

第5章 训练无关的语义概念增强方法

基于语义分析的方法对视觉媒体进行索引已经远远超出了对若干独立概念探测器的简单应用,通过结合多种概念信息及概念探测结果进行后期处理的方法可以进行更加有效的语义检索。由于训练样本集本身的局限性,如人工标注的稀疏性、不准确等缺点,基于特定训练样本集的方法用于对可视媒体进行精度的提高,往往面临着概念相关性(如概念同时出现的规律、本体关联关系等)难以准确获取的问题。

在前面的研究现状中介绍过,可视媒体语义索引的提炼和精度增强是利用概念相关性对索引结果进一步改善的关键技术。这种索引增强技术的作用,是通过给出可视媒体进行概念探测的结果,对该结果应用概念的相关性进行调整和改善,以达到对大量可视媒体有效索引的目的。随着图像、视频等可视媒体在互联网上体量的快速增长,以及在行为自动识别和分析中的多种应用,这种索引的增强方法不可避免地需要满足三点要求:①对语义概念索引的灵活性;②对不同数量和质量标注数据的适应性;③对大规模样本数据的扩展性。在可视媒体语义索引增强的过程中,应该尽量满足上述三点要求,这样才能保证增强方法在视觉媒体大数据上的灵活应用,否则在进行索引增强的过程中,很难有效利用语义概念的相关性,从而影响预期的效果。

为满足实际应用的要求,本章将详细介绍一种灵活有效的可视媒体语义索引的精度增强算法。该算法将全局上下文语义关系、本体相关性、局部关联关系等语义内容结合起来,在大大提高索引精度的同时,减少了对标注数据集和外部知识库的依赖,在很大程度上提高了方法的灵活性。并且本方法的一个突出的优点是,能在不受训练数据集和外部知识限制的情况下充分发掘概念本身存在的出现规律,以实现概念探测结果的自学习和自调整。

5.1 方法出发点

融合概念探测的结果以提供更高质量的语义分析和索引结果是充满挑战的问题。当前的研究更多地集中于从训练数据集中学习得到概念间的显式(Explicit)关系,并将这种量化关系应用于测试数据集。由于原始的语义概念探测的结果通

常是不准确和充满噪声的,在实际研究中很少有学者直接采用原始的探测结果进行相关性分析并用于探测结果的增强。然而,依据 TRECVID 国际评测的结果,概念探测的精度已经逐步达到可以接受的水平,尤其对于一些具有足够多标注样本的概念^[1-2]。尽管进行概念相关性的准确量化仍存在很多困难,但是这些具有更高概念探测精度的结果应该可以用来作为提高整个索引效果的线索和依据,这是因为概念之间不是相互独立而是紧密相关的。

对于日常行为记录,这些媒体内容往往存在很强的时间特性和局部相似特性。视频记录中邻近帧之间存在很强的时间相关性,这种相关性也是很多压缩算法实施的基础。在日常行为记录中,可穿戴式记录设备通过持续捕捉外界环境的变化从而使镜头或关键帧之间在语义内容是紧密相关的。这种相关性往往由可穿戴式传感器所记录的相同场景或者穿戴者所持续从事的行为活动之间的关联特征所决定,例如“做饭”行为中视觉概念如“冰箱”“微波炉”等出现的关联性。除了日常行为记录,在社交媒体事件中持续拍摄或共享的静态图像间由于共同的位置、活动或用户间的社交关系等也会产生很强的语义关联。对于这种语义关联的视觉媒体处理,可以很自然地联想到利用它们内在的时序关联关系进行概念索引的后期处理,即充分利用这些视觉媒体的局部邻近关系对语义识别的结果进行增强。

本章所介绍的训练无关的语义概念增强算法(Training Free Refinement, TFR)受如下4点启发。

(1) 可靠性(Reliability)。对于一些概念的探测结果,其精度已经达到足够高的程度,并可以用来作为进行其他概念结果精度增强的可靠依据。

(2) 相关性(Correlation)。多个概念往往同时在单幅画面中出现,或者存在一些概念相互排斥,即不能同时存在于同一幅画面中。也就是说,概念是相关的,却不是独立的。

(3) 紧凑性(Compactness)。由于概念的出现模式不是相互独立的,因而概念索引的结果可以被映射到一个更加紧凑的语义空间,而不破坏原有的相关性。

(4) 重复性(Re-Occurrence)。同一批概念经常在语义相近的媒体样本中重复出现。因此,在具有时间关系的可视媒体,如邻近的视频关键帧、图像序列等,这种邻近关系可以进一步得到利用。

