近期工作总结

2019-08-30

一、关于自监督学习

传统的监督学习：需要大量labeled data，人工标注数据费时费力且昂贵。

Self-Supervised：提出pretext tasks，在训练Objective functions的过程中学到features。

> The networks can be trained by learning objective functions of the pretext tasks and the features are learned through this process.

在训练的过程当中自动生成“伪标签”。

自监督学习：使用无标签的数据，通过自动生成的标签来训练（比如旋转图片，以旋转角度为标签；或snowball中的设计一种vector格式，以句子相应的vector作为标签），学到数据的feature后，迁移到其他任务上。

个人理解，自监督学习是没有label的，但是是通过输入数据按某些规则自己生成一些label作为参照（比如snowball里面的pattern，是通过seed set去和语料库按模式匹配而生成的，反过来又用它来生成新的seed）。

在Github上找到了一个仓库，里面收集了Self-Supervised相关的论文资源。为了了解自监督的思想，我先找了下面这篇视觉领域的综述文章读。

1. Self-supervised Visual Feature Learning with Deep Neural Networks: A Survey

文章地址：https://arxiv.org/pdf/1902.06162.pdf

文章指出，采用自监督可以省去收集并人工标注数据的时间、金钱上的浪费，可以利用大规模无标签的数据集，不用人工标注，而学到图像的特征。本文主要介绍了一些CV里面常用的卷积网络模型以及Task，后面又介绍了和自监督相结合的深度学习网络结构，以及使用到自监督的方法和评价指标。

一些深度卷积网络模型很广泛地被用作pre-trained的模型，经过微调来应用到其他任务上，因为：（1）从大规模数据集上学来的参数提供了一个良好的起始点，可以帮助模型在具体任务上训练时**收敛的更迅速**；（2）经过大规模数据集训练过后，模型已经学到了一些层次的特征，可以在训练其他任务时**防止过拟合**，尤其是对应数据比较少的时候。然而，深度神经网络的效果十分依赖于数据量的大小，但制作这些大数据集（比如ImageNet，OpenImage）需要大量的人工标注，十分费时费力。而自监督就是这样一种方法，它利用大规模无标签的数据，通过设计一个pretext task（这里我暂且翻译成“辅助任务”），使网络在训练的过程中学到相应的feature。这种“辅助任务”需要满足两个条件：一是卷积网络需要通过捕捉到图像/视频的visual features来解决辅助任务，二是辅助任务的“伪标签（pseudo labels）“能够通过原始数据image/video的自身属性自动生成出来，而不需要人工标注。这就是文中所说的自监督的基本思想。

- 文章第二部分主要是介绍一些概念，如监督学习、半监督学习、弱监督学习、无监督学习、自监督学习。这里作者将自监督归进无监督中，因其不需要人工为数据标注出label。
- 第三部分是将常见的深度神经网络架构。处理图像常见的有AlexNet、VGG、ResNet、

GoogLeNet、DenseNet；介绍了一些处理视频的方法；以及使用RNN等等。

- 第四部分介绍了一些常用的下游任务。
- 第五部分介绍数据集，暂略。
- 第六部分介绍学习Image Feature的方法。主要包括三部分：Generation-based methods，context-based methods以及free semantic label based methods。
    - 第一种的Generation-based methods主要包括：基于GAN的图像生成，高分辨率图像生成，图像缺失区域补全，灰度图像染色等。这些任务的pseudo label一般就是其原图片本身。这方面的前序工作是Auto Encoder。
    - 第二种的context-based methods主要是将图像的上下文信息加入到辅助任务设计中去，如context similarity、spatial structure、temporal structure作为监督信号。在训练辅助任务时这些feature会被卷积网络捕捉到。
    - 第三种的“Free Semantic Label”是指具有语义的，但不需要人为标注而是自动生成的标签，例如segmentation masks，depth images，optic flows，以及可以被游戏引擎或硬编码方法生成出来的surface images。
- 第七部分介绍学习Video Feature的方法。主要包括四部分：generation-based methods, context-based methods, free semantic label-based methods以及cross modal-based methods。
    - 第一种的generation-based methods,也是主要包括这几个方面，video generation with GAN，video colorization，video prediction。对于这些任务，其pseudo label一般是视频本身。因为这些伪标签不需要被人工标注，因此可以被认为是自监督学习。
    - 第二种的context-based methods：视频包含不同长度的frames，包含丰富的时间、空间信息。视频里这些继承下来的时间信息可以被看成监督信号，用作自监督来学习feature。主要包括temporal order verification（判断一组frame的顺序是不是正确的）和temporal order recognition（给一组frame排序）。
    - 第三种Cross Modal-based是指其可以从多个数据流的共同出现来一起学习video features，包括RGB frame sequence，optical flow sequence，audio data，camera pose等。
- 第八部分对效果进行了简单的分析。这种利用自监督模式，通过pretext task学习feature的效果一般比不上监督学习。但在几个well-designed的pretext的作用下，学习image feature的能力是可以和监督学习媲美的。
- 第九部分是说作者对自监督学习未来的一些想法（应该是针对图像领域的），有这三点：通过人工合成的数据来学习、通过Web数据来学习、用Multiple Pretext Task来学习。感觉对我们的问题帮助不大。
- 总结：感觉这篇文章介绍的图像/视频领域的自监督思想，都是通过设计一个good pretext task——用label-auto-generated的数据训练模型，使其学得feature，然后再应用到对应的下游任务上。暂时没想到和snowball之间有什么关联。。-_-||

