一种基于反馈模糊图论的视频多语义标注算法

朱宇光[1,2] 闫 婷[3] 张建明[3] 杨 雄[2] 胡维礼[1]
(南京理工大学自动化学院 南京 210094)[1] (常州工学院计算机信息工程学院 常州 213002)[2]
(江苏大学计算机科学与通信工程学院 镇江 212013)[3]

摘 要 为了弥补视频语义检索中视频底层特征与高层语义概念之间的"语义鸿沟",提出了一种基于反馈模糊图论的视频多语义标注算法。该算法首先构造一个包括所有数据的时间和空间分布信息的小样本集,据此进行人工标注并将其作为训练集。然后将模糊算子引入图论中,将语义概念间的关系模糊化,以实现模糊推理。最后将标注完成的测试集中的样本加入到训练集中,以完成视频标注的反馈。实验结果表明,使用反馈的模糊图不仅可以很好地建立语义概念间的关系,还能提高视频标注的准确率,表现出良好的性能。

关键词 视频标注,模糊图,多语义标注,语义鸿沟
中图法分类号 TP391 **文献标识码** A

Video Multi-semantic Annotation Algorithm Based on Feedback Fuzzy Graph Theory

ZHU Yu-guang[1,2] YAN Ting[3] ZHANG Jian-ming[3] YANG Xiong[2] HU Wei-li[1]
(School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China)[1]
(School of Computer Information & Engineering, Changzhou Institute of Technology, Changzhou 213002, China)[2]
(School of Computer Science & Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China)[3]

Abstract For bridging semantic gap between video low-level features and high-level semantic concepts in the semantic-based video retrieval system, the video multi-semantic annotation algorithm based on feedback fuzzy graph theory was proposed. First, a training set which includes most temporal and spatial distribution of the whole data is made up and it will achieve a satisfying performance even in the case of limited size of training set. Secondly, the fuzzy operators are applied to graph theory to achieve fuzzy reasoning by using fuzzy semantic. Last, in order to finish the feedback of video annotation, some temples from the testing set that have finished annotation are selected and added into the training set. Experimental results indicate that feedback fuzzy graph not only sets up the relationship between semantic concepts well, but also improves the precision of annotation and shows good performance.

Keywords Video annotation, Fuzzy graph, Multi-semantic annotation, Semantic gap

1 引言

视频数据将呈现爆炸式增长是数字化存储技术和多媒体技术发展的必然结果。对视频数据的应用首先需在海量数据中对其进行有效的筛选和检索,而视频固有的底层特征和用户实际需求之间必然存在着差别,此即所谓的"语义鸿沟"。人们往往采取在视频里找出能对视频的内容进行语义层面解释的原始数据,再用这些特定的原始数据来缩小视频检索方法的"语义鸿沟"。通常被称为视频概念检测、高层语义特征提取的视频语义标注是得到上述原始数据的一种有效方法,简称为视频标注。它实际上是把一些与相关视频关联的语义的概念赋给视频片段的行为,可分为基于机器学习和基于人工的两种方法。基于人工的视频标注可得到非常精确的结果,但只适用于小规模的数据概念集。实现视频的自动标注一般使用机器学习的方法,可完成从视频固有的底层特征到语义概念之间的一一对应,从而使"语义鸿沟"分为从视频固有的底层特征到语义概念之间及语义概念到用户实际需求之间的更窄的鸿沟[1]。

基于机器学习的视频标注主要采用有监督学习、半监督学习和主动学习等方法。在很多实际应用中,随着数据采集技术和存储技术的发展,获取大量的无标记的样本易于实现,而获取有标记样本通常需要付出很大的代价。半监督学习[2]的主要思想是利用大量无标记数据自发地学习出数据的内在结构规律,再利用无标记数据与少量有标记数据之间的关系或相似度来进行指导以期达到更好的学习效果,因此,半监督学习成为模式识别和机器学习中的重要方法。文献[3]提出了基于图学习的SSMR方法来对视频语义概念进行检测,它将半监督学习的结构性假设融入到现存的相似性度量的计算之中;文献[4]使用基于图的直推式学习方法对视频进行多语义概念的标注。

到稿日期:2013-03-01 返修日期:2013-05-15 本文受江苏省高校自然科学研究面上项目(11KJD520002),常州市科技计划项目(CC20120030)资助。

朱宇光(1966-),男,硕士,副教授,主要研究方向为光电通信技术、下一代网络技术、智能控制技术,E-mail:zyg_victor@sina.com。

本文提出了一种基于反馈模糊图论的视频多语义标注算法，首先根据分类器的泛化性能构造一个样本数目小但是能代表整个样本集分布的最优训练样本集；然后将模糊算子引入图中，将语义概念之间的关系模糊化以便实现模糊推理；最后选出一部分标注完成的测试集中的样本加入到训练集中，完成视频标注的反馈，以提高视频标注的准确率。模糊图是图的一种扩展，实验结果表明，本文提出的算法可以很好地建立语义概念间的关系，运用到视频标注中时表现出了良好的性能。