5.2 方法描述

依据 5.1 节的假设,我们提出了训练无关的索引增强算法,该算法可以整合不同精度概念探测结果的相关性,对整体概念索引的效果进行提高。在原始概念探测过程中,单个概念探测器在单幅图像上的识别结果通常由返回的置信度(C Confidence)值来表示,对初始的概念探测置信度结果进行整合,可以构建结果矩阵 C 。在该结果矩阵中,每行 c_i ($1 \leq i \leq N$) 代表可视媒体的一个样本,如可穿戴式

视觉设备记录的一幅图像或一个视频镜头；每一列 v_j ($1 \leq j \leq M$) 代表用于语义索引的单个概念。其中, N 和 M 分别表示样本和词汇的个数。

如图 5-1 所示,本章提出的可视媒体语义索引的精度增强算法主要包括全局增强(Global Refinement,GR)和近邻传播(Neighborhood Propagation,NP)两部分。这两部分分别通过本体支持的非负矩阵分解方法和基于相似度的图传播方法进行,二者分别利用了概念出现的全局上下文特征和局部相似度特征来实现对原始概念探测精度的改善。

GR 的出发点是,首先选择出具有更高可能性是正确的概念探测结果,这些结果可以构建一个不完全但更可靠的矩阵,该矩阵可以通过分解的办法得到补全。在图 5-1(b)中,GR 通过加权的矩阵分解对原始探测结果矩阵 C 中原本不精确的元素进行估计。如果存在概念本体结构,那么本体关系可以在这个分解过程中用于选择与 C 对应的合适权重矩阵元素。在图 5-1(c)中,以重新构建的矩阵 C' 为基础计算样本间相似度,并根据该相似度为目标样本 c_i 确定一定数量的最邻近样本。然后,通过应用传播算法根据与每个样本互联的邻近样本对概念标签进行迭代推理。

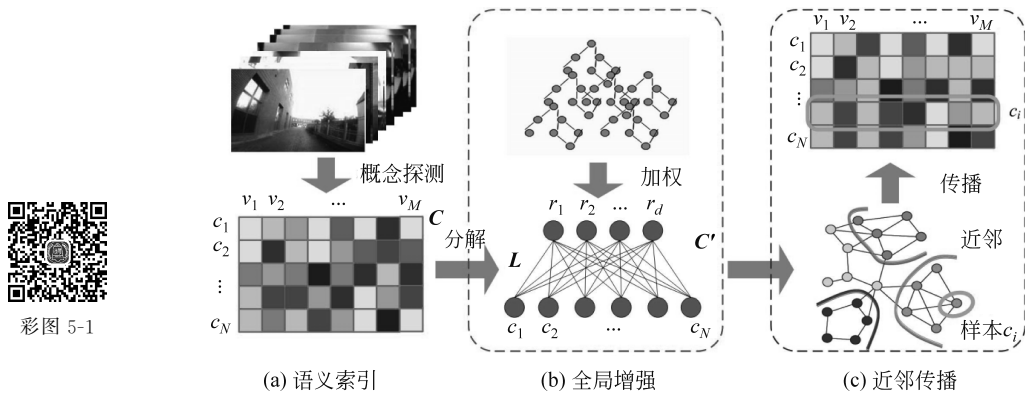


图 5-1 训练无关的索引增强算法框架示意图

图 5-1(a)为语义索引,视觉媒体样本通过概念探测后返回结果矩阵 C 。图 5-1(b)为全局增强,对 C 进行调整并通过全局上下文模式修改为 C' 。图 5-1(c)为近邻传播,进一步通过相似度传播方法对 C' 进行修正。

5.2.1 概念探测结果分解

在图 5-1 中,用不同的灰度值表示原始概念探测结果矩阵 C 中置信度元素的大小。在 GR 过程中,对概念探测结果分解的目的是对矩阵 C 进行调整,以使从调整后结果中反映出的概念上下文关系与实际情况更加一致。非负矩阵分解方法(Non-Negative Matrix Factorization,NMF)在从具有稀疏特性的输入数据中获得

