指导教师：杨涛　　提交时间：2015.3.29

The task of Digital Image Processing

数字图像处理

School of Computer Science

No: 01

姓名：吉智宇

学号：2012302390

班号：10011201

为基于网络社区照片的图片推荐应用进行故事情节图重建

Gunhee Kim
Disney Research Pittsburgh
gunhee@cs.cmu.edu

Eric P. Xing
Carnegie Mellon University
epxing@cs.cmu.edu

摘要

在本文中，我们研究了一种从大规模网络图片集合或者其它可选的辅助信息如友谊图中重建故事情节图的方法。故事情节图是对使某一个主题的输入图片集中那些经常发生的事件或活动的发散性叙述结构形象化的一种有效的总结。为了进一步拓展故事情节图的实用性，我们利用它们来执行图片序列预测的任务，从中可使图片推荐应用获得启发。我们把故事情节图的重建问题作为稀疏时变图的推论，并制定优化算法成功地解决了一些关键的有挑战性的网络规模的问题，包括全局最优，线性复杂性，而且易并行。通过在24类超过330万的图片上的实验和亚马逊的Mechanical Turk的用户研究，表明本算法在故事情节图重建和图片预测任务方面优于其他算法。.

1. 引言

拍照设备和高速互联网的广泛传播使之与猖獗的社交网络相结合，在众多的网络平台上产生爆炸性的图片共享。这种大规模的和不断增长的图片数据已导致一个信息超负荷问题：用户往往被洪水般的图片搞得不知所措，努力掌握甚至他们最亲密的朋友的各种活动，事件和故事的照片。因此，以高效但全面的方式自动总结一大组的图片变得越来越艰难但是却有必要。

在本文中，如图1所示，我们研究了一种方法用于从一个主题中多个用户贡献的一大组照片流中推断故事情节图（例如，独立+一天），其中的照片流是一个拍照者在固定时间段（例如一天）内拍摄的一组图像。故事情节图通常是指一系列具有时间顺序或因果关系的事件，这通常由一个有向图[11，17]表示。同样，本文的目标是从一大组照片流中自动推断出这样的有向故事情节图。从概念上讲，该图中的顶点对应于数据集上的主图集群，并且边缘连接在许多照片流顺序复发的顶点。更严格的定义将本文进行阐述。

故事情节图作为图片数据库的结构化总结传达了如下一些独特的优势。首先，在一个图片流中，许多感兴趣的话题通常组成了重复的事件或活动序列。一些典型的例子包括休闲活动，节假日和体育赛事。例如，在独立日，全美国各种事件和活动被数百万人捕获并形成照片流集合，这可能有着共同的故事情节：早上游行，下午烧烤派对，晚上有烟花。这样的故事情节可以通过图片的图结构来更好描述而不是由过渡图像检索方法独立地检索到的一组图像。第二，故事情节图可以表征与主题相关的不同分支化叙事结构。一张照片流由故事的单个线性线索作为时间轴的图象序列。通过由不同的用户聚合这些图片，我们的算法可以揭示故事情节的各种可能的线索，它帮助用户了解底层大局周边的话题。.

我们的目的不同于私人故事情节[15]，这只有一个单一的用户的相册的摘要。在这种情况下，脸部识别是重要的，以便故事情节勾画出自己和她的亲密中心朋友。虽然私人故事情节也是需要的，但我们在这里的目标是通过利用集体建设故事情节图所有可用的照片集。此外，我们还将讨论弱个性化的故事情节图表，我们在其中着重讨论友谊图，使我们对对特定用户的亲密朋友的照片流有更高的权重。

。

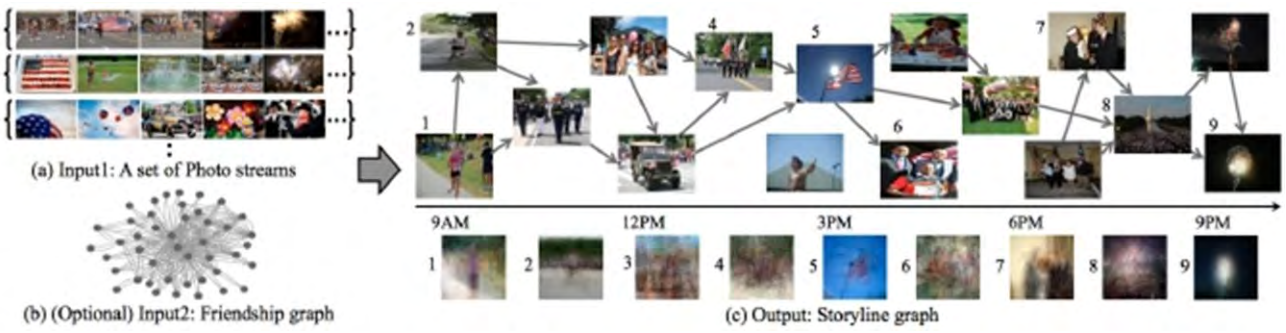


图1。大型成套Web照片故事情节重建流图与（独立+一天）的例子。输入是双重的：（a）组独立地采取多个用户在不同的时间和地点的照片流，（b）是任选的友谊曲线图。（c）的输出是故事情节图形作为结构概要。顶点是图像集群的例子，和边缘连接顺序经常性节点跨越的照片流。我们显示九个所选节点簇在底部的平均图像。

