

1 论文：基于对比预测的缺失视图聚类方法

1.1 原文资料

英文名称: Incomplete Multi-view Clustering via Contrastive Prediction

中文名称: 基于对比预测的缺失视图聚类方法

第一作者: 林义杰 (2020级直博研究生)

第一作者单位: 四川大学计算机学院

通讯作者: 彭玺

出处: 国际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021)

时间: 2021年3月

相关代码<https://github.com/XLearning-SCU/2021-CVPR-Completer>
(<https://github.com/XLearning-SCU/2021-CVPR-Completer>)

论文原文<http://pengxi.me/wp-content/uploads/2021/03/2021CVPR-completer.pdf>
(<http://pengxi.me/wp-content/uploads/2021/03/2021CVPR-completer.pdf>)

VALSE 论文速览<https://www.bilibili.com/video/BV1Ub4y1a7Zy/>
(<https://www.bilibili.com/video/BV1Ub4y1a7Zy/>)

1.2 相关知识:

常见的多视图学习:

1. Joint Representation:

各个视图分别投影到联合子空间中, 去学习多视图联合表示

2. Coordinate Representation:

分别映射到各自子空间, 但这个各自映射的表示之间要满足相关的约束条件

跨视图的一致性假设: 数据的完备性: Completeness of data 即: 要求所有样本点在所有视图都存在: It requires that all examples appear in all views

解决一个机器学习问题主要有两部分: 数据和算法。而算法又有三个部分组成: 假设函数、损失函数、算法优化。机器通过损失函数进行学习。这是一种评估特定算法对给定数据建模程度的方法。如果预测值与实际结果偏离较远, 损失函数会得到一个非常大的值。在一些优化函数的辅助下, 损失函数逐渐学会减少预测值的误差

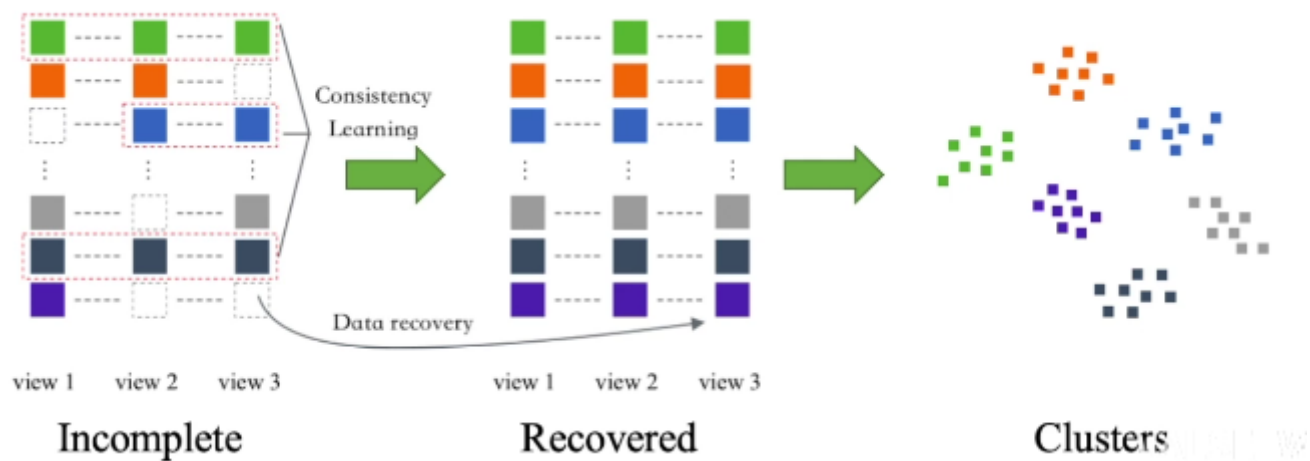
1.3 背景:

然而现实中，监控的部分摄像头会出现损坏或者噪声等问题，或者自动驾驶的部分传感器失灵，需要我们推测反演出没有照射到的部分

不完全多视图的聚类存在两个挑战问题：

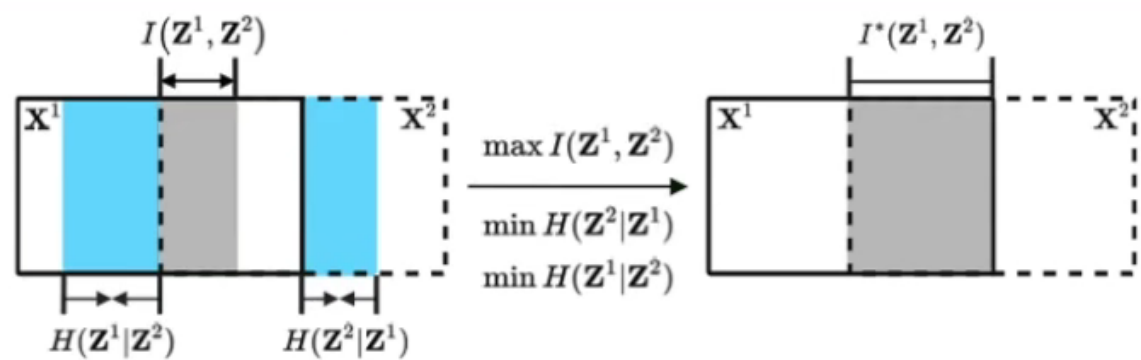
- 1.不利用标签信息来完成学习的一致性 即：无监督一致性学习， 从不完全的多视图学习中，对未缺失数据间的进行一致性学习，即充分利用部分完整的多模态数据进行一致性学习（不考虑缺失数据的一致性）
- 2.消除缺失视图的影响，从而恢复数据

如下图所示：



1.4 方法：

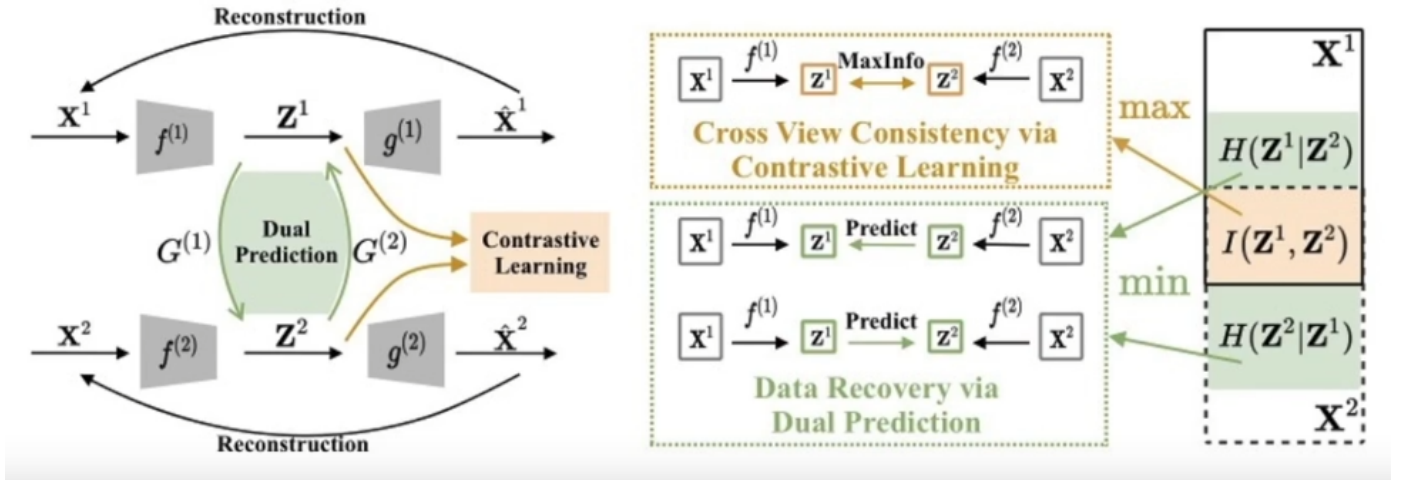
对于目前科研研究现状：跨视图一致性学习和数据恢复都是**分开处理**的，没有考虑相关结合，缺乏统一理解
作者提出：从**信息论**的角度来分析问题，即不完全多视图聚类中的数据恢复和一致性学习是一体两面的，两者可统一到信息论的框架中，且两者相辅相成，**相互促进**。



如上图所示， x^1 和 x^2 分别用实线和虚线表示两个视图，**目标就是最大化灰色部分，最小化蓝色部分**。

它的一致性表示可以由互信息量 $I(Z^1, Z^2)$ 来衡量，即灰色部分 最大化互信息可以增加共享信息量，有助于数据恢复

数据的可恢复性用条件熵来表示，即蓝色区域，对于 $H(Z^i|Z^j)$ 当且仅当条件熵为0， Z^i 完全被 Z^j 所决定 促进数据的恢复性，即可以将缺失数据表示出来， 同时，蓝色部分属于是 x^1 和 x^2 互不包含的部分，因此最小化条件熵也有利于其一致性学习



