

Task 5: AI Books—stage 3

任务描述

- 选择Task5(AIbooks), 第二阶段选择的书为Unsupervised Learning in Space and Time A Modern Approach for Computer Vision using Graph-based Techniques and Deep Neural Networks
- 第三阶段翻译的部分为本书1、2章
 - Unsupervised Visual Learning: From Pixels to Seeing (无监督视觉学习: 从像素到视觉)
 - Unsupervised Learning of Graph and Hypergraph Matching (无监督学习: 图和超图匹配)

二、图和超图匹配的无监督学习 (报告人: 李逸博)

图和超图匹配是计算机视觉中的重要问题, 应用的任务包括: 图像对齐、3D重建、动作识别等。图匹配控制了局部特征之间几何关系和外观的成对约束。同时, 超图匹配包含了对特征集计算的高阶关系, 可以同时获取几何信息和外观信息。时空域中的对象可以用带有局部信息和全局信息的图表示。匹配问题的目的是在局部提取的特征和高阶边和超边计算的信息达成一致。在本章中提出了图和超图匹配、优化、学习的方法。提出了光谱匹配算法和整数投影算法。

1.原理介绍

1.1 与无监督学习的关系

在图匹配问题中, 对齐通常指的是两个图的结点和边之间在集合和外观上的一致性。节点和边的对齐在两个图的节点之间的候选分配的特殊图中建立了强链接。通过这样的强一致性链接, 最有可能是正确的候选任务会形成一个强大的集群。因此, 找到正确的赋值就变成了一个聚类问题。该问题通常需要遵守一定的匹配约束, 这样一个图中的一个节点最多只能匹配另一个图中的一个节点。

1.2 图匹配

带有成对约束的图匹配问题包括求解指示器向量 x^* , 使它最大化具有特定映射约束的二次分数函数:

$$x^* = \operatorname{argmax}(x^T M x) \text{ s.t. } Ax = 1, x \in \{0, 1\}^n \quad (2.1)$$

x 为指示向量, $x_{i,a} = 1$ 如果特征 i 和另一幅图像的特征 a 匹配, 否则为0; $Ax = 1, x \in \{0, 1\}^n$ 对 x 施加一对一的约束, 这样一幅图像中的一个特征最多可以与另一张图像的另一个特征相匹配。 M 为

一个包含正元素的矩阵，其中包含两两计分函数使 $M_{ia;jb}$ 度量一对特征 (i, j) 与另一幅图像的一对候选匹配 (a, b) 在几何和外观方面的一致性。

1.3 超图匹配

超图匹配可以将上面的公式拓展到包含高阶约束的张量公式。给定两组特征，一组来自模型图像 I_m ，另一组来自测试图像 I_t ，三阶（超图）匹配包括寻找这两个集合之间的对应关系，从而使作为候选匹配三元组之和的匹配分数最大化。分数可以考虑潜在赋值的三元组上的几何和外观信息，扩展用于图形匹配的成对分数。

设置一个候选赋值列表 $L_{ia} = (i, a)$ ，每个 (i, a) 对应都有一个唯一的索引 ia ，我们用一个指示符向量 x （长度与 y 相同）来表示一个可能的解，这样如果模型中特征 i 与测试图像中的特征 a 匹配，则标记 $x_{ia} = 1$ ，否则，为0。通常还会施加一对一的映射约束，这样一个图像中的一个特征可以匹配另一个图像中的单个特征。映射约束以矩阵的形式表示

$Ax \leq 1, x \in \{0, 1\}^n$ ， A 为一个二值矩阵。三级匹配分数为：

$$S(\mathbf{x}) = \sum_{ia;jb;kc} H_{ia;jb;kc} x_{ia} x_{jb} x_{kc}. \quad (2.3)$$

利用特征 (a, b, c) 求解超图匹配意味着在一对一映射的约束下，找到最大化 $S(x)$ 的解。

$$\mathbf{x}^* = \arg \max_{\mathbf{x}} \sum_{ia;jb;kc} H_{ia;jb;kc} x_{ia} x_{jb} x_{kc}. \quad (2.4)$$

2 求解图匹配

2.1 光谱匹配

贪婪离散化算法：我们解释主元的每个元素 M 的特征向量 v 作为相应赋值正确的置信度。我们选择最大置信度的元素作为开始，然后我们删除所有冲突的赋值，并选择正确的赋值，然后重复这个过程，直到所有赋值都被标记为正确或不正确。

2.2 整数投影不动点法 (IPEP)