为了进一步显示出的故事情节图的实用性，我们利用它们来进行图像序列预测的任务，这与图片推荐应用直接相关。例如，一旦我们获得人们在滑雪之旅中经常做什么的故事情节的汇总，我们可以给将要开始滑雪旅行的用户推荐一部分经历。这类似于亚马逊的“购买此产品的客户还购买了”功能。

我们把故事情节图重建作为稀疏时变有向图的衍生问题（例如，[22]）。然后，我们提出了一种优化算法，对于大型的问题具有若干有吸引力的特性如最优保障，线性复杂度，易并行，和渐进的一致性。作为评估，我们收集了 24 个主题四万多个图片流超过 330 万的 Flickr 图片。在我们的实验中，我们首先证明，我们的故事情节与其他基线相比是较为成功的结构，同时使用从亚马逊的 Mechanical Turk 获得的注释。我们也定量证明我们的方法在图片序列预测任务方面优于其他候选方法。

与以前的工作的关系。在最近的Web挖掘研究中，很多人已经做了从在线文本语料库提取不同的故事线的工作，如新闻文章和科学论文[1，5，19，20]。虽然从这方面的研究获得一些灵感，我们的工作从根本上不同于他们，我们采用Web图像集而不是文本数据。在[25]，图像会同文本结合起来用于生成故事情节；然而，只有原始图像特性的使用，更重要的是，该算法被用355幅的小数据集图片测试。

在计算机视觉，故事情节挖掘一直在积极研究体育[6]和新闻[13]的视频。然而，视频通常只含有少量的指定演员在固定场景同步的声音和字幕，所有这些都无法在网络社区照片获得。相关研究的另一个线索是探索游客拍的地标性图片集合。在这个方向上，故事情节隐式地以几何的方式实现，如地标三维模型[21]或游客的路径[2,7]。我们的工作不同之处在于，我们的目标是一般主题建筑剧情图，其中没有几何限制（如蝇+钓鱼）。其他值得注意的相关工作总结如下。在[15]，一个基于故事情节的总结讨论了小型私人相册。然而，它仅测试了约200幅图片的数据集，并且它不能正确地处理多个用户的图片。在[9]中，随时间演化的Web图像的副主题在时间线上被可视化。但它的输出是图片之间的相似性图，这样故事的概念没有实现。[8]的工作以户外活动为主题重建照片故事情节。然而，这是一个初步的研究，仅仅着眼于调整和分割照片流;没有探讨故事情节的重建。

贡献。本文的主要贡献可总结如下。

（1）据我们所知，我们的工作是至今为止第一次尝试从大型在线图片解决故事情节的自动重建，尤其是对于休闲活动，节假日，体育主题事件。我们的方法提供了一种新的结构化总结，它不仅能以分支网络的形式使相关联的各种事件或活动联系起来，而且也使应用可实现，如图片推荐应用。

（2）我们开发的优化算法，在辅助信息下从大规模图片流推断稀疏时变有向故事情节图。同时实现几个关键的 Web 规模的挑战，包括全局最优，线性复杂度，而且易并行。随着 24 类超过 330 万图像实验和通过亚马逊的 Mechanical Turk 的用户研究，我们证明，该方法在图像预测任务和重建故事图比其他候选方法成功。

2. 问题描述

我们的算法输入是双重的。第一重输入是一组特定主题的照片流。它由 $P = \{P^1, \cdots, P^L\}$表示，其中 L 是照片流的数量。每个照片流$P^l = \{P_1^l, \cdots, P_{L^l}^l\}$是由单个摄影师在一段时间[0,T]内拍摄的顺序图像集，它被设置为一天。因此，所得的故事情节图表中被定义在[0,T]的范围内。我们假设每个图像p_i^l与用户 ID u^l 和时间戳t_i^l相关，每个照片流中的图像通过时间戳排序。第二个可选的输入是一个友谊图$G_F = (u, \varepsilon_F)$，这是一个加权对称图。该顶点集是一组用户，并且边的权指示友谊图的度。

由于图像集合由大而多图像的高度重叠，对单个图像建设的故事情节图是低效的。故事情节图的顶点优先对应在输入图像集中的复发图像簇。我们通过神经编码的编码和解码想法实现这样的图像群[16]。概念上，该编码通过一个小的码字集合表示每个图像。然后故事情节图在码字集合上定义。解码可以把图从码字集合到图像实例化。

图像编码。为了捕捉图像的各种视觉信息，我们使用四种不同的图像描述符，这是由（SIFT），（HOG2x2），（Tiny）和（Scene）表示。其中（SIFT）和（HOG2x2）分别是三层空间金字塔直方图密集提取 HSV 颜色 SIFT 和导向边缘（HOG）功能直方图。（Tiny）表示微小的图像特征[24]，这是一个 32x32 调整后的图像的 RGB 值。由于这三个特点是高维，我们用软矢量量化的紧凑表示；为每个要素类型构造D_J（= 600）图像集群通过应用 K 均值对随机取样的图像特征，然后将每个图像被分配到 c 最接近图像集群高斯加权。最后，（Scene）表示线性一对多 SVM 分类的得分向量对于 397 的场景类别 SUN 数据集[26]。该（Scene）传达了一个有意义的高层次的描述图像，因为 Web 图像中包含很多场景。同样，我们限制（Scene）向量只保留顶部-C 最高值。我们用 c = {1,3,5} 和τ-1 规范化所有四种描述向量。