关键特征具有很强的优势,这种方法更加适用于像语义索引增强这种自身的标注数据集具有稀疏性且 \mathbf{C} 中的置信度元素非负的情况。对 \mathbf{C} 应用非负矩阵分解的目的是将其近似为 $\tilde{\mathbf{C}} = \mathbf{L}\mathbf{R}$, 其中, 向量 $\mathbf{L}_{N \times d}$ $\mathbf{R}_{d \times M}$ 及分别代表 d 维样本相关和概念相关的隐含成分。通过应用自定义的优化规则, \mathbf{C} 中的每个置信度值可以被调整为 $\tilde{c}_{ij} = \sum_{k=1}^d l_{ik} \times r_{kj}$ 。在 GR 全局调整过程中, 可以采用加权低秩约束的方法优化上述的分解问题, 以反映不同精度概念探测结果对优化函数的影响。由于 \mathbf{C} 中每个置信度元素 c_{ij} 表示概念 v_j 在样本 c_i 中出现的概率, 当 c_{ij} 的值较高时, 对概念 v_j 存在的估计可能更加正确。这个假设同样在本章文献[3]和[4]中被采用, 即如果概念探测器返回的置信度值高于某个阈值时, 那么最初的探测器相对来说更加可靠。为区分不同概念探测器对优化函数的贡献, 我们在研究中采用一个权重矩阵 $\mathbf{W} = (\omega_{ij})_{N \times M}$ 进行约束, 并对如下加权最小二乘函数进行优化。

$$F = \frac{1}{2} \sum_{ij} \omega_{ij} (c_{ij} - \mathbf{L}_i \mathbf{R}_j)^2 + \frac{\lambda}{2} (\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2) \quad (5-1)$$

其中, $\|\mathbf{L}\| \geq 0$, $\|\mathbf{R}\| \geq 0$, 表示 Frobenius 范数。二次正则项 $\lambda(\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2)$ 用于防止在优化过程中出现过拟合。在实际应用过程中, 可以在权重矩阵 \mathbf{W} 中为更加可靠的探测结果对应的元素赋以较高的权值, 而为不是很可靠的探测结果对应的元素赋以较低的权值。在完成分解后, 对探测结果的调整可以表示为对前后两个置信矩阵的融合, 即

$$\mathbf{C}' = \alpha \mathbf{C} + (1 - \alpha) \tilde{\mathbf{C}} = \alpha \mathbf{C} (1 - \alpha) \mathbf{L} \mathbf{R} \quad (5-2)$$

在求解上述矩阵分解问题的过程中, 采用了乘法方法 (Multiplicative Method)^[5]。不同于采用固定优化步长的方法, 这种方法具有自动缩放学习速率的优势。在不失普遍性的前提下, 在下面的推导中以对矩阵 \mathbf{R} 的更新为例, 对矩阵 \mathbf{L} 的更新可以用相似的方法获得。受本章文献[5]的启发, 本章为函数 $F(r)$ 构建一个辅助函数 (Auxiliary Function) $G(r, r^k)$, 在这个函数中将矩阵 \mathbf{L} 以及矩阵 \mathbf{R} 、 \mathbf{C} 和 \mathbf{W} 中的对应列 r 、 c 、 ω 固定。函数 $G(r, r^k)$ 应该满足条件: $G(r, r^k) \geq F(r)$ 和 $G(r, r) = F(r)$ 。因此, 函数 $F(r)$ 按照如下的更新规则是不增的^[5], 即

$$r^{t+1} = \underset{r}{\operatorname{argmin}} G(r, r^t) \quad (5-3)$$

其中, r^t 和 r^{t+1} 表示在连续的两次迭代中对 r 值的更新。对上述定义的 F , 本章构建 G 函数为

$$G(r, r^t) = F(r^t) + (r - r^t)^T \nabla F(r^t) + \frac{1}{2} (r - r^t)^T \mathbf{K}(r^t) (r - r^t) \quad (5-4)$$

其中, r^t 是当前优化过程中的更新状态。定义 $\mathbf{D}(\cdot)$ 为由向量构建的对角矩阵, 上述式(5-4)中的 $\mathbf{K}(r^t)$ 可以定义为

$$\mathbf{K}(r^t) = \mathbf{D} \left(\frac{(\mathbf{L}^T \mathbf{D}_\omega \mathbf{L} + \lambda I) r^t}{r^t} \right) \quad (5-5)$$

其中, $\mathbf{D}_w = \mathbf{D}(\tau_w)$, 上述的除法执行的是元素除法。因此, r 可以通过优化 $G(r, r')$ 进行更新。通过求解

$$\frac{\partial G(r, r')}{\partial r} = 0$$

可以获得

$$\nabla F(r') + K(r')r - K(r')r' = 0 \quad (5-6)$$

其中, $\nabla F(r') = \mathbf{L}^T \mathbf{D}_w (\mathbf{L}r' - c) + \lambda r'$, 代入式(5-6)可以获得对 \mathbf{R} 的更新规则为

$$\mathbf{R}_{kj}^{t+1} \leftarrow \mathbf{R}_{kj}^t \frac{[\mathbf{L}^T (\mathbf{C} \circ \mathbf{W})]_{kj}}{[\mathbf{L}^T (\mathbf{L}\mathbf{R} \circ \mathbf{W})]_{kj} + \lambda \mathbf{R}_{kj}} \quad (5-7)$$

