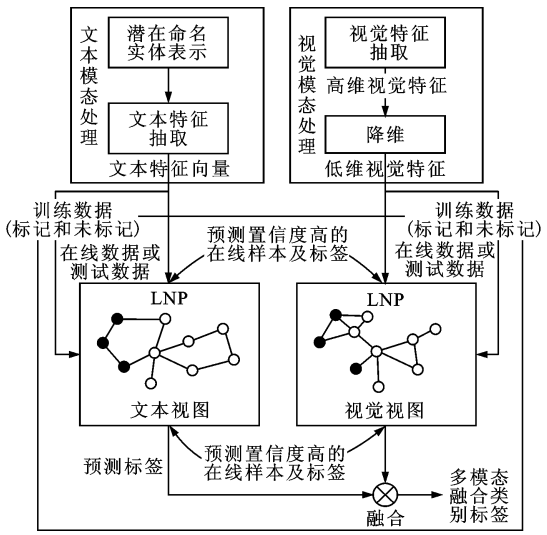


图 1　网络视频的自动分类框架图

2　在线半监督分类模型

2.1　基于图的半监督基分类器

在每个视图上，将样本作为顶点，样本之间的相似度作为边，构建图 $G=(V,\boldsymbol{W})$，其中集合 V 包含了初始的 l 个标记样本的集合，即 $L=\{(\boldsymbol{x}_1,\boldsymbol{y}_1),(\boldsymbol{x}_2,\boldsymbol{y}_2),\cdots,(\boldsymbol{x}_l,\boldsymbol{y}_l)\}$，以及 u 个未标记样本集合 $U=\{\boldsymbol{x}_{l+1},\boldsymbol{x}_{l+2},\cdots,\boldsymbol{x}_{l+u}\}$，其中：输入样本向量 $\boldsymbol{x}_i\subset\mathbf{R}^d$，$d$ 为维数；类别标签向量 $\boldsymbol{y}_i=[y_{ic}]_{1\times C}\in\mathbf{R}^C$，若 $\boldsymbol{x}_i$ 属于第 c 类，$y_{ic}=1,c=1,2,\cdots,C$；$y_{ic'}=0(c'\neq c)$。

假定图上每个 $\boldsymbol{x}_i$ 能由 K 个领域点 $\boldsymbol{x}_{i_j}\in N(\boldsymbol{x}_i)$ 线性重建，重建权重系数可通过最小化重建误差获得

$$\left.\begin{aligned}&\varepsilon_i=\Big\|\boldsymbol{x}_i-\sum_{j:\boldsymbol{x}_{i_j}\in N(\boldsymbol{x}_i)}w_{ii_j}\boldsymbol{x}_{i_j}\Big\|^2\\&\text{s.t.}\sum_{j:\boldsymbol{x}_{i_j}\in N(\boldsymbol{x}_i)}w_{ii_j}=1,\ w_{ii_j}\geqslant 0\end{aligned}\right\}\tag{1}$$

式中：w_{ii_j} 为重建权重系数；$\boldsymbol{x}_i$ 的权重向量 $\boldsymbol{w}_i=[w_{ii_1},w_{ii_2},\cdots,w_{ii_K}]^{\mathrm{T}}$；权重矩阵 $\boldsymbol{W}=[w_{ij}]_{n\times n}$。通过迭代的线性邻域传播[8]，计算输入样本的预测值

$$\boldsymbol{f}_i^t=\sum_{i_j:\boldsymbol{x}_{i_j}\in N(\boldsymbol{x}_i)}w_{ii_j}\boldsymbol{f}_{i_j}^{t-1}\tag{2}$$

式中：$\boldsymbol{f}_{i_j}^{t-1}$ 为第 $t-1$ 次迭代时 $\boldsymbol{x}_j$ 的预测值。将 $\boldsymbol{W}$ 分解为

$$\boldsymbol{W}=[\boldsymbol{W}_{ll},\boldsymbol{W}_{lu};\boldsymbol{W}_{ul},\boldsymbol{W}_{uu}]$$

式中：$\boldsymbol{W}_{ll}$、$\boldsymbol{W}_{lu}$、$\boldsymbol{W}_{ul}$、$\boldsymbol{W}_{uu}$ 为 4 个子矩阵。在 U 上的预测结果收敛为

$$\boldsymbol{F}_u=(\boldsymbol{I}-\boldsymbol{W}_{uu})^{-1}\boldsymbol{W}_{ul}\boldsymbol{Y},\quad \boldsymbol{Y}=[y_{ic}]_{n\times C}\tag{3}$$