Algorithm 2.1 Sequential second-order expansion for higher order matching

Initialize $\mathbf{x}^* = \mathbf{x}_0$, $S^* = \mathbf{x}_0^T \mathbf{M} \mathbf{x}_0$, $k = 0$, where $x_i \geq 0$ and $\mathbf{x} \neq \mathbf{0}$;

repeat

- 1) Let $\mathbf{b}_{k+1} = P_d(\mathbf{M} \mathbf{x}_k)$, $C = \mathbf{x}_k^T \mathbf{M}(\mathbf{b}_{k+1} - \mathbf{x}_k)$, $D = (\mathbf{b}_{k+1} - \mathbf{x}_k)^T \mathbf{M}(\mathbf{b}_{k+1} - \mathbf{x}_k)$;
- 2) If $D \geq 0$, set $\mathbf{x}_{k+1} = \mathbf{b}_{k+1}$. Else let $r = \min(-C/D, 1)$ and set $\mathbf{x}_{k+1} = \mathbf{x}_k + r(\mathbf{b}_{k+1} - \mathbf{x}_k)$;
- 3) If $\mathbf{b}_{k+1}^T \mathbf{M} \mathbf{b}_{k+1} \geq S^*$ then set $S^* = \mathbf{b}_{k+1}^T \mathbf{M} \mathbf{b}_{k+1}$ and $\mathbf{x}^* = \mathbf{b}_{k+1}$;
- 4) Set $k = k + 1$

until $\mathbf{x}_{k-1} = \mathbf{x}_k$

return \mathbf{x}^*

整数投影不动点法 (IPEP) 将任何初始的解 (连续或者离散) 作为输入, 可以快速找到问题2.1的初始离散约束的解, 该解比初始解的得分通常显著更高。

3 求解超图匹配

超图匹配和图匹配一样, 都是 $NP - hard$ 的问题, 因此人们通常需要高效的算法来寻找好的近似解。目标是最大化原始目标得分, 同时尽可能保留原始离散一对一匹配约束。我们通过实验证明, 原始区域的局部优化比松弛区域的全局优化更有效。

提出的方法是一个迭代的过程, 在每个步骤 k , 通过其围绕当前解 \mathbf{x}_k 的二阶泰勒展开来近似高阶分数 $S(\mathbf{X})$ 。这将超图匹配问题转化为一个整数二次规划 (IQP), 它是围绕当前解决方案 \mathbf{x}_k 定义的。二阶近似也可以局部优化, 在连续域 $A\mathbf{x} = 1$, $\mathbf{x} \in [0, 1]^n$ 中是有效的。

定义匹配分数如下:

$$S(\mathbf{x}) \approx S(\mathbf{x}_0) + 3(\mathbf{x}^T \mathbf{M}(\mathbf{x}_0) \mathbf{x} - \mathbf{d}(\mathbf{x}_0)^T \mathbf{x}). \quad (2.10)$$

要局部最大化该分数, 利用高阶匹配的顺序二阶展开算法处理上面的匹配分数, 该算法在有限次迭代后停止, 可以得到至少比初始值好的解。

4.用于图匹配的无监督和半监督学习

理论分析:

提出的算法是基于矩阵 M 和主特征向量 v 的统计特性实现的, 主特征向量 v 是由谱匹配算法给出的连续解。

$$v_r \approx \frac{(k-1)p_r - 1 + \sqrt{(1 - (k-1)p_r)^2 + 4(k-1)p_r^2}}{2(k-1)p_r}. \quad (2.12)$$

$$v_r^* = \frac{m/(k-1) + (1 - m/(k-1))v_r}{1 - m + mv_r}, \quad (2.13)$$

2.12建立了 p_r 和 v_r 之间的关系，表明，与不正确分配之间的成对协议相比，正确分配之间的成对协议越有可能(由 p_r 选择)，匹配性能越好。

2.13将 v_r^* 与 v_r 和误差率 m 联系起来。每个 p_r 是从训练序列上的每个单独的矩阵 M 中凭经验估计的，类似地，每个 v_r^* 和 v_r 来自每个单独的特征向量。基本上显示了为什么无监督算法相对于基本事实确实将最大化性能。

在无监督学习中，要最大化如下式子：

$$J(\mathbf{w}) = \sum_{i=1}^N \mathbf{v}^{(i)}(\mathbf{w})^T \mathbf{b}(\mathbf{v}^{(i)}(\mathbf{w})). \quad (2.20)$$

这里的困难是 $b(v(i)(w))$ 不是一个连续的函数，也不可能用 w 来表示，因为 $b(v(i)(w))$ 是一个迭代离散化过程的结果。然而，重要的是 $b(v(i)(w))$ 是分段常数，并且除了有限组不连续点之外，在任何地方都为零导数。因此，我们可以预期，我们将只在 b 为常数且导数为零的点上计算梯度。