其中, \circ 表示 Hadamard 即元素乘法。矩阵 \mathbf{L} 中的每个元素可以通过相似的更新规则进行迭代, 即

$$\mathbf{L}_{ik}^{t+1} \leftarrow \mathbf{L}_{ik}^t \frac{[(\mathbf{C} \circ \mathbf{W})\mathbf{R}^T]_{ik}}{[(\mathbf{L}\mathbf{R} \circ \mathbf{W})\mathbf{R}^T]_{ik} + \lambda \mathbf{L}_{ik}} \quad (5-8)$$

需要说明的是, 可以通过证明函数 G 是 F 的辅助函数进而证明上述的更新过程是收敛的, 这个证明将在 5.2.3 节给出。

5.2.2 集成概念本体

在 5.2.1 节, 应用了加权非负矩阵分解的方法进行了低精度概念探测结果的调整, 这种调整基于假设: 如果 \mathbf{C} 中元素高于特定的阈值, 则对应的概念探测将更加可靠。如果为低置信度值都设置相同的权值, 上述的方法将以相等的机会对这些概念探测的结果进行调整。然而, 在实际应用中并非如此, 不同概念往往需要进行不同力度的调整。为了从权值矩阵 \mathbf{W} 中反映概念的相关性, 本章为基于非负矩阵分解的方法引入一种新的本体加权策略 (Ontological Weighting Scheme, OWS)。

为建模概念语义, 首先构建了概念本体结构以方便从该本体结构中为不同的概念推理出适当的权值, 这种思路与本章文献[6]相似。其目的是更准确地构建矩阵 \mathbf{W} 以反映不同概念之间的交互关系和各自不同的探测精度。基于这种启发, 将概念 v 的祖先节点 (Ascendant) 和后代节点 (Descendant) 分别表示为 $\text{ASC}(v)$ 和 $\text{DES}(v)$ 。类似地, 将以显式方式在本体中建模的互斥 (Disjoint) 概念标记为 $\text{DIS}(v)$ 。由概念探测器返回的样本 x 中出现概念 v 的置信度值为 $\text{Conf}(v|x)$ 。引入多类边缘 (Multi-Class Margin)^[7] 为

$$\text{Conf}(v|x) - \max_{v_i \in D} \text{Conf}(v_i|x) \quad (5-9)$$

其中, D 为 v 的互斥概念全集, 即所有与 v 不能同时出现的概念的集合。需要说明的是, 由于在本体中存在隐含的与 v 互斥的概念, 因此有 $D \supseteq \text{DIS}(v)$ 。例如, 在本体结构中只描述了“室内”和“室外”, 二者是互斥的概念, 其中“树”“天空”和“路”是“室外”的后代概念。因此, $\text{DIS}(\text{室外})$ 只包括“室内”这一概念, 而“室内”概念所

有互斥概念的全集 D 包括“室外”和“室外”的所有子孙节点,如“树”“天空”和“路”。事实上,全集 D 包括 $\text{DIS}(v)$ 和 $\text{DES}(\text{DIS}(v))$ 以及 $\text{DIS}(\text{ASC}(v))$ 。其中, $\text{DES}(\text{DIS}(v))$ 表示 v 的互斥概念的全部子孙概念, $\text{DIS}(\text{ASC}(v))$ 表示 v 的祖先概念的全部互斥概念。所有这些对概念关系的陈述 (Statement) 可以在本体中直接声明或进行推理。然而,在推理过程中需要采用推理机 (Reasoner) 对隐含的互斥陈述进行逻辑上的推理。RDFS^[8] 和 OWL^[9] 等不同的推理机可以在本体推理过程中直接采用,并在算法中执行获得隐含概念关系的任务。通过引入本体结构,可以将 \mathbf{W} 中的元素进一步赋值为

$$w_{ij} \propto 1 - (c_{ij} - \max_{v_k \in D} c_{ik}) \quad (5-10)$$

这种加权策略可以解释为: 如果与概念 v_j 互斥的概念具有更高的探测置信度值,则 v_j 在样本 x_i 中出现的可能性将会减小。在这种情况下,概念 v_j 的权值应该适当增大。在相反的情况下,将会应用多类边缘的定义将权值减小。

