

CCF-CV 走进高校海大站

常琳

2016 年 8 月

今天上午和实验室的同学去参加中国计算机视觉专委会走进高校报告会。说实话，一开始我是拒绝去听的，因为要走很远的路去图书馆，报告内容还都是有关深度学习的。要是不去的话，我可以趁着这个时候把 MAC 捣鼓好，当然，这只是个借口，每次去听报告之前我都会习惯性的拧巴一会儿，找各种不去的理由，然而到最后还是要去的，因为确实能学到东西，这叫什么？叫“命”！

以上，凑字数。

因为这次报告确实是听得不好，听得不懂，没啥东西写。

第一位报告人王亮博士的报告内容是深度学习及视觉模式分析，对于一个初学者来说，这大概是三篇报告中最好的了，因为内容浅显易懂。王亮博士首先介绍了首先介绍了大规模特征表达和深度学习能够火起来的原因，大规模特征表达其实也就是大的数据的表达，我们需要解决的问题是鲁棒性特征表达；我们知道，深度学习能够得到较好的效果，需要有大量数据的输入，而深度学习能够火起来，也是因为有了大数据。接着讲了近十年深度学习火起来之后的几个发展阶段，概括地讲就是从 RBM+Autoencoder 到 CNN 再到 RNN 阶段，RBM+Autoencoder 是我目前正在研究的一个小知识点，暂时还没有很清晰的认识。CNN 到 RNN 阶段，侧重点是从静态图到动态视频，CNN 侧重于静态图的分析，RNN 与时间序列有关，可以用来分析视频内容。最后还讲到了大规模知识迁移的问题，其中的迁移学习与传统学习比起来确实有优势，如图 1。这种优势主要体现在学习的前后联系之中，更接近于人脑的学习方式。

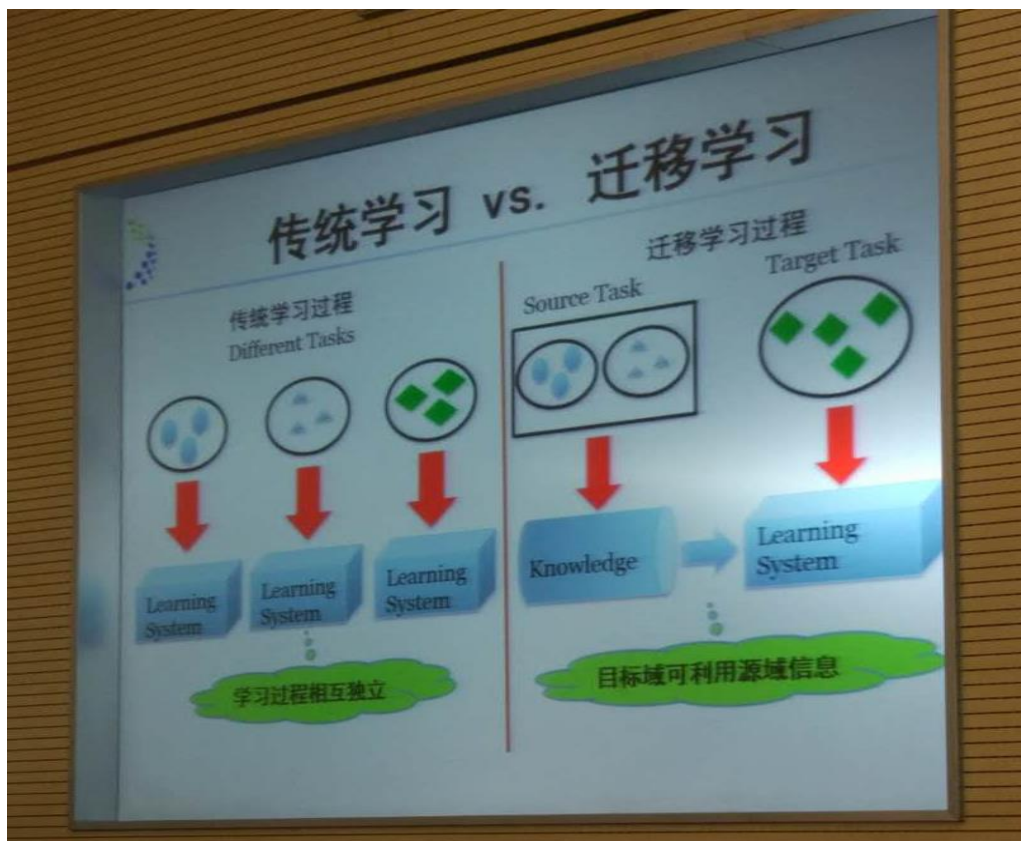


图 1 传统学习与迁移学习

由于时间比较紧迫，这次没能给我带来思考的内容我不想啰嗦太多，听完王亮博士的报告，接下来一周我会根据报告内容研究下玻尔兹曼机 (BM) 和 autoencoder。对了，还有一个惊喜的发现就是，原来降维也可以集成的方法，也许真能将传统方法和深度学习的方法结合起来做尝试。