因此，每个图像被分配到 J 集描述符的矢量$x_j \in R^{D_j}$与每个非零 C。虽然我们在这里使用 J = 4 $\left[D_j\right]_{j=1}^{4}$= [600,600, 600,397]，但是可以附加任意数量的不同图像描述符。我们可以连接 J 与向量 x，其中|x|=$\sum_{j=1}^{J} D_j$，由于第 3 节讨论的独立性假设，这并不影响我们的图形推理算法。

故事图的定义。故事情节图 G =（O，ε）定义如下。在顶点集合 O 中的每个节点对应于码字（即|O| = D）和所述边缘集 ε∈OxO 包括指示他们之间的有向边缘。我们让故事情节的曲线图是稀疏和随时间变化的[10，22]。稀疏性避免任何不必要的每个节点的复杂的故事分支中的任何图像可以后继任何图像。随时间变化的曲线图意味着我们允许的ε^t随着时间的推移在 T∈[0,T]平滑改变 。这是基于该图像的码字之间的过渡流行可随时间变化;例如，在肺+潜水主题中，水下图像可以在中午跟明亮的天空而在晚上，则是夕阳影像。

因此，我们的算法的输出是一组故事情节的图$\{A^t\}$在 t∈[0,T]，其中在为邻接的ε^t矩阵。虽然我们可以计算A^t在任何点 t，在实践中，我们统一分裂[0,T]分成多个时间点（例如，每 30 分钟），将在估计。稀疏鼓励每个在具有小数目的非零元素，而平滑增强之间的边缘结构连续在和在A^{t+1}的变化平稳。

解码：解码步骤中检索最适合图像由所定义的码字之间的过渡在时间 t。我们采用连续错误纠正的方法输出码（ECOC）[3]，与直方图交集作为解码度量。任何码字或其组合在可以通过h ∈ R^D表示。因此，我们可以附近吨等级的图像通过计算的总和要素明智

最低：PDD = 1 分钟（HD，XD），并取回排名第一形象 h 的代表。

3.估计照片故事情节图

通过下列图表推理的一般方法，我们首先进行结构学习探索拓扑故事情节图，然后参数学习同时固定图形的拓扑。在数学上，前者是识别每个$\{A^t\}$的非零元素，而后者是估算其实际相关联的权重。

对于统计易处理性和可扩展性，我们的算法建立在对照片流四个假设是合理的。在实践中。其中的三个在下面进行介绍，和第四个稍后呈现。（A1）所有的照片流被假定为彼此独立地服用。（A2），我们采用第 k 阶马尔可夫假设在连续的图像中的光流 1 之间。（A3）的图是稀疏，不同顺利跨越时间。

作为图像编码的结果是，每个图像相关联有一个描述符向量x ∈ R^D。因此，我们可以通过$P^l = \{(x_1^l, t_1^l), \ldots (x_{L^l}^l, t_{L^l}^l)\}$表示照片流。

1在这里，我们使用一阶马尔可夫假设为简单我们的讨论。延伸到第k阶马尔可夫模型是直转发，并且将在后面讨论.

我们通过导出一组观察到的照片流 P 的可能性 f（P）开始我们的模型。基于假设（A1）和（A2），可能性 f（P）的定义如下。

$$f(P)=\prod_{l=1}^{L} f(P^l), f(P^l)=f(x_1^l,t_1^l)\prod_{i=2}^{L^l} f(x_i^l,t_i^l|x_{i-1}^l,t_{i-1}^l)$$

其中$f(x_i^l,t_i^l|x_{i-1}^l,t_{i-1}^l)$是连续发生的图像$x_{i-1}^l$在时间$t_{i-1}^l$到$x_i^l$在$t_i^l$照片流 L，其尺寸为$L^l$的条件的可能性。第四假设施加在过渡模式。（A4）$x_i^l$的码字是有条件地相互独立给出$x_{i-1}^l$。也就是说，过渡似然因子超过单独的码字：$f(x_i^l,t_i^l|x_{i-1}^l,t_{i-1}^l)$

$$=\prod_{d=1}^{D} f(x_i^l,t_i^l|x_{i-1}^l,t_{i-1}^l)$$

举一个简单的过渡模式$f(x_i^l,t_i^l|x_{i-1}^l,t_{i-1}^l)$，我们使用线性动力学模型:$x_i^l=A_e x_{i-1}^l+\epsilon$其中$\epsilon$是高斯噪声具有零均值和方差的矢量$\sigma^2$（即$\epsilon$~N(0, $\sigma^2 I$)）。为了t_{i-1}^l和t_i^l之间的时间信息编码成$A_e\in R^{DxD}$，我们使用一个两参数率模型，指数和瑞利模型，它已被广泛地用于表示扩散网络时间动态[18]。其中$\Delta_i=t_i^l-t_{i-1}^l$的（X，Y）被定义为

$$a_{xy}=\begin{cases}\alpha_{xy}\exp(-\alpha_{xy}\Delta_i) & (Exponential)\\ \alpha_{xy}\Delta_i\exp(-\alpha_{xy}(\Delta_i^2/2)) & (Rayleigh)\end{cases}$$