5.2.3 收敛性证明

前面的推导可以直接得出 $G(r, r^t) = F(r)$ 是满足的,因此对 $G(r, r^t)$ 是函数 $F(r)$ 的辅助函数的证明,事实上仅需要证明 $G(r, r^t) \geq F(r)$ 。为此,将函数 $F(r)$ 扩展为如下形式。

$$\begin{aligned} F(r) &= \frac{1}{2}(c - \mathbf{L}r)^t \mathbf{D}_w (c - \mathbf{L}r) + \frac{\lambda}{2} r^t r + \mathbf{C}(\mathbf{L}) \\ &= F(r^t) + (r - r^t)^t \nabla F(r^t) + \\ &\quad \frac{1}{2}(r - r^t)^t (\mathbf{L}^t \mathbf{D}_w \mathbf{L} + \lambda \mathbf{I})(r - r^t)^t \end{aligned} \quad (5-11)$$

其中, \mathbf{I} 是 $d \times d$ 单位矩阵且 $\mathbf{C}(\mathbf{L})$ 只与 \mathbf{L} 有关。只需证明

$$(r - r^t)^t (F(r^t) - \mathbf{L}^t \mathbf{D}_w \mathbf{L} - \lambda \mathbf{I})(r - r^t) \geq 0 \quad (5-12)$$

这就等同于证明 $\mathbf{D} \left(\frac{\mathbf{L}^t \mathbf{D}_w \mathbf{L} r^t}{r^t} \right) - \mathbf{L}^t \mathbf{D}_w \mathbf{L}$ 是半正定的 (Positive Semi-definite)。为此定义一个缩放矩阵为

$$\begin{aligned} \mathbf{M} &= \mathbf{D}(r^t) \left[\left(\mathbf{D} \frac{\mathbf{L}^t \mathbf{D}_w \mathbf{L} r^t}{r^t} \right) - \mathbf{L}^t \mathbf{D}_w \mathbf{L} \right] \mathbf{D}(r^t) \\ &= \mathbf{D}(\mathbf{L}^t \mathbf{D}_w \mathbf{L} r^t) \mathbf{D}(r^t) - \mathbf{D}(r^t) (\mathbf{L}^t \mathbf{D}_w \mathbf{L}) \mathbf{D}(r^t) \end{aligned} \quad (5-13)$$

对任意向量 \mathbf{v} , 由于 \mathbf{M} 是对称的矩阵,可以得到

$$\begin{aligned} \mathbf{v}^t \mathbf{M} \mathbf{v} &= \sum_{ij} v_i M_{ij} v_j \\ &= \sum_{ij} [r_i^t (\mathbf{L}^t \mathbf{D}_w \mathbf{L})_{ij} r_j^t v_i^2 - v_i r_i^t (\mathbf{L}^t \mathbf{D}_w \mathbf{L})_{ij} r_j^t v_j] \\ &= \sum_{ij} (\mathbf{L}^t \mathbf{D}_w \mathbf{L})_{ij} r_i^t r_j^t \left(\frac{1}{2} v_i^2 + \frac{1}{2} v_j^2 - v_i v_j \right) \end{aligned}$$

$$= \frac{1}{2} \sum_{ij} (\mathbf{L}^T \mathbf{D}_w \mathbf{L})_{ij} r_i^t r_j^t (v_i - v_j)^2 \geq 0 \quad (5-14)$$

至此,可以判断 $\mathbf{D} \left(\frac{\mathbf{L}^T \mathbf{D}_w \mathbf{L} r^t}{r^t} \right) - \mathbf{L}^T \mathbf{D}_w \mathbf{L}$ 是半正定的,因此 $G(r, r^t)$ 是 $F(r)$ 的辅助函数。这就证明了前面推导的式(5-7)和式(5-8)更新迭代规则是收敛的。

5.2.4 近邻相似性传播

如图 5-1(c)所示,基于近邻关系的传播方法可以进一步对中间结果 \mathbf{C}' 进行修正并利用语义上更加相近的邻近样本对索引结果进行提高。这个过程主要包括两部分,即基于相似度的邻近样本定位和图传播方法。

1. 相似度度量

对原始概念探测结果通过全局调整 GR 方法进行修改之后,置信度值已经通过加权非负矩阵分解进行了优化,并使结果反映出来的概念模式与建模后的隐含特征相一致。尽管这个过程在全局对上下文模式进行了建模,相似度传播可以通过利用样本的局部相关性进一步对结果进行修改,如图 5-1(c)所示。在这个过程中,为基于相似度传播提供更加相关的邻近样本显得至关重要,因此可以采用全局优化后的中间结果 \mathbf{C}' 以提供更精确的相似度度量。