式中：$\boldsymbol{Y}$ 为标签矩阵；$\boldsymbol{I}$ 为单位矩阵。

2.2　基分类器的增量式在线学习

在线学习中，训练集 $X_0=L\cup U$，每次增加新样本后，式(2)应在全部样本上重新计算，由于计算复杂度高，因此不适合在线学习。修改后的权重向量

$$\hat{\boldsymbol{w}}_i=\Phi_i\boldsymbol{w}_i,\quad \Phi_i=\mathrm{diag}(\phi_{i1},\phi_{i2},\cdots,\phi_{iK})$$
$$\phi_{ij}=\exp(-(d_{ii_j}-d_{\min})^2/\rho\sigma^2)\tag{4}$$

式中：d_{ii_j} 为 $\boldsymbol{x}_i$ 到近邻点 $\boldsymbol{x}_{i_j}$ 的距离；$d_{\min}=\min_{i,j}d_{ii_j}$；

σ^2 为样本距离的方差；ρ 为控制参数，本文取 $\rho=10$。

由式(2)、式(4)可知，类别标签传播具有衰减特性，其优点为标签预测可体现出置信度，将标记错误的样本引入的噪声限制在一个局部区域内，这样有利于实现模型的增量更新。假定数据集只含 1 个标记样本向量$(\boldsymbol{x},\boldsymbol{y})$，经过标签传播后 $\boldsymbol{x}_i$ 得到的标签向量值为 $\boldsymbol{f}_i=[f_{i1},f_{i2},\cdots,f_{iC}]$。由式(2)、式(4)可知

$$f_{ic}=\sum_{i_j:\boldsymbol{x}_{i_j}\in N(\boldsymbol{x}_i)}\hat{w}_{ii_j}f_{i_jc}\leqslant(\max_{i_j}f_{i_jc})\cdot\sum_{i_j}\varphi_{ij}w_{ii_j}=\delta_i f_{ic}^{*},\quad c=1,2,\cdots,C;\tag{5}$$

$$\delta_i=\sum_{ii_j}\varphi_{ij}w_{ii_j}<1;\quad f_{ic}^{*}=\max_{i_j}f_{i_jc}$$

重复式(5)，可得从 $\boldsymbol{x}$ 到 $\boldsymbol{x}_i$ 的路径 s_i，s_i 中第 1 个结点为 $\boldsymbol{x}_{s_i(1)}=\boldsymbol{x}$，第 m 个节点的序号为 $s_i(m)=\operatorname{argmax}_{k:\boldsymbol{x}_k\in N(\boldsymbol{x}_{s_i(m-1)})}f_{kc}$。在该路径上，以下不等式成立

$$f_{ic}\leqslant y_c\prod_{k\in s_i}\delta_k\leqslant y_c(\max_k\delta_k)^{l_i}\leqslant y_c(\max_k\delta_k)^{l_i^{*}}\tag{6}$$

式中：s_i^{*} 为从 $\boldsymbol{x}_i$ 到 $\boldsymbol{x}$ 的最短路径；l_i 为路径 s_i 的长度；l_i^{*} 为 s_i^{*} 的长度；y_c 为 $\boldsymbol{y}$ 的第 c 个分量。式(6)表明标签传播以指数函数为上界，将标签传播限制在 $\boldsymbol{x}$ 周围的局部区域 $L_R(\boldsymbol{x})$。训练集增加新标记样本$(\boldsymbol{x}_a,\boldsymbol{y}_a)$时，本文通过如下公式来进行增量更新

$$\boldsymbol{F}_{u,t}=[\boldsymbol{F}_{u,t-1};\boldsymbol{f}_a]+\boldsymbol{F}'_{u,t}$$

$$\boldsymbol{F}'_{u,t}=(\boldsymbol{I}_t-\boldsymbol{W}_{uu,t})^{-1}\boldsymbol{W}_{ul,t}[\boldsymbol{0};\boldsymbol{y}_a-\boldsymbol{f}_a]$$

$$\boldsymbol{f}_a=\sum_{i:\boldsymbol{x}_i\in N(\boldsymbol{x}_a)}w_{ai}\boldsymbol{f}_i\tag{7}$$