但是在实际应用中，用户已经知道了一些正确的匹配，所以半监督学习更加适合。结合监督学习和无监督学习的步骤是：

来自每个步骤的离散解 b 具有固定的分配值，对于这些分配，地面真实信息是可用的，而对于其余的未标记的分配，使用在无监督的情况下由图匹配算法返回的解。

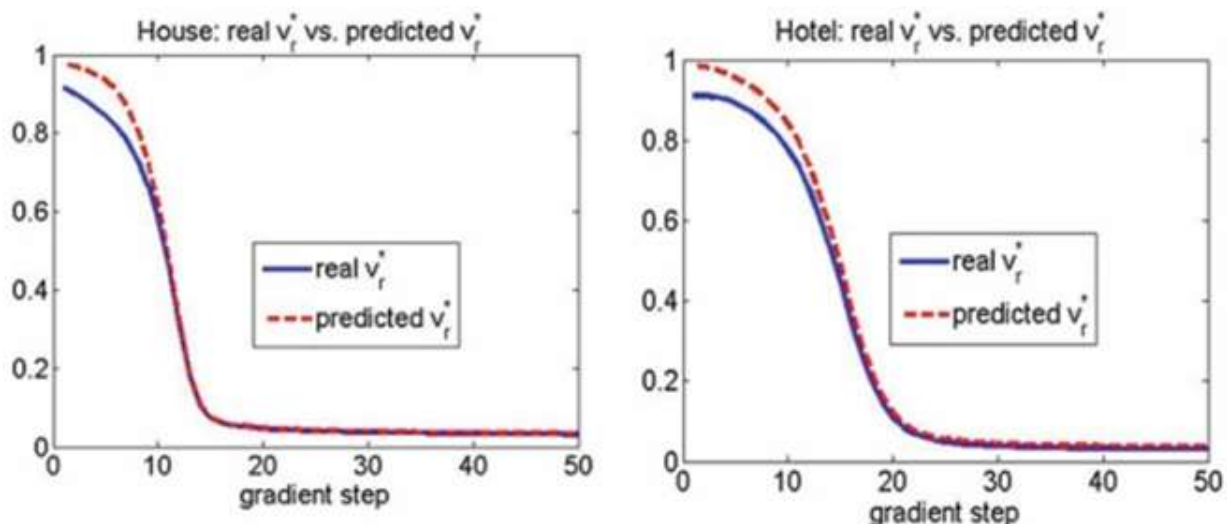
5 图匹配实验(监督学习)

对于图形匹配，关注的两个目标为：

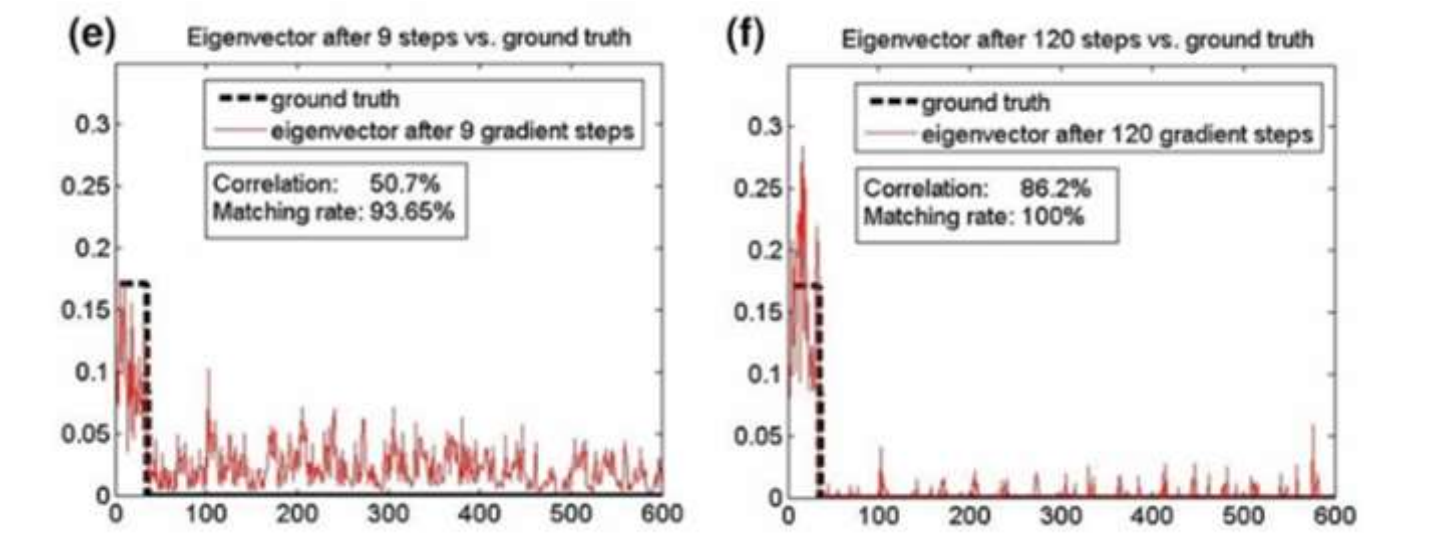
- 上面第四部分理论分析中的2.12式子和2.13式子。在这两个式子的基础上提出匹配率/性能的指标度量正确匹配的特征的比率，定义错误率为1-匹配率。
- 评估学习前后测试图像的匹配性能。目标表明，在测试时，学习后的匹配性能明显优于没有学习的情况。