为得到样本 c_i 和样本 c_j 的相似度量,在中间结果 \mathbf{C}' 上通过计算 Pearson 相关性进行求解:

$$P_{i,j} = \frac{\sum_{k=1}^M (c'_{ik} - \bar{c}'_i)(c'_{jk} - \bar{c}'_j)}{\sqrt{\sum_{k=1}^M (c'_{ik} - \bar{c}'_i)^2} \sqrt{\sum_{k=1}^M (c'_{jk} - \bar{c}'_j)^2}} \quad (5-15)$$

其中, $c'_i = (c'_{ik})_{1 \leq k \leq M}$ 是矩阵 \mathbf{C}' 的第 i 行, \bar{c}'_i 是 c'_i 的均值。为归一化(Normalize)该相似度,采用高斯形式并重新计算相似度为

$$P'_{i,j} = e^{-\frac{(1-P_{i,j})^2}{2\delta^2}} \quad (5-16)$$

其中, δ 是对样本距离进行伸缩的函数。基于这种相似度度量,可以对任意的目标样本 c_i 的 k 个最近邻进行确定。

2. 图传播方法

为执行图传播方法,基于近邻的传播过程首先依据上面的相似度度量确定 k 个最邻近的样本用于传播,这些样本与目标样本通过无向的边构成一个无向图。在这个无向图中,每个边的权重由上面给出的 Pearson 相似度进行量化。根据构建的全连接的无向图,本章在研究中采用了文献[10]的传播算法,更加精准地对概念探测结果做出预测。在数学表示上,可以将上述构建的无向图表示为目标样本

c_i 与 k 个最近邻样本之间的相似度矩阵:

$$\mathbf{G} = (P'_{i,j})_{(k+1) \times (k+1)} \quad (5-17)$$

其中,矩阵 \mathbf{G} 的前 k 行及前 k 列代表目标样本(即待增强样本 c_i)的 k 个最近邻样本,最后一行和最后一列代表目标样本 c_i 。

通过对矩阵 \mathbf{G} 进行列归一化处理,得到的传播概率矩阵 \mathbf{T} 可以进一步构建为

$$t_{i,j} = \frac{P'_{i,j}}{\sum_{l=1}^{k+1} P'_{i,l}} \quad (5-18)$$

这种列归一化方法保证了矩阵 \mathbf{T} 的每一列都具备概率解释。将目标样本 c_i 的 k 个最近邻样本在矩阵 \mathbf{C}' 中的行索引表示为 $n_i (1 \leq i \leq k)$,并将相应的行向量层叠在一起构成新的矩阵 $\mathbf{C}_n = (\mathbf{c}'_{n_1}, \mathbf{c}'_{n_2}, \dots, \mathbf{c}'_{n_k}, \mathbf{c}'_i)$,则可以通过下面的传播方法对矩阵 \mathbf{C}_n 进行更新:

$$\mathbf{C}_n^t \leftarrow \mathbf{T} \mathbf{C}_n^{t-1} \quad (5-19)$$

其中, \mathbf{C} 的前 k 行表示 \mathbf{C}' 中的 k 个最邻近样本,并由下标 n_i 进行索引,最后一行对应于目标样本 c_i 的置信度向量。由于 \mathbf{C}_n 是 \mathbf{C}' 的一个子集,因此构建在 \mathbf{C}_n 上的图 G 事实上是 \mathbf{C}' 构建的全局图的一个子图,如图 5-1(c)所示。在每一次迭代过程中, \mathbf{C}_n 中的向量 \mathbf{C}'_{n_i} 需要锁定并改写为原值,以避免在迭代过程中出现衰退。进行一定数量的迭代之后,算法将收敛于某个新的结果,该结果中的最后一行为通过相似度传播进行局部增强后的结果 \mathbf{C}_n 。通过这个过程,最近邻样本间的局部关系被用来进行更加全面的索引增强。