2. Self-Supervised Relation Extraction from the Web

文章地址：这里

这是找到的一篇用自监督做关系抽取的文章，2007年发表的。应该也是比较古老的方法。目前还没读完、整理完。下次会一起发上来。

二、自监督学习 & NLP

此处待补充。网上关注到的NLP用自监督的工作很少。大部分地方都只举BERT作为例子。

待补充：Multi-Headed Attention笔记、Transformer笔记、ELMo、GPT、BERT笔记

三、自监督学习 & 关系抽取

首先，我去了解了一下什么是关系抽取。网上给出的解释是从大量的非结构化文本中抽取结构化的两个或多个命名实体以及他们之间的关系。

疑问1：snowball最后得到的输出到底是什么？是一堆类似<Microsoft, Redmond>的这种<o, l> tuple吗？文章一直在用Location和Organization举例子。起始过程也是先输入进去几个实例，然后滚雪球。个人理解，他应该只是想用这一对关系来解释一下他这个系统的工作机制，而实际上这个方法可以用来抽取不同的关系。如果是这样的话，那它不是要对每一种关系都手动列出若干个实例？语料库里存在那么多关系，这样也不现实呀？还是说他先人为预设一些关系，比如(location of), (work at)之类的，然后为每个关系分配几个实例进去滚雪球，最后相当于对每种关系都能生成一个tuple的集合？

疑问2：“关系抽取”是指抽出的relation（顾名思义），还是说是包含这个relation的tuple（参考snowball），还是Entity-relation-Entity的三元组？

在图像领域，自监督的常用套路是设计辅助任务，“辅助任务”会在输入数据上加上一个伪标签，然后用一个网络来预测这个伪标签。例如：把输入的图片旋转一个角度，然后预测旋转的角度。（也可以是对图片做一个transformation，然后预测出transformation）。它的主要意义是为传统监督学习构建大规模数据集的成本很大，所以大家逐渐开始关注如何用非监督方法来提取高质量的feature。用自监督训练出学到feature的网络后，再用其作为pre-trained模型，迁移到其他任务上去。

对于关系抽取，输入数据应该是语料库，大规模非结构化的文本数据。因此，类比上面思想，这里自监督要做的事情应该是学到某些“feature”（可能是语言的语义信息），然后再把网络fine-tune到能做关系抽取的任务上来。

COMET做的事情是将监督学习的GPT模型在ATOMIC数据集上训练，利用pre-trained模型具有的词embedding的语义信息，fine-tune到使其能够生成常识。这个相当于是generate的模型，来生成新的三元组。而snowball相当于是从语料库中抽取，来构建结构化的图谱（为什么是图谱？会有多个实体和多种关系构成网络吗？）。

用语言模型生成出一些新的relation，再把这些新生成的relation放到语言模型里tune一tune，然后再生成新的relation。“自监督”。“增量”。

四、之前的讨论整理

本来是在整理一个比较长的笔记，这个只是其中一部分内容。其他的还没整理完，先把这个备忘发给老师看一下，后面的我稍晚一些会发过来。

4.1 第一次讨论：2018-07-23

这次是留完论文任务之后第一次打电话，忘记录音了。印象里这次讨论主要是大致汇报了一下看三篇论文的情况，将三篇论文简单整理成了笔记，捋顺了一下思路，确定了一个初步的idea，以及制定了后续讨论的计划（每周二下午三点）。