其中a_{xy} ≥0 是从码字的传输速率 x 至 y。因为我们感兴趣的是随时间变化的曲线图中，a_{xy}是时间t_{i-1}^l的函数。但是，为了简单起见，我们在这里让a_{xy}静止，它的动态将在下一节讨论。作为a_{xy} →0，连续发生的码字 X 到 Y 的可能性很小。通过让 A=$\{\alpha_{xy}\exp(-\alpha_{xy}\Delta_i)\}_{DxD}$，我们得到如下的过渡模式：

$$x_i^l=g_i A x_{i-1}^l+\varepsilon, g_i=\begin{cases}\exp(\Delta_i) & (Exponential)\\ \Delta_i\exp(\Delta_i^2/2) & (Rayleigh)\end{cases}$$

从方程（3）中，我们可以表示的过渡的可能性为高斯分布：$f(x_i^l,t_i^l|x_{i-1}^l,t_{i-1}^l)=N(x_{i,d}^l;g_i A_{d*}x_{i-1}^l,\sigma^2)$，其中 表示 A 的第 d 个行。最后，方程（P）的（1）的对数可能性是

$$\log f(P)=-\sum_{l=1}^{L}\sum_{i=2}^{L^l}\sum_{d=1}^{D} f(x_{i,d}^l)$$

$$f(x_{i,d}^l)=\left(\frac{L^l}{2}\log(2\pi\sigma^2)+\frac{1}{2\sigma^2}(x_{i,d}^l-g_i A_{d*}x_{i-1}^l)^2\right)$$

3.1 优化

现在我们讨论优化方法，发现在任何 t∈[0,T]的非零元素，通过最大化等式的对数似然方程（4）。这里的一个困难是对于一个固定 t，学习的数据（即在 t 发生的图像）可能比较稀少，因此估计可能遭受极高方差。为了克服这种困难，我们利用假设（A3），其允许估计在通过重新加权邻近吨观测数据相应。此外，由于假设（A4）我们可以分别为每个码字执行优化（d= 1,….,D）。这种方法在图形推理文献[12]被称为邻里的选择。因此，我们反复地解决每个尺寸 D 倍以下的优化问题

$$\hat{A}_{d*}^t=argmin\sum_{l=1}^{L}\sum_{i=2}^{L^l}\omega^t(i)(x_{i,d}^l-g_i A_{d*}x_{i-1}^l)^2+\mu||A_{d*}^t||$$

其中，权重 w(i)是一个观测图象照片流 l 的在时间 t 的权重。也就是说，如果图像的时间戳 ti 靠近 t，权重 w(i)越大，使得在观察有助于更图形估计在 t 上。当然，我们可以定义为权重

$$\omega^t(i)=\frac{k_h(t-t_i^l)}{\sum_{l=1}^{L}\sum_{i=2}^{L^l}k_h(t-t_i^l)}, k_h(u)=\frac{\exp(-u^2/2h^2)}{\sqrt{2\pi}h}$$

其中 k(u)是高斯对称非负核函数，h 为内核的带宽。

在方程（5），我们有 L-1-正规化的稀疏图结构，其中 λ 是控制 At 的稀疏性的参数。这种方法不仅可避免过度拟合，而且是可行的，因为故事情节在每个节点的分支是足够简单可以很容易地理解。因此，我们的图表推断减少解决一个标准加权 L-1-正规化最小二乘问题，其全局最优解可通过高度可扩展的技术来实现，如坐标下降[4]。因此，整体的图表推断可以以线性的时间内进行相对于所有参数，包括的图像的数量和码字 D。我们 MATLAB 代码尺寸花费不到五分钟，以获得一组 40 FAG 为 245K 图像与 D = 1 的冲浪海滩+话题; 800. 请注意，我们的算法的可扩展性，包括线性复杂度和每码字维度琐碎的并行，是利用数以百万计的图像有可能是许多不同的形象描述我们的问题特别重要。我们在补充呈现算法的更多细节，包括伪码及其渐近统计的一致性，从而保证真正的曲线图可以发现作为数据点的增加无限期的数量[22]。

它是直接与上述优化扩展到第 k -th 为了马尔可夫假设。简单地说，式（3）延伸到与所述第 k 阶自回归模型。和式的平方损失函数（ 5）作相应的改变。

一旦{A}被发现，该参数学习更新每个在非零项

的相关权重，同时不改变零元素。由于每个图的结构是已知的，并观察彼此独立选自（A1）和（A4）中，我们可以很容易地解决这个最大似然估计 At 的，这类似于第 k 马尔可夫链的转移矩阵。例如，At 的 xy 与第一阶马尔可夫假设的最大似然估计是观察到的在时间 t 从 x 到 y 转换的分数.

3.2. 结合元数据作为辅助信息

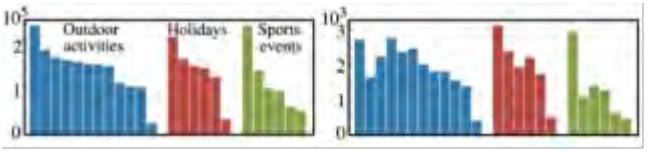
当附加信息是可用的，如一个友谊图，GPS 数据，和其他类型的时间信息，我们就可以定制相应的故事情节的图形。例如，给定一个特定的用户 UQ，故事情节图可以通过在友谊图 GF 加权更多的 UQ 邻域的照片流重铸。另一个例子是一个赛季的具体故事情节图，因为流行的活动或户外活动活动（如蝇+钓鱼）从夏天到冬天将有很大变化。我们利用该产品的内核作为一个统一的框架，把这样的一面信息图形推理。例如，如果特定用户 UQ 和一个月 给出，方程 6 的加权函数被替换为

$$\omega^t(i,u_q,m_q)=\frac{k_h(t-t_i^l)k_s(s_q-s_i^l)k_u\left(\rho(u_q-u^l)\right)}{\sum_{i=1}^{L}\sum_{i=2}^{L^l}k_h(t-t_i^l)k_s(s_q-s_i^l)k_u\left(\rho(u_q-u^l)\right)}$$