5.3 语义平滑的索引增强

5.3.1 算法形式化

按照 5.2.1 节对概念探测结果分解的过程,可以采用加权的 Frobenius 范数^[11-12]区分不同概念探测结果对优化函数的影响。分解得到的隐含因子可以将不同的概念向量映射成语义单元的组合,从而概念间的上下文语义可以通过这种新的编码方式进行评估。如果在这个过程中用外部已知的概念相关性对分解进行约束,分解结果将可以更好地反映上下文信息从而使增强结果得到改善。与 5.2.1 节类似,定义未进行语义平滑的成本函数为

$$G(\mathbf{W}, \mathbf{H}) = \frac{1}{2} \sum_{ij} \omega_{ij} (c_{ij} - \mathbf{W}_i \cdot \mathbf{H}_j)^2 + \frac{\lambda}{2} (\|\mathbf{W}\|_F^2 + \|\mathbf{H}\|_F^2) \quad (5-20)$$

其中,矩阵 \mathbf{W} 和 \mathbf{H} 的维度分别为 $N \times r$ 和 $r \times M$,其他变量的解释与式(5-1)一致。同样,对该问题的最小化求解,可以通过对成本函数的迭代优化,以逐步减小近似的误差。根据前面的解释,矩阵 $\mathbf{H}_{r \times M}$ 代表了 M 个概念对应的低秩隐含特征。因此,概念的相关性可以进一步由这种新的概念表示进行量化。同样,以加权的形式

定义平滑函数 S 以约束这样一个事实,即具有更高相关性的两个概念在空间距离上应该更加接近。在这样的假设下, S 可以由如下方式进行计算。

$$\begin{aligned}
 S(\mathbf{H}) &= \frac{1}{2} \sum_{ij} \|\mathbf{H}_{\cdot i} - \mathbf{H}_{\cdot j}\|_F^2 \mathbf{Correl}(c_i, c_j) \\
 &= \sum_{ij} \mathbf{H}_{\cdot i}^T \mathbf{Correl}(c_i, c_j) \mathbf{H}_{\cdot i} - \sum_{ij} \mathbf{H}_{\cdot i}^T \mathbf{Correl}(c_i, c_j) \mathbf{H}_{\cdot j} \\
 &= \sum_{ij} \mathbf{H}_{\cdot i}^T \mathbf{D}_{ii} \mathbf{H}_{\cdot i} - \sum_{ij} \mathbf{H}_{\cdot i}^T \mathbf{Correl}(c_i, c_j) \mathbf{H}_{\cdot j} \\
 &= \text{tr}(\mathbf{H}(\mathbf{D} - \mathbf{Correl})\mathbf{H}^T) = \text{tr}(\mathbf{H}\mathbf{L}_{\mathbf{Correl}}\mathbf{H}^T) \quad (5-21)
 \end{aligned}$$

其中, \mathbf{Correl} 是概念相关性矩阵,它的每个元素 $\mathbf{Correl}(c_i, c_j)$ 代表两个概念 c_i 和 c_j 的相关性量化。 $\mathbf{D}_{ii} = \sum_j \mathbf{Correl}(c_i, c_j)$ 构成了一个新的对角矩阵,且 $\mathbf{L}_{\mathbf{Correl}} = \mathbf{D} - \mathbf{Correl}$ 表示概念相关性矩阵 \mathbf{Correl} 的拉普拉斯(Laplacian)矩阵。通过对上面定义的函数进行整合,可以得到经过语义平滑后的矩阵分解问题,即

$$\min_{\mathbf{W}, \mathbf{H}} F(\mathbf{W}, \mathbf{H}) = G(\mathbf{W}, \mathbf{H}) + \frac{\beta}{2} S(\mathbf{H}) \quad \text{s.t. } \mathbf{W}, \mathbf{H} \geq 0 \quad (5-22)$$

其中,非负标量 β 用于控制概念相关性在优化 $F(\mathbf{W}, \mathbf{H})$ 过程中的影响力度。在该公式中 $G(\mathbf{W}, \mathbf{H})$ 与 $S(\mathbf{H})$ 之间通过共享概念特征矩阵 \mathbf{H} 进行互相约束,因此,分解后的结果将受到概念相关性矩阵 \mathbf{Correl} 的影响并与其保持一致。在这种情况下, \mathbf{Correl} 通过其拉普拉斯矩阵进一步影响到分解后的特征矩阵 \mathbf{H} 。