因此，提出三个损失函数来解决问题：（ λ 为平衡系数）

$$\mathcal{L} = \mathcal{L}_{cl} + \lambda_1 \mathcal{L}_{pre} + \lambda_2 \mathcal{L}_{rec}$$

(1) **视图内重构损失**，主要用于学习各个视图数据的视图特殊表示，由一系列独自的自编码器重构损失组成；

$$\mathcal{L}_{rec} = \sum_{v=1}^2 \sum_{t=1}^m \|X_t^v - g^{(v)}(f^{(v)}(X_t^v))\|_2^2$$

其中f为编码器，g为解码器，高维数据投影到低维的表示空间，同时进行一些约束从而实现跨视图一致性，各个视图采用的是各自映射 Coordinate Representation

作用：

- 1.约束了子空间的表示，可以**避免平凡解**，需要返回原来视图。
- 2.从表示还原到原始输入空间，其实就具备了**对缺失数据还原的能力**

2) **跨视图对比学习损失**，通过最大化不同视图间的互信息学习多视图一致性；

$$\mathcal{L}_{cl} = - \sum_{t=1}^m (I(Z_t^1, Z_t^2) + \alpha(H(Z_t^1) + H(Z_t^2)))$$

其中，直接利用 X^1 和 X^2 对于 Z^1 和 Z^2 表示的互信息作为损失函数

最大化信息熵有两方面优势：1.增加混乱度，避免平凡解2.使表示具有更多信息量

3) **跨视图对偶预测损失**，通过最小化视图表示的条件熵进而实现视图数据恢复。

$$\mathcal{L}_{pre} = \|G^{(1)}(Z^1) - Z^2\|_2^2 + \|G^{(2)}(Z^2) - Z^1\|_2^2$$

让 Z^1 和 Z^2 互相预测，从而最小化 $G^{(1)}(Z^1)$ 和 Z^2 的差距，以及 $G^{(2)}(Z^2)$ 和 Z^1 的差距

最小化条件熵可以降低不一致，也就相当于增大一致性

1.5 实验部分

1.5.1 数据集验证

Missing Type	Method	Caltech101-20			LandUse-21			Scene-15			Noisy MNIST		
		ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Incomplete	AE ² Nets [51]	33.61	49.20	24.99	19.22	23.03	5.75	27.88	31.35	13.93	38.67	33.79	19.99
	IMG [53]	42.29	58.26	33.69	15.52	22.54	3.73	23.96	25.70	9.21	—	—	—
	UEAF [44]	47.35	56.71	37.08	16.38	18.42	3.80	28.20	27.01	8.70	34.56	33.13	24.04
	DAIMC [11]	44.63	59.53	32.70	19.30	19.45	5.80	23.60	21.88	9.44	34.44	27.15	16.42
	EERIMVC [27]	40.66	51.38	27.91	22.14	25.18	9.10	33.10	32.11	15.91	54.97	44.91	35.94
	DCCAE [43]	40.01	52.88	30.00	14.94	20.94	3.67	31.75	34.42	15.80	61.79	59.49	33.49
	PVC [23]	41.42	56.53	31.00	21.33	23.14	8.10	25.61	25.31	11.25	35.97	27.74	16.99
	BMVC [52]	32.13	40.58	12.20	18.76	18.73	3.70	30.91	30.23	10.93	24.36	15.11	6.50
	DCCA [1]	38.59	52.51	29.81	14.08	20.02	3.38	31.83	33.19	14.93	61.82	60.55	37.71
	PIC [40]	57.53	64.32	45.22	23.60	26.52	9.45	38.70	37.98	21.16	—	—	—
Complete	COMPLETER	68.44	67.39	75.44	22.16	27.00	10.39	39.50	42.35	23.51	80.01	75.23	70.66
	AE ² Nets [51]	49.10	65.38	35.66	24.79	30.36	10.35	36.10	40.39	22.08	56.98	46.83	36.98
	IMG [53]	44.51	61.35	35.74	16.40	27.11	5.10	24.20	25.64	9.57	—	—	—
	UEAF [44]	47.40	57.90	38.98	23.00	27.05	8.79	34.37	36.69	18.52	67.33	65.37	55.81
	DAIMC [11]	45.48	61.79	32.40	24.35	29.35	10.26	32.09	33.55	17.42	39.18	35.69	23.65
	EERIMVC [27]	43.28	55.04	30.42	24.92	29.57	12.24	39.60	38.99	22.06	65.47	57.69	49.54
	DCCAE [43]	44.05	59.12	34.56	15.62	24.41	4.42	36.44	39.78	21.47	81.60	84.69	70.87
	PVC [23]	44.91	62.13	35.77	25.22	30.45	11.72	30.83	31.05	14.98	41.94	33.90	22.93
	BMVC [52]	42.55	63.63	32.33	25.34	28.56	11.39	40.50	41.20	24.11	81.27	76.12	71.55
	DCCA [1]	41.89	59.14	33.39	15.51	23.15	4.43	36.18	38.92	20.87	85.53	89.44	81.87
	PIC [40]	62.27	67.93	51.56	24.86	29.74	10.48	38.72	40.46	22.12	—	—	—
	COMPLETER	70.18	68.06	77.88	25.63	31.73	13.05	41.07	44.68	24.78	89.08	88.86	85.47