2 优化训练样本集的构造

文献[3,4]提到的关于视频标注的学习方法忽略了一个重要的问题：训练样本集的构造。这些算法都随机选择样本，没有代表性。这样，为保证分类器的泛化性能就需要很多的样本，需要大量的人工标注，这无疑是一项费时费力的工作。事实上分类器的泛化性能不取决于训练样本的数量而取决于样本的分布，如果能构造出一个训练样本集，这个样本集的数目相对整个样本集要小，却能逼近整个视频样本集的分布，这就能节省很多人工标注的劳动，且能获得好的分类性能。优化的训练样本集的构造分为两步：(1)对视频镜头进行预聚类；(2)采用改进的 K-means 算法处理预聚类结果，获取聚类中心，从中挑出样本构造样本集。

2.1 视频镜头的预聚类

给定一个视频集合，通过探索整个视频集在空间和时间上的分布，构造一个样本数量小但是有效的训练样本集，并将其提供给人工标注。理论上讲，一个语义概念和它相关的特征的变化在一个相同的视频中要小于不同视频间的变化。因此，可以根据这个理论来观测并提取出聚类信息，也即，基于时间顺序和视觉相似度，视频镜头可以通过一种过分割的形式进行预聚类。

2.1.1 视频镜头的相似性度量

预聚类主要是根据镜头之间的视觉相似度以及它们在时间上的相关性，以过分割的形式进行的。过分割的目的在于使所得聚类中每个镜头表示的语义概念尽可能一致。根据文献[5]中方法，定义镜头之间的相似度如下。

定义 1 给定两个镜头 $S_i=\{f_{i1},f_{i2},\cdots,f_{iK}\}$ 和 $S_j=\{f_{j1},f_{j2},\cdots,f_{jM}\}$。如果镜头 S_i 和 S_j 中分别存在着两个相似帧 f_{ik} 和 f_{jm}（帧 f_{ik} 在镜头 S_i 中，帧 f_{jm} 在镜头 S_j 中），则认为这两个镜头是相似的。

定义 2 镜头 S_i 和 S_j 之间的特征差异。

$$D(S_i,S_j)=\min_{\substack{1<k<K\\1<m<M}} d(f_{ik},f_{jm})$$

式中，$d(.,.)$ 是任意两帧之间的差异。当镜头 S_i、S_j 之间的特征值小于某个阈值时，则认为它们是相似的。

从定义 2 可以看出，对公式的直接计算涉及到镜头中的每一帧，因此计算量很大。而镜头检测主要是比较相邻帧之间的相似度，由于同一镜头中的各帧特征比较相似，因此存在着很大的冗余。为了解决上述问题，可以利用关键帧技术，以数量小的关键帧集合替代镜头中的全部帧。实验证明，这种做法不会降低性能。

2.1.2 视频镜头的时间相关性

上文考虑了镜头之间的相似度，另外一个因素就是概念间的时间相关性。由于在计算机视觉的研究领域中存在不足，从视频中提取的特征无法很好地表征概念，还存在着一些问题，即对于时间上相隔很远的两个镜头，即使它们之间的视觉相似度很高，也还是不能确定它们表达的是否是同一个语义概念，也就是存在着所谓的“语义鸿沟”，亦即：不同的概念在其对应的特征空间中分布可能会比较相似，相同概念的分布却差别很大。对于这种情况，我们将镜头间的时间相关性考虑到视觉相似度量中，也就是引入一个时间窗参数 T，这样得到如下定义。

定义 3(镜头间的相似性度量)

$$D(S_i,S_j)=\begin{cases}\min\limits_{\substack{1<k<K\\1<m<M}} d(f_{ik},f_{jm}), & \text{if } i-j\leqslant T\\ \infty, & \text{otherwise}\end{cases}$$

通过这种方法，我们可以得到视频的预聚类结果。

2.2 改进的 K-means 算法处理预聚类结果

聚类是分析原始数据并从中发现有用信息的一种方法。它将数据对象分组成多个类，每个类中的对象有很高的相似度，不同的类中的对象千差万别。K-means 算法是一种基于划分的方法。其第一步为随机选取 k 个对象的过程，目标为数据集；选取完成后，以被选对象为原始的聚类中心，然后将数据集中剩下的所有对象一一安排至离该聚类中心最近的数据簇中从而得到新的聚类；接着对每个新的聚类中的所有数据对象计算其平均值，从而得到新聚类中心；将 3 个步骤不断迭代至定义 4 的准则函数收敛。

定义 4(平方误差总和)

$$E=\sum_{i=1}^{k}\sum_{p\in C_i}|p-m_i|^2$$

式中，E 为数据集所包含的每个数据对象与聚类中心之间的平方误差的总和；p 为对象；m_i 为聚类中心。

在 K-means 算法中，选择不同的初始点可能获得不同的聚类结果，也就是说聚类结果对初始点有一定的依赖性。为减少依赖性，提高结果的稳定性，采用改进的 K-means 聚类算法。