其中（UQ;）是友谊图用户UQ和 之间的距离。对于用户的距离，我们使用随机游走重启[23]的得分的倒数。因此，这个内核加权技术是灵活的;我们可以很容易地通过包括其它连续旁侧信息执行该平滑变化效果延长产品的内核。

3.3. 用故事情节进行图像推荐

凭借剧情图，我们执行两个连续图像预测的任务，这是紧密相连的照片推荐应用。（Ⅰ）由于采取了通过用户的图像的短序列，我们预测K下一个可能的图像，这可以帮助用户瞬间预览谁已经有类似的经历的其他用户的图像。（Ⅱ）给定时间上远离的图像的两个部分我们估计在它们之间的最可能的路径。这个函数可用于填补缺失的人的照片流的部分，通过参照的故事情节图的其他用户的照片故事图的推理产生一组{At}的，这可以被视为代码字之间的每一个时刻t的状态转移矩阵。因此，状态空间模型（SSM）是很自然的但强大的框架，实现了连续预测的任务[14]之一。对于任务（I），我们使用的正向算法计算最可能的状态向量的第d个元素表示码字D的图像出现在时间i + K的概率。对于每一个K，我们发现



[从左至右]。户外活动（12）：SB（河畔音响NG+沙滩），HR（马+骑），RA（漂流），SN（滑雪），AB（空气+气球），SD（水肺潜水+），YA（游艇），RO（赛艇），MC（山+露营），RC（石+攀岩），SP（野生动物园+公园），FF（FL Y +科幻盛）。假期（6）：CN（中国+新+一年），IN（就职），ID（独立+天），MD（纪念+天），PD（ST+帕特里克+天），ES（复活节+王孙天）。体育赛事（6）：OL（奥运+伦敦），FO（公式+1），OV（奥运+温哥华），TF（旅游+德+法国），WI（温网），LM（伦敦马拉松+）

图2。Flickr的数据集的三大类24个教学班。该图像和照片数据流的数目示于（a）和（b）中，分别。该数据集大小共（3320080，42744）

从排名分数使用在第2中。关于任务（II）中，我们得到的状态向量通过运行正向-逆向算法的EM讨论的解码方法计算中，由于观察光流中间的最佳对应图像丢失。然后，相同的解码方法用于检索最可能的图像。

4.实验

我们首先评估通过使用亚马逊的Mechanical Turk用户研究重建的故事图。然后，我们定量比较我们的方法与其他的候选方法的两个图像预测任务的性能。在MATLAB代码是在我们的网页为更好地了解我们的算法可供选择。

4.1。评估设置

Flickr的数据集。图2总结了我们的Flickr数据集包括大约42K的照片流为24类，其中分为三类图像3.3M：户外休闲活动，节假日和体育赛事。我们使用的主题名称作为搜索关键字，下载包含超过30个图像正确的时间戳和用户信息。

因为 Flickr的官方不提供用户之间的友好图形的所有质疑照片流，我们从用户间接打造信息。我们抓取列表基团的每一个用户是一个成员，使用Flickr的API。然后，我们将一对用户，如果他们是同一组的成员。友好图形GF =的（U; EF），边缘权重表示两个用户一起加入。

基线。由于故事情节重建是一种新颖的任务，也有少数的现有方法来进行比较。因此，我们选择并适应那些没有最初开发的故事情节重建以下三个基准，但呼吁候选人的方法来可视化图像集的话题演变，并执行循序预测任务。第一基线，记为（页），是一个页秩的图像

检索。它是最成功的方法来检索少量典型图像中的一个，

但无法模拟任何结构信息。我们也实现了使用 HMM，这是最流行的框架进行建模游客连续的照片集之一基线（HMM）[2,7]。所述（Clust）是时间轴[8]在一个聚类的基于聚合，其中，所述时间轴上的图像被分成使用 K 均值在每 30 分钟 10 簇。

4.2。结果在故事情节概述

任务。它本质上是难以定量评价所述重构的故事情节图表由于没有地面的。此外，评价人类受试者的
也无可救药地具有挑战性，因为故事情节图都是大图片收藏有可能上百顶点的总结。为了克服这些困难，我们利用众包为基础的评估通过亚马逊的 Mechanical Turk（AMT）。其基本思想是让每个 Turker 的通过我们的算法和基线，而这些全体人群评估。

我们的评价骨料建造的故事情节非常小的部分之间的比较第一次运行我们的算法和基线来生成数据集的故事情节每个类的。然后，我们品尝 100 最经典的图像从数据集作为 IQ.For 每个测试实例 Iq 的 2 智商测试情况下，我们定位节点 VQ，包括智商每个算法的故事图。然后，我们找到的节点已经是最强烈的连接 VQ，并获得一个中心形象即从节点已经。对于评估，我们将展示智商和一对由我们的算法和基线的一个预测图像，并询问 Turker 的选择其中一个最有可能追随智商。我们设计的 AMT 任务成对偏好的测试，而不是一个选择题测试，因为它可能是更容易，不仅为零工与专业知识水平的所有范围，但也为我们进行统计分析的响应。我们得到的 AMT 的注释的有效性，例如成对比较为每个 IQ 的来自至少三个不同零工。总之，我们评估的想法是招聘，而不是整个故事情节的图形，这实际上是不可能的人群注解来衡量每一个重要的优势的偏好。