缺失模式“Incomplete”表示50%的数据有缺失，“Complete”表示数据完整。红色与蓝色分别标记出最优与次优的结果

不难看出，在四个广泛使用的多视图数据集上进行对比实验，取得了比较好的成绩

1.5.2 丢失率和聚类演变

□ Performance with Different Missing Rates

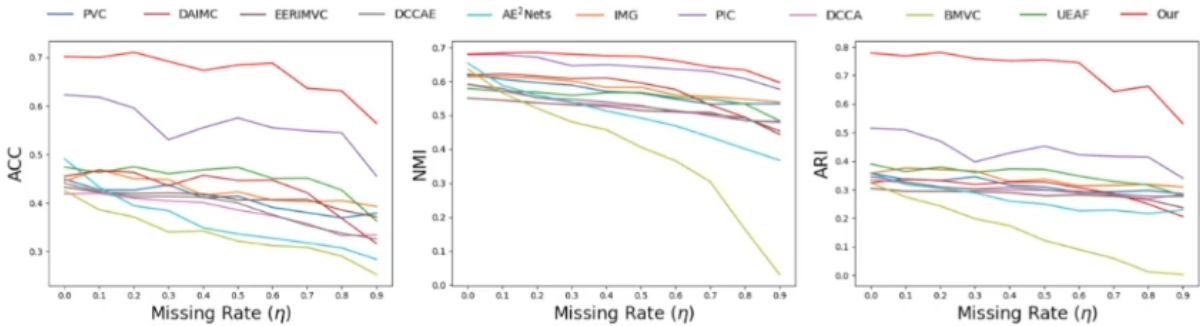


Figure 3. Performance comparisons on Caltech101-20 with different missing rates (η).

□ Evolution of Cluster Assignments

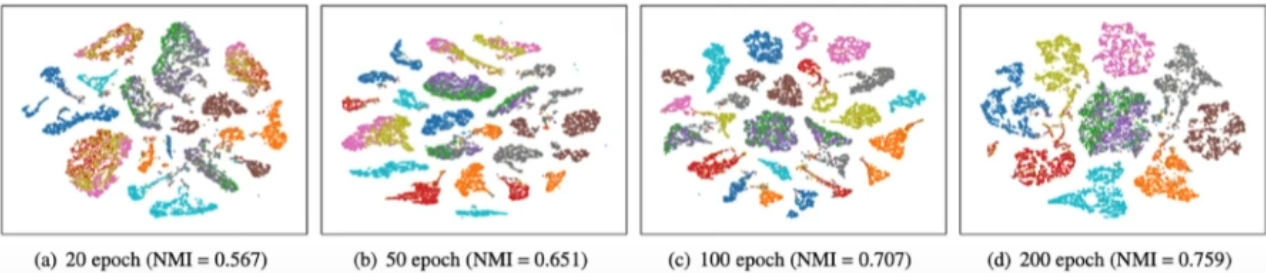


Figure 7. t-sne visualization on the Noisy MNIST dataset with increasing training iteration.