式中：$\boldsymbol{f}_a$ 为 $t-1$ 时刻基于训练集 X_{t-1} 对 $\boldsymbol{x}_a$ 的预测。当训练集加入新标记样本时，标签预测只需在 $L_R(\cdot)$内进行增量计算。图 2 表明了增量更新与在全部数据上进行计算得到的结果非常接近。

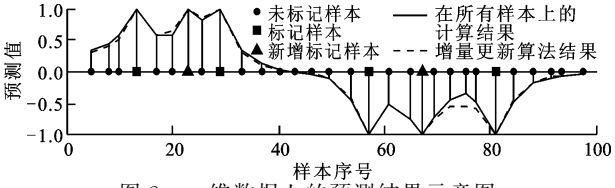


图 2 一维数据上的预测结果示意图

2.3 在线半监督分类

本文的在线半监督分类算法采用了协同训练框架，在线处理过程中收集的未标记样本集记为 U_p。在线半监督学习过程为：①通过文本和视觉视图上的基分类器对 U_p 中的样本分别进行预测和标记；②根据预测结果选取好的未标记样本，其预测标签添加至对方训练集；③在新的训练集上进行基分类器增量更新。

2.3.1 U_p 上的类别标签预测 给定 $\boldsymbol{x}_i\in U_p$，视图 $v(v=1,2)$的标签预测值为

$$\boldsymbol{f}_i^{(v)}=\sum_{j:\boldsymbol{x}_{i_j}\in N(\boldsymbol{x}_i)}w_{ii_j}^{(v)}\boldsymbol{f}_{i_j}^{(v)}\tag{8}$$

式中：$\boldsymbol{x}_{i_j}$ 为 $\boldsymbol{x}_i$ 在训练集 X_t 中的近邻点。根据式(8)计算时不需要改变整个图的结构，因此降低了计算的复杂度。

2.3.2 未标记样本的选取 如何选取未标记样本来扩充训练集，是半监督学习的核心问题，好的未标记样本应满足：①标签预测的置信度高，若加入预测错误的样本相当于引入了大的噪声；②在特征空间中的分布合理，若增加的未标记样本不能较好地代表整个样本分布，则分类器性能的提高会较慢，且推广性较差。

根据式(6)可知，若式(8)中 $\boldsymbol{f}_i^{(v)}$ 的分量 $f_{ic}^{(v)}$ 较大，说明该样本距离与已标记样本较接近。根据流形假设，$\boldsymbol{x}_i$ 的类别与标记样本 $\boldsymbol{x}$ 很相似，因此式(8)中的标签预测结果蕴含了标签预测结果的置信度。例如 $f_{ic}^{(v)}=0.95$ 时，说明 $\boldsymbol{x}_i$ 在 v 上被判成第 c 类的概率较高。另一方面，比 $f_{ic}^{(v)}$ 大的样本通常都靠近标记样本，因此从 U_p 中选取预测值最大的若干样本不利于将类别信息尽快扩展至整个特征空间。如图 3 所示，按如下方法在视图 1 上选取好的样本提供给视图 2，将 U_p 中样本第 c 类的预测值 $f_{ic}^{(1)}$、$f_{ic}^{(2)}$ 分别从大到小排序，取 $f_{ic}^{(1)}$ 最大且 $f_{ic}^{(2)}$ 不在前 5%的 $m_{1,c}$个样本中。其中，图 3 中的黑色和白色结点为 $L\cup U$ 中的样本，灰色结点 $\boldsymbol{x}_4^{(1)}$、$\boldsymbol{x}_4^{(2)}$、$\boldsymbol{x}_5^{(1)}$ 和 $\boldsymbol{x}_5^{(2)}$ 为 U_p 中的样本。在视图 1 上，$\boldsymbol{x}_4^{(1)}$ 和 $\boldsymbol{x}_5^{(1)}$ 的预测较很大，但在视图 2 上，$\boldsymbol{x}_4^{(2)}$ 的预测值偏中等，$\boldsymbol{x}_5^{(2)}$ 的预测值较大。从几何角度看，在视图 2 上引入 $\boldsymbol{x}_4^{(2)}$ 比引入 $\boldsymbol{x}_5^{(2)}$ 使得训练集的分布更符合样本的真实分布。从视图 2 选取好的样本，提供给视图 1 的方法也类似。