定量结果。图 4 示出了我们的方法和三个基线之间的成对的 AMT 偏好测试的结果。的数字表示该选择我们的预测应答的平均百分比为更可能一到每个 IQ 比基线之后来到下一个。也就是说，该数量应该比至少为 50％以上，以验证我们算法的优越性。虽然答案“下一步是什么”是相当主观的，而 AMT 的注释吵闹一定程度是不可避免的，我们的算法显著占主导地位的选票;
例如，我们的算法（我们的）获得选票超过最好的基线（HMM）的平均超过 24 classes 的 66.5％。

4.3 图片序列预测结果

结果在连续图像预测任务。在第二个实验中，我们评估了在相片的建议的范围内，这可以被视为最重要的 1 实际使用故事情节。我们的故事情节的图表进行两图像循序预测任务：（Ⅰ）预测未来可能的图像和（II）的填充在失踪的照片流部分。我们首先随机选择 80％的各类作为训练集的照片流和其它的作为测试集。然后，我们减少每个测试照片流成均匀采样 50 图像，因为连续的图像可以在许多长的照片流非常相似。为任务（I）中，我们随机划分测试照片流成两个不相交的部分。然后，每个算法的目标是，给定的第一部分和接下来的 10 查询时间点 TQ = {tq1,…, tq10}，检索 10 的图像，有可能出现在从训练集中 TQ。测试照片流的 TQ 的实际图像被用作地面。同样地，为任务（II）中，我们随机裁剪出 10 幅图像中的每一个测试照片流的中间。然后，该算法预测为有可能的图像缺少的部分给出的时间点 TQ。我们还进行实验，为弱个性化预测;测试是相同的唯一不同之处在于对用户的查询和月（UQ; MQ）的测试照片流给出。因此，该算法可以利用一个月 MQ 和友谊图找出 UQ 的朋友。总之，我们研究总共多于 10K 试验实例来评估算法的性能。预测质量是用预测和地面图像之间的峰值信噪比（PSNR）进行测定。注意，较高的值表示两个图像更相似。我们描述了如何应用我们的方法和基线的总结的更多细节。

定量结果

图 5 示出具有或不具有弱的个性化我们的方法和三个基线任务（Ⅰ）和（Ⅱ）之间的定量比较。最左边栏设置为 24 个班的平均表现，个体类的 PSNR 值跟随。我们的算法优于所有竞争对手最多的话题类的两项任务。例如，在正常的预测的平均精度，我们的 PSNR 性能增益（以 dB 为单位）比最好基线（HMM）是 0.61 和 0.52（参看准确的数字在图 5 的标题）。有趣的是，弱的个性化预测导致预测精度的仅略有增加。这可能是因为在用户图形邻居之间的拍照行为并不总是相似彼此。友好图形是从用户的 Flickr 的组成员建立，并且因此许多查询用户很可能是因为有自己独特的方式。

图3。 由我们的算法和三个基准预测的图像的例子。在通过AMT每个偏好测试，任务是选择最好的一个，从而可能发生在给定图像之后接下来的一对由我们的算法和基线的一个预测图像中。

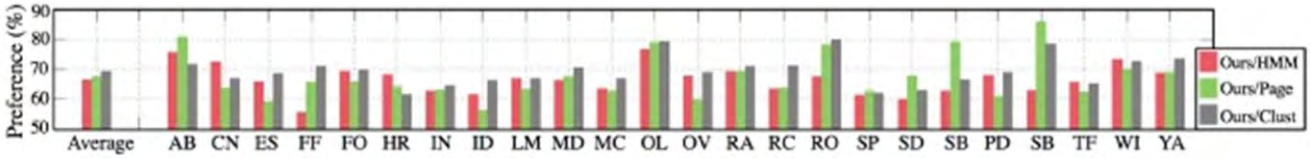


图4。两两偏好测试的结果，通过我们的方法（我们的）和三个基线之间AMT。这些数字表明响应百分比，我们的预测是更可能发生Iq的比基线后下一个。至少数应该高于50%，以验证我们的算法的优越性。最左边一栏集显示了该方法的平均偏好（我们）对所有24类：[66.5，67.5，69.4]在（HMM），（Page），和（Clust）。类的首字母缩写词被称为2所示。

图 6 显示出了通过不同的算法用于预测任务产生了一些选定的实例。在图 6 的第一行(a)-(d)中，我们显示了两个取样给定的图像和图像的位置进行预测的（即{Iq1,…,Iq5}）。然后，我们显示在第二行中的隐藏地面图像，并预测在以下的行由我们的算法的图像和 3 的基准。由于训练和测试集合是不相交的，每一算法只能充其量检索类似（但不相同）的图像从训练数据。所述（HMM）检索相当好的，但高度冗余图像，这是部分由于它不能在表示各种分支结构。在（Clust）基线喜欢时间，连接图像从时间轴上最大的集群，而且性能还不如我们。该（Page）简单地检索排名第一（即代表高品质），在图像的每个查询时间点。由于缺乏使用顺序信息的，存在的预测图象之间没有连接的故事。图 6(e)-(g)显示出小型化的我们的故事情节图表了分别用于图 6 的预测任务的版本（a），（b），（d）。虽然我们在这里显示的简化曲线图只有可能来说明各种方式，其中一些在补充将呈现故事情节的图形。