改进的 K-means 聚类算法在搜索过程中随机选择样本，但要尽量使选择的样本对象既不失真，又能保持原始数据的分布特征。对取样后的样本和原始的样本分别进行 K-means 算法，得到的聚类中心的位置差别不大，这样可以证明改进的 K-means 算法的有效性。为减少选择的样本对初始聚类中心产生的影响，采取 J 次选样，这 J 次选择的样本的集合应尽量等于原始样本集的数目。选样后，分别对这 J 组样本集利用 K-means 算法进行聚类，由此获得 J 组聚类中心，并对其进行聚类准则函数值的比较，确定函数值最小的那组聚类中心作为初始的聚类中心。在算法中，因为采用聚类准则函数，可能会将大的聚类簇分割成小的聚类簇。为解决上述问题，可以设定初始聚类数为 $K'(K'>K)$，这样较大的 K' 值就可以扩大解的搜索范围。对初始的聚类中心采用 K-means 聚类算法，聚类输出 K' 个聚类中心，观察这些聚类中心彼此间的距离，合并那些聚类中心距离值小的聚类簇，直到聚类簇的值达到 K 为止。

改进的 K-means 算法过程如下：

(1)从数据集中取样，得到 J 组样本集 $\{S_1,S_2,\cdots,S_J\}$；

(2)对 J 组样本集分别执行 K-means 算法，获得 J 组 K' 个聚类中心；

(3)对聚类中心根据聚类准则函数计算，选取函数值最小

的那组聚类中心作为初始的聚类中心；

(4)设定初始聚类数 K'，再次执行 K-means 算法；

(5)观察聚类中心间的距离，合并距离最近的一组，重新计算合并后的聚类中心，迭代操作，直到聚类簇的值达到 K 为止。

改进的 K-means 算法在确定初始聚类中心时，计算量比较小，迭代次数少，并且在确定聚类中心时表现出稳定性。将 K-means 算法应用到上文预聚类之后的视频中，可以获得更优的聚类中心。从聚类中心中随机提取样本，将其作为要进行人工标注的训练样本集。

3 基于反馈模糊图的视频多语义标注框架

视频标注本质上是根据视频片段所体现的内容对其进行语义概念上的分类。一般对视频进行有监督分类时，都是先由专家指定若干类别，然后手工标注训练样本集产生分类器，再将待标注的视频提取底层特征放入分类器中进行分类，按照某种规则将未标注的视频分类到最相似的一类中，并将此类的名称作为该视频的语义标注。这种方法存在不足，是因为这些方法设置的类别有排他性，视频只能被标注一个语义概念。而现实世界中的视频内容都很丰富，一个语义概念不能清晰表达视频的内容，一个视频往往包含多个语义概念。如在拍摄的风景视频中有蓝天、白云、山脉，只用一个语义概念表示该视频是不准确的，并且语义概念间是有相关性的，比如“蓝天”一般伴随出现“白云”。同时考虑到语义概念间的相关性是有向的，在一幅图像中，如果检测到一个语义概念为“荒漠”，通过实例可以验证再检测到“天际”的概率会很大；而检测到“天际”这个语义概念后能再检测到“荒漠”的概率非常小。从而可得结论，即：“荒漠”这个语义概念有效于“天际”这个语义概念，而“天际”这个语义概念基本无益于检测到“荒漠”这个语义概念。因而可以得出方向性对于视频语义概念标注至关重要。依靠模糊图论可以将上述这些情况用数学形式表述出来。

3.1 模糊图的概念

模糊图是对一般图模糊化得到的，本质上是一个赋权图。根据文献[6]在一个论域上可以由一个三元组 $\tilde{G}=(V,\tilde{V},\tilde{E})$ 组成模糊图，$V=\{v_1,v_2,\cdots,v_n\}$ 为由 n 个点组成的一个集合(或称为一个论域)，$\tilde{V}$ 为包含于集合 V 中的一个模糊子集，且隶属函数为 $\mu_i(v_i)$，$i=1,2,\cdots,n$，表示结点 v_i 的模糊度，$\tilde{E}$ 为一模糊关系，隶属于论域 $V\times V$，用以下矩阵表示：

$$\tilde{E}=\begin{bmatrix}\mu_{11} & \cdots & \mu_{1n}\\ \mu_{21} & \cdots & \mu_{2n}\\ & \cdots & \\ \mu_{n1} & \cdots & \mu_{m}\end{bmatrix}$$