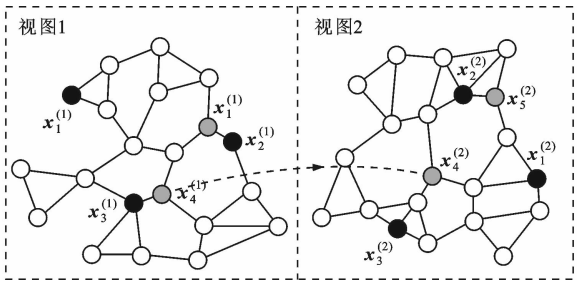


图 3 样本选取示意图

在线学习算法的描述如下：

(1)G_1、G_2 分别是文本和视觉视图对应的基分类器，标记样本 L、未标记样本 U、$X_0=L\cup U$ 为初始训练集；R_1、R_2 是 G_1、G_2 分别挑选出来的好的在

线样本集，初始时 $R_1=R_2=\varnothing$。

(2)由训练数据 X_0 的文本和视觉数据分别构图，计算边权重，根据线性邻域传播算法对样本类别标签进行预测，得到 G_1、G_2。

(3)在线获取的 $U_p=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_p\}$，分别计算文本和视觉特征向量集合，记为 $U_p^{(1)}$ 和 $U_p^{(2)}$。

(4)利用 G_1 对 $U_p^{(1)}$ 进行预测得到的结果 $\boldsymbol{F}^{(1)}=[\boldsymbol{f}_1^{(1)},\boldsymbol{f}_2^{(1)},\cdots,\boldsymbol{f}_p^{(1)}]$，以及 G_2 对 $U_p^{(2)}$ 进行预测得到的标签 $\boldsymbol{F}^{(2)}=[\boldsymbol{f}_1^{(2)},\boldsymbol{f}_2^{(2)},\cdots,\boldsymbol{f}_p^{(2)}]$，融合样本的类别标签得到结果输出。

(5)根据 $\boldsymbol{F}^{(1)}$、$\boldsymbol{F}^{(2)}$，从 2 个视图上分别选择多个未标记样本，记为 R_1 和 R_2。将 R_1、R_2 分别加入 G_2、G_1 的训练数据集，确定每个视图上进行模型更新的局部区域 $L_R(\cdot)$，按式(7)在 $L_R(\cdot)$ 内进行增量式更新基分类器。

2.4　类别相关的多模态融合

不同模态具有不同的分类能力，区分能力强的模态对类别标签的预测置信度相对较高，同时同一模态对于不同类别数据的分类能力也不相同。因此，融合权重系数应是类别相关的，本文基于 $F1$ 值来确定该系数，$F1=2pr/(p+r)$，其中 p、r 分别表示准确率和召回率。

给定验证集 $V\equiv(V_1,V_2)$，V_1、V_2 分别对应于文本视图和视觉视图。根据 G_1、G_2 分别对 V_1、V_2 进行预测，结果为 $F_{11}=(f_{s1}^1,f_{s2}^1,\cdots,f_{sC}^1)$，$F_{12}=(f_{s1}^2,f_{s2}^2,\cdots,f_{sC}^2)$，其中 f_{si}^1 和 f_{si}^2 分别是 G_1、G_2 分类得到的第 i 类样本对应的 $F1$ 值。设 G_1、G_2 的权重系数向量分别为 $\boldsymbol{W}_1=(w_1^1,w_2^1,\cdots,w_C^1)$、$\boldsymbol{W}_2=(w_1^2,w_2^2,\cdots,w_C^2)$，本文确定的权重系数

$$w_c^1=\frac{f_{sc}^1}{f_{sc}^1+f_{sc}^2};\quad w_c^2=\frac{f_{sc}^2}{f_{sc}^1+f_{sc}^2} \tag{9}$$

G_1 对 $\boldsymbol{x}_i$ 的预测结果为

$$\hat{\boldsymbol{y}}_i^{(1)}=(f_{i1}^{(1)},f_{i2}^{(1)},\cdots,f_{iC}^{(1)})$$

G_2 的预测结果为

$$\hat{\boldsymbol{y}}_i^{(2)}=(f_{i1}^{(2)},f_{i2}^{(2)},\cdots,f_{iC}^{(2)})$$

融合后的预测结果为

$$\hat{\boldsymbol{y}}_i=w_1^1\,f_{i1}^{(1)}+w_1^2\,f_{i1}^{(2)},\cdots,w_C^1\,f_{iC}^{(2)}+w_C^2\,f_{iC}^{(2)}$$