5. 结论

我们提出了一个从大集合的照片流可在网络上重建故事情节图的方法。随着实验在超过 3 百万 Flickr 的图像通过 AMT24 类和用户研究，我们验证了我们的可扩展算法可以成功地创建故事情节图作为大规模和不断增长的图像集合的有效结构总结。我们还定量显示出我们的故事图比其他候选人的方法更好地完成两个预测的任务

.

Acknowledgement: This work is supported in part by NSF IIS-1115313, AFOSR FA9550010247, Google, and Alfred P. Sloan Foundation.

References

[1] A. Ahmed, Q. Ho, J. Eisenstein, E. P. Xing, A. J. Smola, and C. H. Teo. Unified Analysis of Streaming News. In WWW, 2011. 2

[2] C.-Y. Chen and K. Grauman. Clues from the Beaten Path: Location Estimation with Bursty Sequences of Tourist Photos. In CVPR, 2011. 2, 6

[3] K. Crammer and Y. Singer. On the Learnability and Design of Output Codes for Multiclass Problems. Machine Learning, 47:201–233, 2002. 3

[4] W. J. Fu. Penalized Regressions: The Bridge Versus the Lasso. J. Computational Graphical Statistics, 7:397–416, 1998. 4

[5] J. Gillenwater, A. Kulesza, and B. Taskar. Discovering Diverse and Salient Threads in Document Collections. In EMNLP, 2012. 2

[6] A. Gupta, P. Srinivasan, J. Shi, and L. S. Davis. Understanding Videos, Constructing Plots: Learning a Visually Grounded Storyline Model from Annotated Videos. In ICCV, 2009. 2

[7] E. Kalogerakis, O. Vesselova, J. Hays, A. Efros, and A. Hertzmann. Image Sequence Geolocation with Human Travel Priors. In ICCV, 2009. 2, 6

[8] G. Kim and E. P. Xing. Jointly Aligning and Segmenting Multiple Web Photo Streams for the Inference of Collective Photo Storylines. In CVPR, 2013. 2, 6

[9] G. Kim, E. P. Xing, and A. Torralba. Modeling and Analysis of Dynamic Behaviors of Web Image Collections. In ECCV, 2010. 2

[10] M. Kolar, L. Song, A. Ahmed, and E. P. Xing. Estimating Time-Varying Networks. Ann. Appl. Stat., 4(1):94–123, 2010. 3

[11] J. M. Mandler and N. S. Johnson. Remembrance of Things Parsed: Story Structure and Recall. Cognitive Psychology, 9(1):111–151, 1977. 1

[12] N. Meinshausen and P. B¨uhlmann. High-Dimensional Graphs and Variable Selection with the Lasso. Ann. Statist., 34(3):1436–1462, 2006. 4

[13] H. Misra, F. Hopfgartner, A. Goyal, P. Punitha, and J. M. Jose. TV News Story Segmentation Based on Semantic Coherence and Content Similarity. In MMM, 2010. 2

[14] K. P. Murphy. Dynamic Bayesian Networks: Representation, Inference and Learning. PhD thesis, University of California, Berkeley, 2002. 5

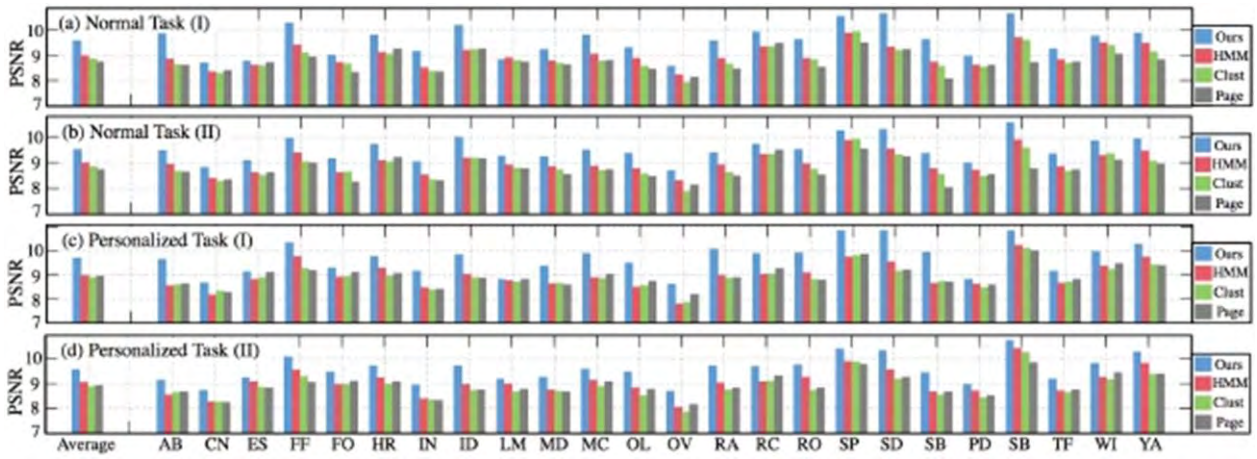


图5。我们的方法的结果，和三个基线为任务（Ⅰ）（即预测可能下一图像），任务（Ⅱ）（即网络连接灌装在缺少份）具有或不具有周个性化。平均PNSR在最左边的栏中设定有 [ours, (HMM), (Clust), (Page)] = [9.60,8.99, 8.86, 8.75] for (a), [9.53, 9.01, 8.85, 8.75] for (b), [9.70, 8.97, 8.89, 8.96] for (c), and [9.57, 9.05, 8.87, 8.93] for (d).