其中，$0\leqslant\mu_{ij}\leqslant 1(i=1,2,\cdots,n;j=1,2,\cdots,n)$。在 $\mu_{ij}\neq 0$ 情形下，v_i 和 v_j 这两个结点之间存在一连接边，μ_{ij} 即定义为 v_i 和 v_j 结点间的连接强度，或其连接边的模糊度。$\mu_{ij}=0$ 表示结点间没有边连接，$\mu_{ij}=1$ 表示结点间的有边连接的状态是完全明确的，$0<\mu_{ij}<1$ 表示结点间有连接强度为 μ_{ij} 的模糊边，$\tilde{E}$ 是对称的，即 $\mu_{ij}\neq\mu_{ji}$。

上述模糊图没有考虑到边的方向性，上文已经讲到方向性是语义概念间的一个重要因素。将方向性考虑到模糊图中可以定义有向模糊图。

定义 5 设在上文模糊图的定义中，结点 v_i 和结点 v_j 之间最多有 m 条边，并且每条边都有方向性，即从 v_i 指向 v_j，或者从 v_j 指向 v_i，将模糊图定义中的矩阵改为广义模糊矩阵，如下：

$$\tilde{H}=\begin{bmatrix}\mu_{11} & \cdots & \mu_{1n}\\ \mu_{21} & \cdots & \mu_{2n}\\ & \cdots & \\ \mu_{n1} & \cdots & \mu_{m}\end{bmatrix}$$

其中，$\mu_{ij}(i=1,2,\cdots,n;j=1,2,\cdots,n)$ 是元素个数不超过 l 的隶属度的集合，$\mu_{ij}=\{\mu_{ij1},\mu_{ij2},\cdots,\mu_{ijk}\}$，$0\leqslant\mu_{ijk}\leqslant 1(k=1,2,\cdots,l)$，则称 $\tilde{G}=(V,\tilde{V},\tilde{H})$ 为一个论域 V 上的具有多重性质的基于方向模糊图。

在定义完基于方向模糊图之后，我们给出模糊图的 λ-截图定义。

定义 6 设 $\tilde{G}$ 是一个基于方向模糊图，$\tilde{G}=(V,\tilde{V},\tilde{H})$，$V=\{v_1,v_2,\cdots,v_n\}$，其中 $\tilde{V}:\mu_{\tilde{V}}(v_i)$，$v_i\in V$，$\tilde{E}:\mu_{\tilde{E}}(v_i,v_j)=\mu_{ij}$，$v_i,v_j\in V(i,j=1,2,\cdots,n)$。

对于一个任意 $\lambda\in[0,1]$，$\tilde{G}$ 的 λ-截图可定义为 G_λ，G_λ 本身不是模糊图，$G_\lambda=(V_\lambda,E_\lambda)$，其中 V_λ 是 $\tilde{V}$ 的 λ-截集。

$$V_\lambda:\mu_{V_\lambda}(v_i)=\begin{cases}1, & \mu_{\tilde{V}}\geqslant\lambda\\ 0, & \text{others}\end{cases}$$

$$E_\lambda:\mu_{E_\lambda}(v_i,v_j)=\begin{cases}1, & v_i,v_j\in V_\lambda,\mu_{\tilde{E}}(v_i,v_j)\geqslant\lambda\\ 0, & \text{others}\end{cases}$$

显然在 λ-截图中，截掉了隶属度小于 λ 的结点，留下的结点其隶属度都被改成 1，且截掉了一些特殊边，这些边因其结点的截除而使其一头甚至两头不和其它边相连，也即隶属度小于 λ 的边均被截掉。这样，剩余隶属度为 1 的边。

3.2 模糊图模型的建立

将模糊图论运用到视频语义标注中，以建立模型。根据视频标注的特点，可将论域 V 作为标注模型中的语义概念集合，提取低层特征后，低层特征与语义概念间有模糊关系用模糊集 $\tilde{V}$ 来表示。上文已经提到，语义概念间存在着相互关系，用模糊关系 $\tilde{E}$ 来表示，即在出现语义概念 A 的情况下出现语义概念 B 的概率。为简单表示，当视频的全部或者局部具有某一低层特征时，就将其定义为某一语义，即为 1。例如，在一个小的样本集中，语义概念有：building，car，mountain，road。这样邻域 V 就可以表示为：$V=\{1/building,1/car,1/mountain,1/road\}$。

若不考虑语义概念间的相互关系，也就是说：在出现 building 的情况下出现 car 的概率与在出现 car 的情况下出现 building 的概率是一样的，这样模糊关系 $\tilde{E}$ 表示如下：

$$\tilde{E}=\begin{bmatrix}1 & 0.5 & 0.1 & 0.4\\ 0.5 & 1 & 0.3 & 0.6\\ 0.1 & 0.3 & 1 & 0.2\\ 0.4 & 0.6 & 0.2 & 1\end{bmatrix}$$