在200个epoch就趋于收敛

1.5.3 数据恢复（针对于下游任务）

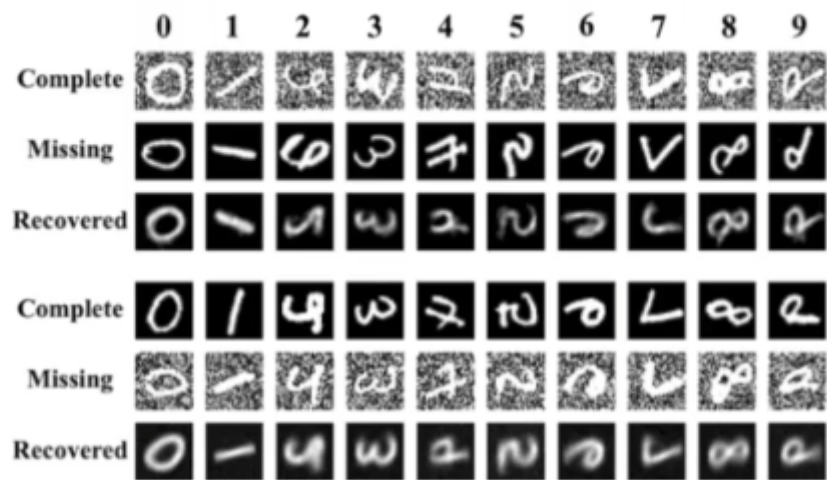


Figure 6. Data recovery on Noisy MNIST. Line 1 and 4 are complete views, Line 2 and 5 are missing views, and Line 3 and 6 are the recovered results from the complete view.

我们期待的结果是：第三行通过第一行恢复，结果应该与第二行接近，四五六行类似

现象结果是：

从第六行与第五行对比，没有恢复出噪声背景， 这呈现出它会恢复重要信息，如语义信息，而不会恢复出噪声信息，因此第六行与第四行一样干净。

通过**语义信息和噪声背景**，可以看出两个视图的一致性和不一致性

然而这种数据恢复，面向更多的是下游任务，只会恢复共享信息

1.5.4 消融分析和参数分析

□ Ablation Study

Table 2. Ablation study on Caltech101-20. In the table, “✓” denotes COMPLETER with the component.

\mathcal{L}_{pre}	\mathcal{L}_{cl}	\mathcal{L}_{rec}	ACC	NMI	ARI
		✓	33.65	31.60	16.43
✓			38.61	37.65	26.50
	✓		46.69	58.03	41.86
	✓	✓	55.75	59.35	58.88
✓		✓	54.70	52.63	43.49
✓	✓		64.59	62.11	71.07
✓	✓	✓	68.44	68.57	75.44

$$\mathcal{L} = \mathcal{L}_{cl} + \lambda_1 \mathcal{L}_{pre} + \lambda_2 \mathcal{L}_{rec}$$

□ Parameter Analysis

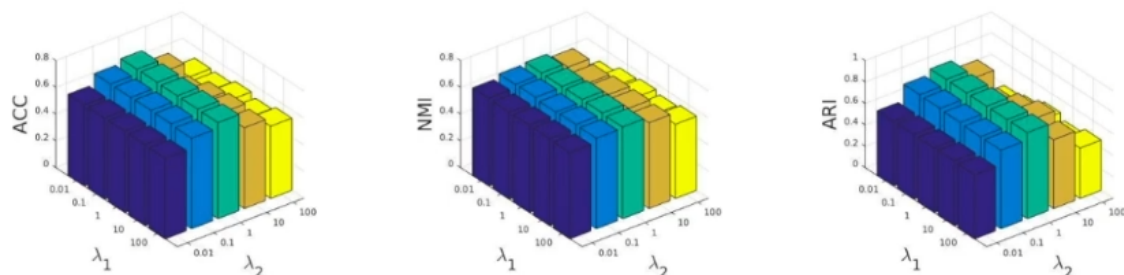


Figure 4. Parameter analysis on Caltech101-20.

消融学习和参数分析，证明其每个损失都很重要，而且参数选的合适

1.6 总结

针对不完全多视图聚类方法的挑战：1) 如何在不利用标签信息的情况下学习一致的多视图公共表示；2) 如何从部分缺失的数据中还原完整的数据。作者提出了独立的见解，采用信息论的角度提出了：数据恢复和一致性学习是**一体两面**的，两者可统一到**信息论的框架**中。

其中，互信息能用于衡量其跨视图表示间的一致性，而条件熵可用于衡量跨视图的可恢复性。最大化互信息可以增加共享的信息量，通过其共享信息量也有助于数据恢复；最小化条件熵能够有利于数据恢复，反过来也降低不一致性，从而有助于提高一致性；**两者相互促进**。

最后实验部分也取得很好的效果，其数据恢复能很好地验证其成果：通过语义信息和噪声背景，可以看出两个视图的一致性和不一致性，

但需要注意的是，其数据恢复面向的更多的是**下游任务**，如聚类，因此考虑的是共享信息量作为恢复来源