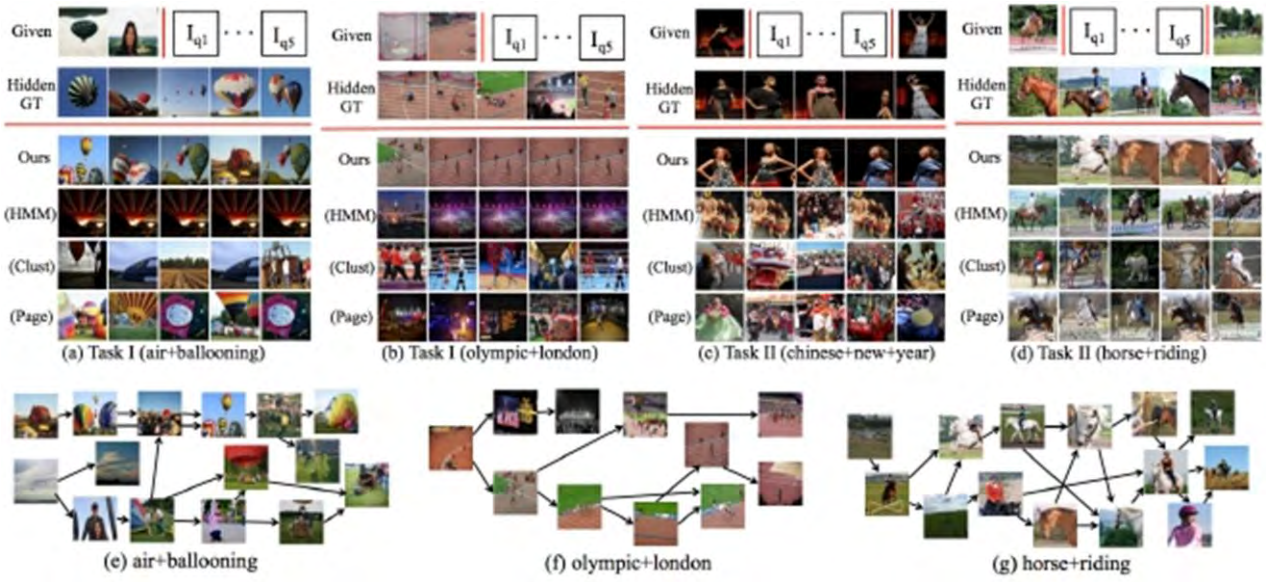


图6。两个预测的任务的示例的结果：(a)-(b)在任务(I)，和(c)-(d)在任务（Ⅱ）。每个算法的目标是预测为使用它的故事情节和给图像连接第一个行。在每一组中，我们显示在第二行中的隐藏地面图像和由其他行不同的算法预测的图像。 我们对我们的故事情节的图表（e）中也存在小型化版本），它被分别用于第（一），（b）和（d）该预测的任务。

[15] P. Obrador, R. de Oliveira, and N. Oliver. Supporting Personal Photo Storytelling for Social Albums. In MM, 2010. 1, 2

[16] B. A. Olshausen and D. J. Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1? Vision Research, 37(23):3311–3325, 1997. 3

[17] M. O. Riedl and R. M. Young. From Linear Story Generation to Branching Story Graphs. IEEE Computer Graphics and Applications, 26(3):23–31, 2006. 1

[18] M. G. Rodriguez, D. Balduzzi, and B. Sch¨olkopf. Uncovering the Temporal Dynamics of Diffusion Networks. In ICML, 2011. 4

[19] D. Shahaf and C. Guestrin. Connecting the Dots Between News Articles. In KDD, 2010. 2

[20] D. Shahaf, C. Guestrin, and E. Horvitz. Trains of Thought: Generating Information Maps. In WWW, 2012. 2

[21] N. Snavely, I. Simon, M. Goesele, R. Szeliski, and S. M. Seitz. Scene Reconstruction and Visualization from Community Photo Collections. Proceedings of the IEEE, 98(8):1370–1390, 2010. 2

[22] L. Song, M. Kolar, and E. Xing. Time-Varying Dynamic Bayesian Networks. In NIPS, 2009. 2, 3, 4

[23] J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos. Neighborhood Formation and Anomaly Detection in Bipartite Graphs. In ICDM, 2005. 5

[24] A. Torralba, R. Fergus, and W. T. Freeman. 80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition. IEEE PAMI, 30:1958–1970, 2008. 3

[25] D. Wang, T. Li, and M. Ogihara. Generating Pictorial Storylines Via Minimum-Weight Connected Dominating Set Approximation in Multi-View Graphs. In AAAI, 2012. 2

[26] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba. SUN Database: Large-scale Scene Recognition from Abbey to Zoo. In CVPR, 2010. 3