第二位报告人是姜育刚博士，报告的主要内容是用深度学习对视频内容进行识别，还有一堆我没太看懂的神经网络结构图。在我看来，相比于图像的场景识别，对视频内容进行识别更加有意义，除了识别人脸或者其他目标物识别这种任务，我们对于某些事件的判断还是需要一段视频做参考的，由于视频是由许多帧图像组成的，它可以看做深度学习里的大数据样本，单个图片就是小数据了。不过报告中提到，视频识别要比图像识别难，我本来以为视频识别就是把视频的每一个帧当做一幅图像来识别，再把识别结果累加，但似乎不是这样，还需要考虑视频轨迹这些相关内容。此外，报告中还提到了类别关系对于识别的应用，以及在特征之间加一些关联关系的约束，这两个方法在这里是以深度学习为背景提出的，我做的分类实验还是太少，目前还不能在传统分类方法上实践这两个方法，以后在研究分类问题时，在学习理论知识的基础上，也要看下是否能将这两个方法融合到特征提取与分类中。

第三位报告者是李玺博士，这里我没有接收到太多信息，因为走神了，这里只想记录一下我走神的过程：

李博士提到他们的关注点已经从 Image Feature 设计到 Image Matching，虽然之前我已经知道深度学习不必关注特征的设计，只考虑网络的设计，但这里还是刺激到我了，我该怎样把自己的识别方式交给计算机呢？应该教它识别什么特征，效果才会秒杀深度学习？人工设计特征什么时候才能设计出好的特征呢？设计好了特征还要设计更好的分类器和核函数，到底是降维管用还是升维管用，需不需要把降维和升维一并研究了昵？

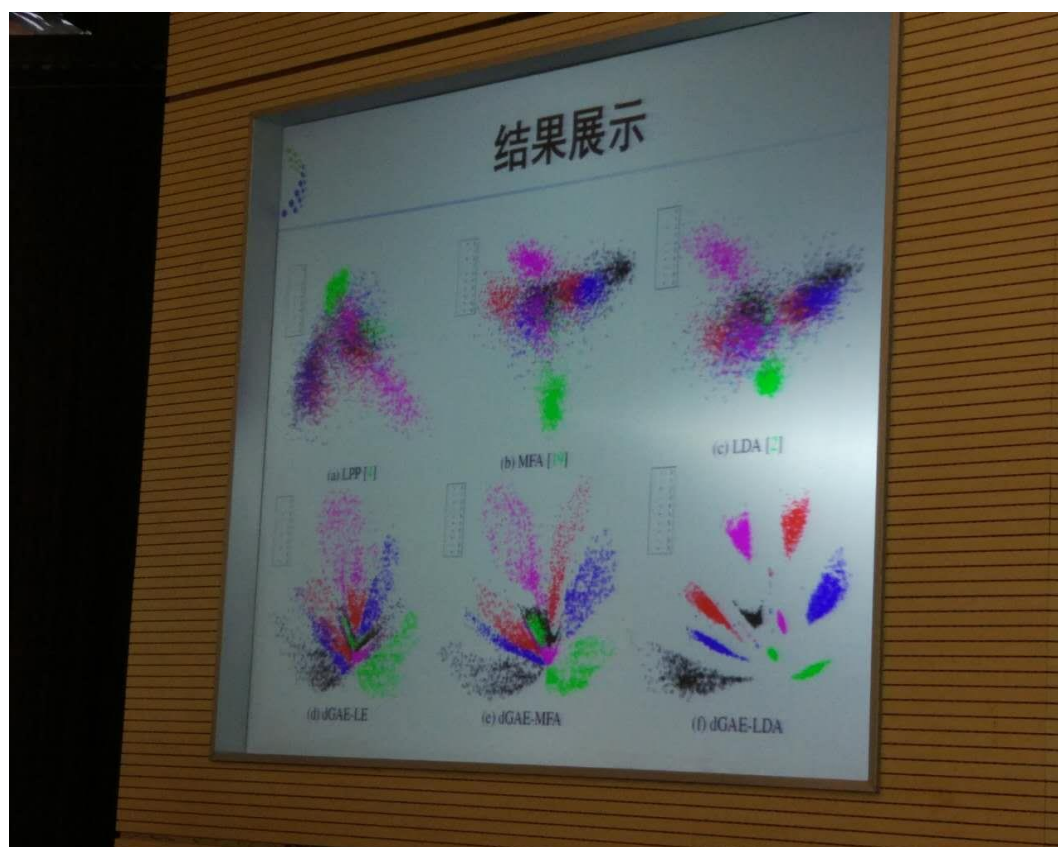


图 2 几种降维方法的结果展示

图 2 似乎把传统方法核深度方法给结合起来用了，看来我也需要研究一下深度学习了，但

是只有一年的时间啊,时间这么短我又这么笨,要不要考虑读个博继续研究呢?等等,读博?我怎么会在这里读研呢?我怎么会踏进科研行业呢?别说降维的知识没有看多少,数据集还没弄好呢,编程还不会呢,RBM 的各种变形又该怎么去研究,还要不要和 autoencoder 一起研究?我还要用机器学习分析开家煎饼果子咖啡铺子呢,到时候传统的机器学习方法和深度学习的预测结果不一致怎么办呢……至此,我的思绪已经足以打发剩下的时光了。回神后细想,我思考了这么多,对今后的科研一点帮助都没有,还白白浪费了倾听大牛声音的机会,实在是太不值得。

学渣和学霸的差距就是在这种情形下拉开的,关于走神这种事情究竟是不是可控的我还不明白,大家都说是可控的,那么姑且就算它是可控的吧。而听报告时跟着报告人的思路一起走下去,每当发现自己走神的时候就用上面的说法把自己拉回来,也许就不会总走神了。其他人每次开完会都能记住每个报告的要点和思路,我却只能记住自己走神想了些什么,这里还是需要每日反思的。