矩阵中的行列值 $\mu_{ij}=p(v_i,v_j)/p(v_j)$，其中 $p(v_i,v_j)$ 表示语义概念 v_i,v_j 同时出现的概率，而 $p(v_j)$ 表示语义概念 v_j 出现的概率。如果将语义概念间的关系考虑到模型中，基于方向模糊图的模糊关系 $\tilde{H}$ 就可以表示为：

$$\tilde{H}=\begin{bmatrix}1 & 0.5 & 0.1 & 0.4\\ 0.4 & 1 & 0.2 & 0.6\\ 0.1 & 0.3 & 1 & 0.2\\ 0.3 & 0.5 & 0 & 1\end{bmatrix}$$

如此我们可以给出简单的基于方向模糊图，如图 1 所示。

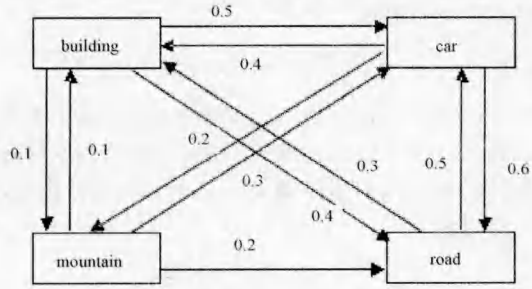

图 1　简单的基于方向模糊图

3.3　基于反馈模糊图的视频多语义标注

在前文中，已经讲述了模糊图的概念以及建立将模糊图运用到视频标注中的模型的方法，以下给出基于反馈模糊图论的视频多语义标注算法的步骤：

(1)按照文中第 2 节所述构造出优化的训练样本集，对样本进行人工标注，选出训练集，统计语义出现的概率以及相互间的关系，建立语义模糊图；

(2)对测试集的样本统计其低层特征，标注出第一语义 f_1 以及第二语义 f_2；

(3)对模糊图进行截图操作，形成语义关系图；

(4)按照第一语义 f_1 统计测试样本的标记信息 U；

(5)判断标记信息集合 U 中是否包括第二语义 f_2，若包含，则将与第二语义 f_2 相关的语义概念添加到标记信息 U 中，此时更新标记信息为 $U=\{f_2,U\}$；若不包含第二语义 f_2，则标记信息仍为 U；

(6)此时测试集中的样本完成标注，选取一部分此时标注完的测试集中的样本添加到训练集中，完成标记信息的反馈。

下面给出算法说明：

算法中所采用的样本的低层特征为 HSV 直方图和边缘直方图。其中语义概念 a 和语义概念 b 之间的边的权值为 $p(a|b)$，其计算方式为：$p(a|b)=p(a,b)/p(b)$，其中 $p(a,b)$ 表示语义概念 a 和语义概念 b 同时出现的概率，$p(b)$ 表示语义概念 b 出现的概率。在确定第一语义和第二语义时，先对待标注的样本统计其低层特征，第一语义与第二语义的计算是先计算出待标注样本与哪些测试集中样本的距离最近，然后统计这些测试集样本的语义，语义概念出现次数最多的语义作为第一语义 f_1，出现次数次多的作为第二语义 f_2，其中距离 sim 的计算公式为：

$$\mathrm{sim}(F_i,F_j)=w_1\sqrt{\sum_{k=0}^{31}(f_{i,k}^c-f_{j,k}^c)^2}+w_2\sqrt{\sum_{k=0}^{79}(f_{i,k}^t-f_{j,k}^t)^2}$$

计算公式中前一部分表示 HSV 距离，HSV 颜色直方图特征根据人类对色调(H)、饱和度(S)、纯度(V)感知能力不同，将图像按 8×3×3 进行非等量量化，即 H 被分成 8 个等级、S 被分成 3 个等级、V 被分成 3 个等级。由此得到一个 72 维的颜色特征向量，表示为 f^c。后一部分表示边缘距离，对关键帧提取边缘直方图，得到一个 80 维的纹理特征向量，表示为 f^t。由于维数不相同，对这两种特征进行高斯归一化，从而使得特征向量的取值在一个相同的尺度内。$f_{i,k}^c$ 表示第 i 帧的第 k 个颜色特征分类，$f_{j,k}^t$ 表示第 j 帧的第 k 个纹理特征分类。w_1 和 w_2 为距离的加权值，本算法中都将其设置为 0.5，表示低层颜色特征和纹理特征对确定语义方面有相同的作用。算法中排序找到 50 个最小的 sim，即离待标注样本最近的 50 个样本，统计样本须带第一语义概念 f_1 和第二语义概念 f_2。