$\boldsymbol{x}_i$ 所属的类别为

$$\operatorname{argmax}(w_c^1\,f_{ic}^{(1)}+w_c^2\,f_{ic}^{(2)}),\quad 1\leqslant c\leqslant C$$

3　实验结果及分析

3.1　实验数据集

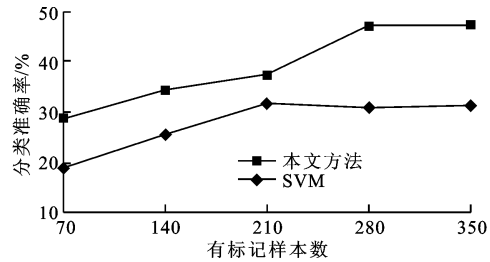
由 Youtube 数据构建了网络视频数据集(MCG-WEBV)[13]，其视觉特征包括颜色直方图、颜色矩及边缘直方图等。本文信息采用词袋模型表示，并选取了 7 个类别、29 437 个网络视频作为实验数据，如表 1 所示。

表 1　实验数据集

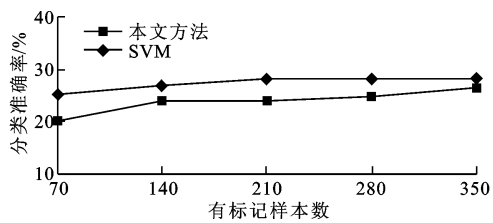
类别序号	视频类别	样本数	样本数比例/%
1	Entertainment	6 907	23.46
2	Music	5 186	17.62
3	News&Politics	4 638	15.76
4	Sports	3 892	13.22
5	Gaming	3 249	11.04
6	Autos&Vehicles	3 213	10.91
7	Pets&Animals	2 352	7.99

3.2　基分类器的分类结果及分析

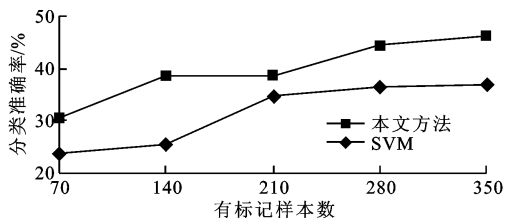
在少量有标记样本的情况下，对 SVM 与本文方法进行了实验对比(见图 4)。将数据划分为有标记样本集与未标记样本集 2 部分，未标记样本集有 2 964个数据，标记样本集由大到小分别为 70、140、210、280 和 350，2 种方法均在有标记样本集上进行训练，并在未标记样本集上进行预测。SVM 通过 Libsvm 工具包[14]实现，对于文本数据、视觉数据分别采用线性核函数和径向基(RBF)核函数，通过实验确定，局部邻域 $L_R(\cdot)$ 的大小为 3.5 kB。利用式



(a)文本视图



(b)视觉视图



(c)两视图融合后的结果

图 4　本文方法与 SVM 的分类性能对比结果

(1)构建图时,文本视图采用向量夹角余弦作为距离度量,视觉视图采用欧氏距离度量,实验结果如图 4 所示。由图 4 可见,在少量有标记样本的状态下,本文方法在文本视图上显著优于 SVM,但在视觉视图上二者的性能相近。当文本、视觉 2 个视图融合后,本文方法的分类性能指标高于 SVM 约 8.3%。

3.3 在线学习的有效性验证

将分类器在线更新及未标记样本对分类性能的影响进行如下的分析。

3.3.1 *分类器在线更新对分类性能的影响* 对分类器在线更新与离线不更新的方式进行在线数据的预测性能实验,将实验数据按月份划分为训练数据与在线数据 2 部分。训练数据包含第 1、第 2 个月的数据,在线数据包含从第 3 到第 8 个月的所有数据,其中每个月的在线数据都随机划分出一部分作为测试集,用以测试在线学习时分类器对当月数据(即在线数据)的预测性能。在线学习共进行了 80 次,得到的实验结果如图 5 所示。

