

关于论文《Exclusivity-Consistency Regularized Multi-view Subspace Clustering》的学习报告

这是一篇发表在 2018 CVPR 上的论文，主要解决的问题是多视图聚类，采用的方法是将单视图子空间聚类拓展到了多视图子空间聚类。在单视图到多视图的拓展中，为了利用多视图数据的互补性与一致性性质，作者提出了两个惩罚函数项，分别是排他项和一致项。

在当前存在的多视图聚类方法中，普遍存在两大缺点：首先是大多数的工作都是倾向于从多源的数据中学习一个共有的数据表示，而忽略了不同视图之间的互补性信息。其次是当前存在的应用了子空间学习和谱聚类的方法都是把这两个步骤分开执行，当成两个串行的步骤；但没有考虑到这两个步骤更应该是并行的，每个步骤高度依赖于另一个步骤的影响。

因此，在论文中作者提出了 representation exclusivity 和 indicator consistency 这两项，分别对应多视图数据的互补性和一致性性质，并且把这两项共同加入到一个目标函数中，对目标函数的优化就相当于把子空间学习和谱聚类当成并行的步骤执行。

接下来详细介绍算法 ECRMS 的框架：

1. 子空间聚类：子空间聚类分为子空间学习和谱聚类两个阶段。

- a) 首先子空间学习基于的是数据的自表示过程，基本思想是假设高位空间中的数据本质上属于低维子空间，能够在低维子空间中进行线性表示，能够揭示数据所在的本质子空间，有利于数据聚类。即数据 x_i 可以表示为同一个子空间的其他数据的线性组合的形式。因此得到以下的目标函数：

$$\min_{Z, E} \|E\|_k + \lambda \|Z\|_l \text{ s.t. } X = XZ + E, \text{diag}(Z) = 0, \quad (1)$$

其中 $Z \in R^{n \times n}$ 为子空间表示矩阵，假如数据 i 和数据 j 不在同一个子空间中，则 z_{ij} 为 0。因此 XZ 的意义就是使数据能表示为同一子空间下其他数据的线性组合的形式。目标函数中的 $\|Z\|_l$ 是为了避免平凡解，即避免数据样本只用自身来进行表示。

这里的子空间学习思想与 LLE (locally linear embedding) 的思想很像，LLE 的基本思想是假设某个数据样本可以通过其邻近的数据样本的线性组合表示，而这里的邻近数据样本在子空间学习中就是相同子空间下的数据样本。

- b) 那么子空间表示矩阵 Z 的真实意义是什么呢？我个人理解为 z_{ij} 就表示数据样本 i 和数据样本 j 属于同一个子空间的可能性，因为 i 和 j 属于同一个数据子空间的可能性越大，则对 i 的表示中 j 占的权重就越大，这样也可以理解为样本 i 和样本 j 的相似性就越大。因此，可以将子空间表示矩阵 Z 理解成样本之间的相似度矩阵。

- c) 得到样本相似度矩阵, 我们就能运用一般的谱聚类算法, 因此以上所有步骤统称起来就是子空间聚类, 得到的目标函数如下:

$$\min_{\mathbf{F}} \|\mathbf{Z} \odot \Theta\|_1 \text{ s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}.$$

$$\text{其中 } \Theta_{ij} = \frac{1}{2} \|f^i - f^j\|_2^2$$

2. ECRMS 框架

- a) Representation Exclusivity

矩阵 \mathbf{U} 和矩阵 \mathbf{V} 的排他性定义为

$$H(\mathbf{U}, \mathbf{V}) = \|\mathbf{U} \odot \mathbf{V}\| = \sum_{i,j} (u_{ij} * v_{ij} \neq 0)$$

\mathbf{U} 和 \mathbf{V} 的排他性可以这样理解, 要是 $H(\mathbf{U}, \mathbf{V})$ 最小, 最理想就是每个矩阵中的所有元素为 0, 即 i 和 j 属于不同的子空间。那么假如在矩阵 \mathbf{U} 中 i 和 j 属于同一个子空间, 那么这个惩罚函数就会使矩阵 \mathbf{V} 中 i 和 j 尽量不属于同一个子空间。从而使得矩阵 \mathbf{U} 和矩阵 \mathbf{V} 就可能地不像。多视图中的互补性说白了就是不像性, 即多个视图中的表示应就可能地不像, 这样才能形成互补的效果。因此将排他性这一定义引入到多视图子空间聚类, 得到以下的目标函数:

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{Z}_v} \sum_{v=1}^V \left(\|\mathbf{E}_v\|_1 + \lambda_1 \|\mathbf{Z}_v\|_1 + \lambda_2 \sum_{w \neq v} \|\mathbf{Z}_v \odot \mathbf{Z}_w\|_1 \right) \\ \text{s.t. } \forall v, \mathbf{X}_v = \mathbf{X}_v \mathbf{Z}_v + \mathbf{E}_v, \text{diag}(\mathbf{Z}_v) = \mathbf{0}. \end{aligned} \quad (7)$$

- b) Indicator Consistency

在多视图聚类的情況下, 同一个数据样本应该聚类到同一个簇中, 因此这篇论文中引入了标签一致性这个概念, 即无论是哪个视图下, 同一个数据样本的标签 (即聚类分配的结果) 都是相同的。而这一点可以结合子空间聚类的步骤得到, 因此得到以下的目标函数:

$$\min \sum_{v=1}^V \|\mathbf{Z}_v \odot \Theta\|_1 \text{ s.t. } \mathbf{F}^T \mathbf{F} = \mathbf{I}, \quad (8)$$

其中 \mathbf{F} 是所有视图共用的聚类指示矩阵。这个共用指示矩阵要和每个视图下学习到的子空间表示结合起来进行聚类分析。

最终整个框架的目标函数如下:

$$\begin{aligned} \min_{\mathbf{F}, \mathbf{Z}_1, \dots, \mathbf{Z}_V} \sum_{v=1}^V \|\mathbf{E}_v\|_1 + \lambda_1 \|\mathbf{Z}_v\|_1 + \lambda_2 \underbrace{\sum_{w \neq v} \|\mathbf{Z}_v \odot \mathbf{Z}_w\|_1}_{\text{Exclusivity}} \\ + \lambda_3 \underbrace{\|\mathbf{Z}_v \odot \Theta\|_1}_{\text{Consistency}} \\ \text{s.t. } \mathbf{X}_v = \mathbf{X}_v \mathbf{Z}_v + \mathbf{E}_v, \text{diag}(\mathbf{Z}_v) = \mathbf{0}, \mathbf{F}^T \mathbf{F} = \mathbf{I}, \end{aligned} \quad (9)$$

接下来就是我的想法：以前我们遇到的多视图聚类方法都只是从多视图的数据中学习到一个公共的表示，这篇论文是我第一次见到还用到了多视图的互补性的性质。并且这篇论文的互补性和一致性是相互影响的，首先是得到所有视图下数据的子空间学习表示，使不同视图的表示就可能地多样化，达到多视图互补的目的，然后在互补的情况下学习到一个公共的指示矩阵 F 。这里我们发现整个过程涉及两个互为对立的群体，一个是使得各个视图的表示尽可能不相同的互补性；另一个是得到多个视图共同的表示的一致性。

老师上次说的在以图为例的多视图聚类中我们要解决两个问题，一个是如何度量两个图的相似性；一个是如何利用多视图下的互补性与一致性。对于第二个问题，现在我的想法是：我们可以把多视图的互补性和一致性理解为两股互相博弈的敌对势力，一个是为了相同；一个是为了不同。然后我就想到了对抗性神经网络 (GAN)，初步了解了一下 GAN，发现 GAN 就是通过两股势力的互相博弈，然后各自进步，企图战胜对方的过程。因此能不能通过对抗性神经网络对多视图数据的一致性与互补性进行训练，通过互补性这一股势力逼迫一致性这一股势力学习得越来越好，也就得到了更充分和更全面的数据一致性，这样就能够融合多个视图更充分的信息。

昨天在 2017 的 CIKM, AAAI 和 CVPR 上找了一下，发现还没有人做过利用 GAN 提取多视图数据的一致性与互补性的工作。