对于图 1 给出的基于方向模糊图，如取 $\lambda=0.5$，则截图之后如图 2 所示。

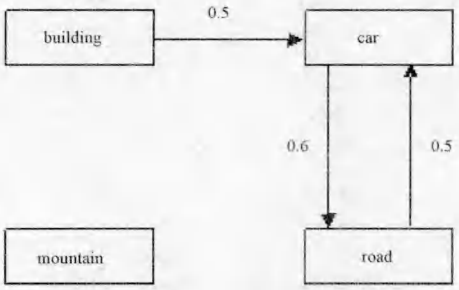

图 2　简单基于方向模糊图的 $\lambda=0.5$ 的截图

若以 building 作为第一语义 f_1，car 作为第二语义 f_2，则首先将与第一语义相关的语义概念加入到标记信息中，此时 $U=\{\text{building,car}\}$；检测 f_2 在 U 中，将与 f_2 相关的语义概念添加到 U 中，更新标记信息，此时的标记信息 $U=\{\text{building,car,road}\}$ 作为最终测试集样本的标记信息。若以 mountain 为 f_1，则没有与其相关的语义概念，此时标记信息 $U=\{\text{mountain}\}$。

4　实验结果与分析

本文设计了实验来验证算法性能，所用样本来自标准视频数据集 TRECVID2007 中的视频片段，为视频片段划分镜头提取关键帧，形成样本集。实验中使用的训练集包含 263 幅关键帧图片、测试集包含 134 幅。基于构造优化训练样本集的原理，首先对这 263 幅关键帧图片进行预聚类处理，然后再采用改进的 k-means 算法，找出聚类中心，从聚类中心中选出一部分样本形成优化的训练样本集，对样本进行人工标注，形成训练集。在人工标注时，根据视频实际内容为每一个关键帧标注 1～4 个语义概念，语义概念有 14 个分别为：sports、weather、outdoor、building、mountain、road、sky、face、person、car、animal、meeting、boat-ship、explosion-fire。从与各语义概念相关的镜头中提取出部分关键帧，如图 3 所示。

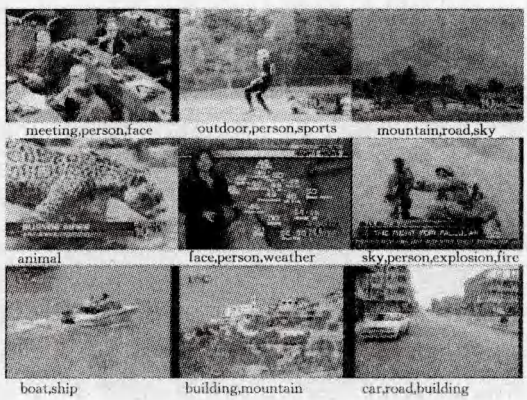

图 3　训练集中手工标注的样本示例

在建立语义关系模糊图时，要对语义概念出现的次数进行统计，在上文进行完优化训练样本集的构造后，从训练样本集中挑选出 145 幅关键帧图片并统计出累计语义出现 359 次，则可以统计出 14 个语义概念出现的概率，如表 1 所列。

表1 14个语义概念出现的概率表(%)

语义概念	出现次数	概率
sports	3	2.07
weather	47	32.4
outdoor	10	6.9
building	22	15.2
mountain	18	12.4
road	14	9.7
sky	25	17.2
face	49	33.8
person	30	41.4
car	25	20.7
animal	15	17.2
meeting	15	10.3
boat-ship	17	11.7
explosion-fire	24	16.6

本实验首先要建立这14个语义的关系,从而构造出语义关系图。对于阈值λ的选择,目前还没有标准。根据实验经验,应使形成的λ-截图尽量建立所有语义的连通。实验中,得知在$\lambda=0.5$时可以对语义概念间的关系作出良好的推理和分析,在$\lambda=0.5$时截图如图4所示。

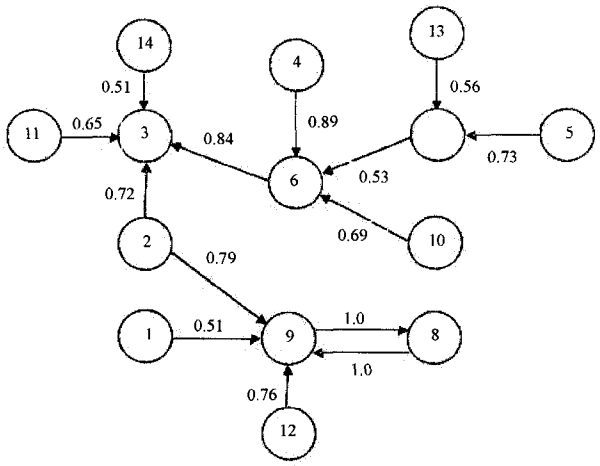

图4 实验的样本集的$\lambda=0.5$的截图

其中结点1到14分别代表语义sports、weather、outdoor、building、mountain、road、sky、face、person、car、animal、meeting、boat-ship、explosion-fire。当确定第一语义为sports时,自动将outdoor和person加入到信息标记集合中;若第二语义是outdoor,则最终的标记信息集合$U=\{$sports, outdoor, person$\}$,若第二语义为person时,此时更新标记信息集合$U=\{$sports, outdoor, person, face$\}$,若第二语义不是outdoor和person,最终的标记信息集合为$U=\{$sports, outdoor, person$\}$。确定好截图后,按照上文所述的基于反馈模糊图的视频多语义标注算法完成标注。