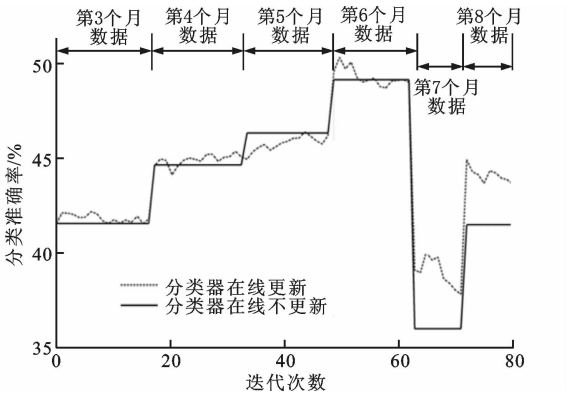


图 5 多模态融合后分类器的在线更新结果

由实验结果可见,分类器在线更新与离线不更新方式的性能在开始几个月相差很小,大致保持一致,但随着时间的延续,在后几个月它们之间的性能差异较为明显。就性能变化趋势而言,分类器在线更新后逐渐表现出了更好的分类性能。

3.3.2 *未标记样本对分类性能的影响* 实验分析了下面 3 种未标记样本的选择方式对在线学习的性能影响:①在线学习时随机增加有标记样本至训练集;②在线学习时选择预测置信度高的未标记样本及其预测标签扩充训练集;③在线学习时随机选择未标记样本的训练集。第②种方式即为本文在线学习所采用的方法,第①种方式相当于监督学习方法,第③种方式相当于分类器完全不预测,随机选择未标记样本扩充训练集。

实验中在线学习了 80 次,所得结果如图 6 所示。在线学习伊始,由于每种方式下的分类器和训练数据差异很小,所以性能差异不大,但随着在线更新的不断进行,分类器性能差异逐渐增强。由图 6 可以看出,第②种方式的性能介于第①、③种方式之间,在第 7 个月和第 8 个月中,方式②要比方式③有较明显的提高,说明本文方法能够在线提升分类器的性能。该性能的提高与基分类器有密切联系:若基分类器的准确性高和推广性好,则在线半监督学习有较好的性能提升。

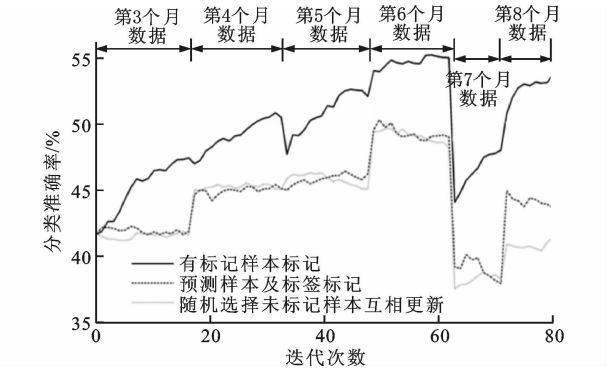


图 6 3 种样本选取方式在多模态融合下的数据变化曲线

3.4 类别相关的多模态融合方法实验

将相同权重系数与类别相关权重系数进行对比,实验所用的训练集共有样本 6 157 个,其中包含 2 894 个有标记样本和 3 263 个未标记样本,用来计算类别相关的融合权重系数的验证集包含有标记样本 2 895 个,测试集样本 2 484 个。如表 2 所示,类别统一融合权重(10%,20%,…,90%)通过在网格上搜索确定,使得融合后的分类性能最高,最终确定视觉和文本的权重均为 50%。

表 2 中 W_1、W_2 分别为文本、视觉视图的权重。从实验结果中可以看到,2 种融合方法得到的结果均比单一视图的分类结果有较大的提高(见表 2 平均值),性能提高幅度均在 10%以上。其次,类别相关权重的融合方法比类别统一权重有一定提高(平

表 2 2 种权重的融合性能对比 %

视频类别	类别统一权重			类别相关权重		
	W_1	W_2	$F1$	W_1	W_2	$F1$
Entertainment	50	50	36.66	50.91	49.09	36.77
Music	50	50	53.37	53.27	46.73	53.14
News&Politics	50	50	40.49	58.10	41.90	40.21
Sports	50	50	63.22	54.16	45.84	63.13
Gaming	50	50	42.82	55.90	44.10	45.01
Autos&Vehicles	50	50	64.20	58.35	41.65	64.46
Pets&Animals	50	50	45.56	52.80	47.20	46.28
平均值	50	50	49.47	54.78	45.22	49.86

均高约 0.8%）。实验结果表明，文本模态比视觉模态有更好的区分能力，在融合时也占有比较大的权重。在实际应用中，类别相关的融合方法需要根据单一模态的预测结果来选择可能的几组权重，最后选取融合后可能性最大的类别作为最终结果。