由于该优化问题目前没有解析解,因此可以采用经典的梯度下降法,沿着梯度相反方向不断更新矩阵 \mathbf{W} 和 \mathbf{H} ,在迭代过程中逐渐寻找局部的优化结果。 $S(\mathbf{H})$ 对 \mathbf{H} 的梯度是 $\nabla_{\mathbf{H}} S(\mathbf{H}) = 2\mathbf{H}\mathbf{L}_{\mathbf{Correl}}$,因此,函数 F 对应于 \mathbf{W} 和 \mathbf{H} 的梯度可以量化计算为

$$\begin{aligned}
 \frac{\partial F}{\partial \mathbf{W}} &= [(\mathbf{W}\mathbf{H} - \mathbf{C} \circ \mathbf{W})\mathbf{H}^T + \lambda \mathbf{L}] \\
 \frac{\partial F}{\partial \mathbf{H}} &= \mathbf{W}^T [(\mathbf{W}\mathbf{H} - \mathbf{C}) \circ \mathbf{W}] + \beta \mathbf{H}\mathbf{L}_{\mathbf{Correl}} + \lambda \mathbf{H} \quad (5-23)
 \end{aligned}$$

其中, \circ 表示 Hadamard 乘法。通过语义平滑求解得到的矩阵 \mathbf{W} 和 \mathbf{H} 也可以进一步应用到 5.2.4 节的传播方法中,以对求解的中间结果进行进一步的语义增强,如本章文献[11]详细介绍的方法。

5.3.2 概念相关性的外部推理

在一些上下文特征显著的应用领域,如可穿戴式行为感知过程中,概念在单个事件(或行为)主题中具有更强的上下文关联性。概念往往在表示事件语义中扮演不同的角色,并且其中某些概念之间会依据其上下文关系进行交互。这就意味着,如果将概念投射到一个向量空间中,概念向量之间由于本身存在的关联性导致它们之间相互依赖,而这种关联性可以间接地通过事件的主题衔接起来。如图 4-3

所示,事件的语义空间可以定义为由一组概念作为基构成的线性空间。从图 4-3 可以看出,由于一些概念在描述同一个事件过程中存在高度关联性,本章提出通过事件主题关联(Topic-related)方法推理概念之间的相关性。

为合理量化上面描述的概念间相关性,本章采用了第 4 章中事件主题相关的用户实验的结果。在实验中参与者对他们认为与特定事件主题关联的概念进行建议,最终依据所有参与者的输入统计得到概念的相关性。

所提出的计算方法基于这样一种假设,即如果更多的人认为一对概念同时与一个给定的行为类型相关,那么就意味着这对概念之间有很强的相关性。通过遍历所有的目标行为类别,并对这种相关性累计求和,这种综合后的相关性值可以得到量化。为此,一对概念 c_i 和 c_j 的相关性值可以表示为

$$\text{Correl}(c_i, c_j) = \frac{\sum_{\text{act} \in A} \min(v_{\text{act}}(c_i), v_{\text{act}}(c_j))}{\sum_{\text{act} \in A} v_{\text{act}}(c_i) \times \sum_{\text{act} \in A} v_{\text{act}}(c_j)}, \quad i \neq j \quad (5-24)$$

其中, A 表示在用户实验中进行调查的所有行为类型的全集。 $v_{\text{act}}(c_i)$ 是概念 c_i 在特定的行为类型 act 上获得的投票总数, $\min(v_{\text{act}}(c_i), v_{\text{act}}(c_j))$ 反映了一对概念在行为类型 act 同时获得投票的个数,即投票的重叠情况。由于这种重叠的投票结果反映了多个实验参与者对特定行为上下文中两个概念相关性强弱的评估,两个概念的相关性在所有行为事件的重叠投票总数高时应该更强。 $\sum_{\text{act} \in A} v_{\text{act}}(c_i)$ 和 $\sum_{\text{act} \in A} v_{\text{act}}(c_j)$ 表示两个概念 c_i 和 c_j 在所有行为类型中各自获得的投票总数,并用于对相关性的值进行归一化处理。

5.4 实验及结果讨论

在实验中,对训练无关的索引增强算法在两种不同的数据集上进行了验证:数据集一(Dataset1)是由可穿戴式视觉传感器所收集的一系列静态图像;数据集二(Dataset2)是 TRECVID 2006 国际评测中使用的视频数据集。实验中采用了 AP(Average Precision)和 MAP(Mean Average Precision)用于对比手工标注的真实结果进行客观评估。

5.4.1 在数据集一上的评估结果

在这个评估过程中,本章对提出的算法采用与本章文献[4]相同的数据集进行评估。该数据集包含由 4 位真实用户用可穿戴式相机采集的 12248 幅图像,并采用 85 个日常生活中出现的视觉概念对数据集进行标注。为测试在不同概念探测精度下的算法表现,本章采用了本章文献[13]中的蒙特卡洛(Monte Carlo)方法对概念探测器进行仿真。在仿真过程中,概念探测器的结果精度基于手工对概念是