我们进行上述实验是为了对标注性能作出客观评判,是设计实现标注系统过程的一个不可或缺的环节。我们用一种平均查全率(Average Recall, AR)和平均查准率(Average Precision, AP)来评估设计系统的性能。前者表示标注是否全面,反映了系统漏检的能力;后者表示标注的准确程度,反映了系统减少噪声干扰的能力。计算方法如下:

$$AR=\frac{1}{n}\sum_{i=1}^{n} recall[c_i]$$

$$AP=\frac{1}{n}\sum_{i=1}^{n} precision[c_i]$$

其中,n为语义的总数,c_i表示第i个语义。$precision[c_i]$,$recall[c_i]$计算公式为:

$$precision[c_i]=\frac{N_{correct}[c_i]}{N_{plabel}[c_i]},recall[c_i]=\frac{N_{correct}[c_i]}{N_{label}[c_i]}$$

为了评价本文提出的基于反馈模糊图的视频多语义标注算法的性能,用本文方法对实验数据进行语义提取后得到的指标值与经典的SVM算法以及K近邻算法进行比较,结果如表2所列。

表2 算法性能比较表

算法	AP	AR
K近邻算法	0.342	0.469
SVM算法	0.357	0.653
模糊图算法	0.359	0.722

表2给出了基于反馈模糊图论的视频多语义标注算法与K近邻算法和SVM算法的AP和AR值。从表中可以看出经典的SVM算法总是优于K近邻算法,从机器学习的角度分析这是合理的,SVM算法在选取了合适的核函数之后,更能体现出优越性。本文所提出的算法在AP值上略高于SVM算法。基于反馈模糊图论的视频多语义标注算法在确定第一语义时是非常关键的,也就是说第一语义在视频标注中起到了至关重要的作用。从AR值的比较中可以看到,SVM算法的性能仍然要比K近邻算法的好,而基于反馈模糊图论的视频多语义标注算法在AR值上也高于其它两种算法,该算法在提高标注标全率方面有了很大进步,与其它方法相比表现出了良好的性能。因而,本文所提出的基于反馈模糊图论的视频多语义标注算法在以后的研究应用中有很大的发展空间。

结束语 如何利用大量的未标注的样本来改善学习已经成为当前视频检索研究中备受关注的问题。本文的贡献是提出了一种新的基于反馈模糊图论的多语义标注方法来进行自动视频标注。首先根据分类器的泛化性能构造一个样本数目小但是能代表整个样本集分布的优化训练样本集;然后将模糊算子引入图论中,将语义概念之间的关系模糊化以实现模糊推理;最后为了提高视频标注的准确率,选出一部分标注完成的测试集中的样本加入到训练集中,完成视频标注的反馈,从而实现一种基于反馈模糊图论的视频多语义标注算法。实验结果表明:本文提出的基于反馈模糊图论的视频多语义标注方法在视频标注应用中具有良好的性能。

为使得视频标注的结果更为准确,未来可以研究如何将语义概念间的关联性融入到张量学习中[7-15]。

参考文献

[1] Alexander G, Hauptmann. Lessons for the future from a decade of informedia video analysis research[J]. Image and Video Retrieval Lecture Notes in Computer Science, 2005, 3568: 1-10

[2] 黄树成,朱宇光,董逸生. 基于半监督学习的数据流分类方法[J]. 计算机研究与发展,2007,44(z2):225-229

[3] Wang Meng, Hua Xian-sheng, Song Yan, et al. Automatic video annotation by semi-supervised learning with kernel density estimation[C]//MULTIMEDIA'06 Proceedings of the 14th annual ACM international conference on Multimedia. 2006: 967-976

[4] Liu Jing, Li Ming-jing, Ma Wei-ying, et al. An adaptive graph

model for automatic image annotation[C]// MIR'06 Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval. 2006:61-70

[5] Yeung M M, Yeo B L. Time-constrained and Clustering for segmentation of video into story units[C]//Proceedings of the 13th International Conference on Pattern Recognition. Vienna, 1996, 3:375-380

[6] Tang Jin-hui, Hua Xian-sheng, Wang Meng, et al. Correlative Linear Neighborhood Propagation for video annotation[J]. IEEE transactions on systems, man, and cybernetics-part B: cybernetics, 2009, 39(2):409-416

[7] Wang Fei, Zhang Chang-shui. Label propagation through linear neighborhoods[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(1):55-67

[8] Saul L K, Roweis S T. Think globally, fit locally: unsupervised learning of low demnsional manifolds[J]. The Journal of Machine Learning Research, 2003, 4:119-155

[9] Zha Zheng-jun, Mei Tao, Wang Jing-dong, et al. Graph-based semi-supervised learning with multi-label[J]. Journal of Visual Communication and Image Representation, 2009, 20(2):97-103