4　结　论

为了解决多模态网络视频在线半监督的分类问题，本文提出了多图上的协同训练方法，所建模型采用图作为基分类器，以及带衰减的线性邻域传播来预测样本。本文方法中，视觉与文本模态对应的基分类器按照协同训练策略，对在线未标记样本进行预测、选取、互相标记，以扩充训练集，使得分类器能够在线学习，并与在线数据分布保持更好的一致性。多图上的协同训练具有如下优点。

(1)由于局部学习通常比全局学习效果更好，因此本文方法能得到更好的分类结果。

(2)将局部学习用于协同训练能够简单有效地进行增量学习，因此非常适合于大规模数据的在线学习。

实验结果表明，在少量有标记样本的情况下，本文方法优于广泛使用的 SVM 算法，与分类器离线不更新的情况相比，分类器在线学习显现出性能逐渐的提高。因此，类别相关的多模态信息融合可实现更高效的多模态信息融合。

参考文献：

[1] 刘安文，支琤，张瑞，等. 基于语义概念的视频检索系统的设计与实现 [J]. 中国图形图象学报，2008，13(10)：2055-2058.
LIU Anwen，ZHI Cheng，ZHANG Rui，et al. Design and implementation of semantic concept based viseo retrieval system [J]. Journal of Image and Graphics，2008，13(10)：2055-2058.

[2] YANG Linjun，LIU Jiemin，YANG Xiaokang，et al. Multi-modality web video categorization [C]// Proceedings of the International Workshop on Workshop on Multimedia Information Retrieval. Madison，Wisconsin，USA：ACM，2007：265-274.

[3] CUI Bin，ZHANG Ce，CONG Gao. Content-enriched classifier for web video classification [C]// Proceedings of the 33rd international ACM SIGIR Conference on Research and Development in Information Retrieval. Madison，Wisconsin，USA：ACM，2010：619-626.

[4] ZHANG Xu，SONG Yicheng，CAO Juan，et al. Large scale incremental web video categorization [C]// Proceedings of the 1st Workshop on Web-Scale Multimedia Corpus. Madison，Wisconsin，USA：ACM，2009：33-40.

[5] WU Xiao，ZHAO Wanlei，NGO C W. Towards Google challenge：combining contextual and social information for web video categorization [C]// Proceedings of the 17th ACM International Conference on Multimedia. Madison，Wisconsin，USA：ACM，2009：1109-1110.

[6] CHEN Zhineng，CAO Juan，SONG Yicheng，et al. Web video categorization based on Wikipedia categories and content-duplicated open resources [C]// Proceedings of the International Conference on Multimedia. Madison，Wisconsin，USA：ACM，2010：1107-1110.

[7] LEUNG J K W，LI C H，IP T K. Commentary-based video categorization and concept discovery [C]// Proceedings of the 2nd ACM Workshop on Social Web Search and Mining. Madison，Wisconsin，USA：ACM，2009：49-56.

[8] WANG Jingdong，WANG Fei，ZHANG Changshui，et al. Linear neighborhood propagation and its applications [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2009，31(9)：1600-1615.

[9] BLUM A，CHAWLA S. Learning from labeled and unlabeled data using graph mincuts [C]// Proceedings of the 18th International Conference on Machine Learning. Madison，Wisconsin，USA：ACM，2001：19-26.

[10] ZHOU D，BOUSQUET O，LAL T，et al. Learning with local and global consistency [C]// Advances in Neural Information Processing Systems. Denver，CO，USA：MIT Press，2004：321-328.

[11] BLUM A，MITCHELL T. Combining labeled and unlabeled data with co-training [C]// Proceedings of the Eleventh Annual Conference on Computational Learning Theory. Madison，Wisconsin，USA：ACM，1998：92-100.

[12] LEON B，VLADIMIR V. Local learning algorithms [J]. Neural Computation，1992，4(6)：888-900.

[13] CAO Juan，ZHANG Yongdong，SONG Yicheng，et al. MCG-WEBV：a benchmark dataset for web video analysis，tech rep ICT-MCG-09-001 [R]. Beijing：Institute of Computing Technology，2009.

[14] CHANG C C，LIN C J. LIBSVM：a library for support vector machines [EB/OL]. (2005-01-01) [2012-08-01] http：// www. csie. ntu. edu. tw/cjlin/libsvm，2001.

(编辑　管咏梅　赵大良)