下面的图分别显示的是房屋、九点、人脸、汽车和摩托车实验的结果。

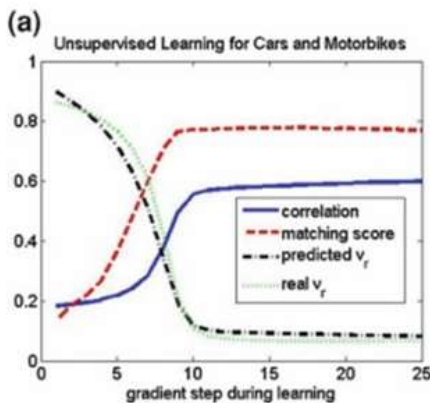
无监督学习阶段：每个图像获得的特征向量 v^* 和 v_r ，判断是否一致。



在人脸上的结果：左图显示的是特征向量和真值之间的相关性，右图显示的是训练期间的匹配率。（为30个不同实验的平均值）



相关性和匹配率：在汽车和摩托车上进行无监督学习的结果，将真实值和预测值进行比较。



- 实验的细节如下：
- v 和真实值 t 之间的相关性在每个步长都增加
 - 匹配率和相关性同时以相似的速率提高，表明最大化这种相关性也会最大化最终性能。
 - 在所有单个实验中，方程2.12和2.13的两个近似值随着每个梯度步长而变得越来越精确，从第一次迭代的不到10%的精度到最后的不到0.5%的误差。

6 图匹配实验（用未标记的对应关系学习）

使用完全相同的图像序列(房屋:110幅图像，酒店:100幅图像)进行训练和测试，并使用由卡埃塔诺等人手动选择的相同特征。

下图为在“房屋”训练集上的结果，显示的是匹配的运行时间和暂停时间。符号：“S”和“U”分别表示“有监督”和“无监督”。

Dataset	Ours: S(5) (%)	Ours: U(5) (%)	Caetano et al. [17]: S(5) (%)	Caetano et al. [17]: S(106) (%)
House	99.8	99.8	<84	<96
Hotel	94.8	94.8	<87	<90

将监督学习和无监督学习进行比较，进行了70个不同的实验，随机选择10个训练图像，得到的结果是：无监督学习更慢，但是参数学习是基本相同的。虽然无监督学习较慢，但是在经过几次迭代之后，它们的性能会收敛到相同的值。在测试期间，这两种方法在匹配性能方面表现相同(在所有70个实验中正确匹配的特征的平均百分比)。与没有学习参数的相同匹配算法相比，这两种算法的性能明显更好。

下图比较的是学习参数的情况下进行无监督学习的结果和无参数学习的结果。

Datasets	Unsupervised learning	No learning
House + Hotel (%)	99.14	93.24

下图显示的是，摩托车图像对的匹配结果。



7 结论

在这一章中，提出了基于二次整数规划公式和图/超图匹配和推理映射问题的高效学习和优化方法。通过大量的实验证明，这些算法可以显著提高匹配的准确性。匹配的两个重点是：

- 在连续域中找到松弛问题的最优解
- 获得具有重要理论意义的高质量离散解

8 总结

通过第三阶段这本书的阅读和翻译，我对于无监督学习在图和超图匹配方面有了一定的了解，这本书难度比较大，但是我从中也学到了如何组织自己的实验验证理论和猜想并提出改进的完整过程。

在本学期的task5项目中，我完成的部分是：

- 《Artificial Intelligence Machine Learning and Deep Learning》的第二章
- 《Artificial Intelligence Machine Learning and Deep Learning》的第四章
- 《Unsupervised Learning in Space and Time A Modern Approach for Computer Vision using Graph -based Techniques and Deep Neural Networks》的第二章

在翻译的过程中，我对于机器学习有了更进一步的认识，同时还对机器学习领域的具体应用和具体实验有了一定了解，收获很大。