[10] Jain R, Hong Ri-chang, Yan Shui-cheng, et al. Image Annotation By kNN-Sparse Graph-based Label Propagation Over Noisily-Tagged Web Images[J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(2):111-115

[11] Angelova R, Weikum G, et al. Graph-based Text Classification: Learn from your Neighbors [C]//SIGIR'06 Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Seattle, 2006: 485-492

[12] Liu Qing-shan, Huang Yu-chi, Metaxas D N. Hypergraph with sampling for image retrieval [J]. Pattern Recognition, 2011, 44(10/11):2255-2262

[13] Wang Jing-dong, Zhao Ying-hai, Wu Xiu-qing, et al. A transductive multi-label learning approach for video concept detection [J]. Pattern Recognition, 2011, 44(10/11):2274-2286

[14] Tang Jin-hui, Hua Xian-sheng, Mei Tao, et al. Video annotation based on temporally consistent Gaussian random field [J]. Electronics Letters, 2007, 43(8):448-449

[15] Song Yan, Hua Xian-sheng, Dai Li-rong, et al. Semi-automatic video annotation based on active learning with multiple complementary predictors [C]//Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval. Singapore, 2005:97-104

[16] 袁正午，朱冠宇，丰江帆，等. 基于支持向量机的视频语义场景分割算法研究[J]. 重庆邮电大学学报：自然科学版，2010，22(4)：458-463

(上接第 238 页)

改变并且一旦误报就被清除)，在具体应用中存在检测半径设定困难、检测性能低等问题。为改进 ARTIS 模型存在的上述问题，受生物免疫受体编辑和免疫抑制机制的启发，本文提出了一种新的人工免疫系统模型 REISAIS，模型既包括具有检测功能的检测器，也包括具有免疫抑制功能的抑制器。模型中检测器具有一定的主动学习能力(产生之后可以通过受体编辑扩大对非自体空间的覆盖，误报之后可以通过受体修正进行持续学习)。本文给出了两种受体编辑(RARE 和 DRINNS)和受体修正(RARR 和 DRR)的具体实现，对模型的有效性进行了分析和证明，在两个有代表性的数据集上进行的对比实验结果表明，与 ARTIS 模型相比，所提模型无需设定检测半径并且具有更好的检测性能。

参 考 文 献

[1] Dasgupta D. Advances in artificial immune systems[J]. IEEE Computational Intelligence Magazine, 2006, 1(4):40-49

[2] Forrest S, Beauchemin C. Computer immunology[J]. Immunol Rev, 2007, 216(1):176-197

[3] Timmis J, et al. Theoretical advances in artificial immune systems[J]. Theoretical Computer Science, 2008, 403(1):11-32

[4] Hofmeyr S, Forrest S. Immunity by Design: An Artificial Immune System[C]//Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 1999). 1999:1289-1296

[5] Hofmeyr S, Forrest S. Architecture for an artificial immune system[J]. Evolutionary Computation, 2000, 8(4):443-473

[6] 李涛. 计算机免疫学[M]. 北京：电子工业出版社，2004：147-159

[7] Dasgupta D. Immunity-based intrusion detection system: A general framework[C]//The 22nd National Information Systems Security Conf. 1999:147-160

[8] Harmer P K, et al. An artificial immune system architecture for computer security applications[J]. IEEE Transaction on Evolutionary Computation, 2002, 6(3):252-280

[9] Kim J, Bentley P. Towards an artificial immune system for network intrusion detection: An investigation of dynamic clonal selection[C]// Congress on Evolutionary Computation (CEC 2002). 2002:1015-1020

[10] Kim J, Bentley P. Immune memory and gene library evolution in the dynamic clonal selection algorithm[J]. Genetic Programming and Evolvable Machines, 2004, 5(4):361-391

[11] Kim J, et al. Immune system approaches to intrusion detection-a review[J]. Natural computing, 2007, 6(4):413-466

[12] 李涛. 基于免疫的计算机病毒动态检测模型[J]. 中国科学 F 辑：信息科学，2009，39(4)：422-430

[13] Kim J, Bentley P. An evaluation of negative selection in an artificial immune system for network intrusion detection[C]// Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2001). 2001:1330-1337

[14] Timmis J. Artificial immune systems—today and tomorrow[J]. Natural Computing, 2007, 6(1):1-18

[15] Li Gui-yang, et al. An Outlier Robust Negative Selection Algorithm Inspired by Immune Suppression[J]. Journal of Computers, 2010, 5(9):1348-1355

[16] 李贵洋，郭涛. 一种基于受体编辑的实值阴性选择算法[J]. 计算机科学，2012，39(8)：246-251

[17] 罗微，马骊，王小宁. T 细胞受体编辑与修正[J]. 中华微生物学和免疫学杂志，2008，28(003)：278-281

[18] Stibor T, et al. A comparative study of real-valued negative selection to statistical anomaly detection techniques[C]// Proceedings of Second International Conference on Artificial Immune System (ICARIS 2005). 2005:262-272