4.2 第二次讨论：2018-08-06

这次微信电话我进行了录音，方便后面回过头来听，回顾一下当时讨论了些啥。这次主要是复述了一下对问题的理解。对比了snowball工作与COMET工作的可比性、异同。老师的答复是：多走了一步

COMET：利用人工标注的Transformer模型进行迁移学习，使其迁移到生成常识上来

snowball：初始有一些种子，在语料库里不断迭代，也能发现许多不同的关系

需要思考的是：snowball和COMET这两个模型本质的异同。

snowball里面隐含的是：self-supervised的思想。不依靠人工标注。用余弦相似度，设定某一阈值，高于某一阈值的补充进来，不断地滚。目前的研究套路大多也是这样，利用embedding空间里的夹角余弦值，补充进来，relation也可以做embedding。通过某种方式发现新的关系，然后在关系的空间里分类，设定一些评估，然后不断的滚，也可以实现目的。

Transformer线的思想：为什么用这样的预训练模型也可以发现关系，而且结果是靠谱的，他背后的本质是什么?

snowball本质：有一个自监督机制，里面有一个对抗生成的思路，通过监督，使得更好的样本能够进来，不断地自增长。

用“两条腿”来理解——从自监督的角度来理解Transformer，从Transformer的角度来理解自监督学习，最后看看能不能把它们融合在一起，然后可以做的比原来更好一点，比如可以引入人工？因为用Transformer就不能引用人人工了，不能引入专家知识。在这个过程中假如可以引入专家知识，达到更大的效果，也是很好的。用到的知识面面很宽广，可以用到生成对抗（GAN），也有自监督的思路，跟Transformer也有关系，把这些线都搞清楚。先以知识为主，把这两条线都走通，先学的很清晰，然后再讨论研究。

后面又讨论了一些关于长线研究和短线、快速占坑的话题。提到了夏令营座谈会那次那个发ACL的同学混淆“数据增强” 和 “对抗样本”的问题。突然有一点感受就是，做科研还是应该先沉淀许多知识，然后在idea的层面上多思考思考问题的本质，后面再关注具体实现的问题。而且由于老师这边压力不是很大，不急着出成果，也就有了更多时间用来学习、思考，我个人认为这一点我以后会很受益。谈到想idea，感觉自己数学基础知识差的确实有点大，就上次讨论时曾经提到过“EM算法”、“核函数”、“随机过程”等等概念，我回去查了一下，发现需要学习《随机过程》、《统计学习方法》等等。本科阶段也只是学了高等数学、线性代数、概率统计、数值分析这四门课，会的也都是一些基础的东西，感觉我这些数学基本功不深的话，对问题很难有一个全面的理解。不知道研究生会不会有这些课程，或者需要自学一下某些数学课。感觉自己需要补一下数学知识。（老师能不能给列几门课程，我回去学一一下？）

后面说到，目前生命科学那边的一个分支也是自监督的线，学出一个知识图谱，在其上做推理等等。目前还是先学习，把这些基本概念都搞清，等到差不多年末的时候也可以有准备的进来。

COMET是利用Transformer这个语言模型，用task训练使其带着语义信息迁移过来，相当于snowball里面的哪一一步？——snowball通过现有的tuple去语料库里“学得”上下文vector，再通过这些vector去从语料库里生成新的tuple。这里的vector其实就相当于在这个问题下抽象出来的，能够生成tuple的“语言模型”。他的测试相当于人工监督，去筛掉一些不好的样本，而机器学习里面相当于是利用最小化loss，优化模型参数。一个是人工监督，一个是非人工干预。snowball相当于是一轮一轮的在训练，而数据是增量的。

这两个工作各有其长短的地方。Transformer这个语言模型本身非常复杂，但是机制比snowball简单；snowball的过程有它领先的意义，是一种自监督的、增量的训练。我们的目的就是在这两个点中间夹出一个新的点。

4.3 第三次讨论：2018-08-13

这周回去主要是沿着Transformer和Pre-trained模型这条线捋的。首先解释了一下Transformer的原理。

第一个问题：什么叫Attention？详细讨论了一下Attention机制。相当于是一个衡量词之间重要程度的关系矩阵，然后这个矩阵是可以被学出来的。这里提到，了解一个机制，一是了解其原理本身，二是其怎么实现的。

第二个问题：什么叫Multi-Headed Attention。为什么要Multi-Headed。为什么8个Head随机初始化之后可以分别关注到不同方面的特征？这里提到了EM算法。（我回去查了查，还没弄懂。。）分成Multi-Head以后，维度变小了，而且还能表达不同的方面。一是为什么分成了8个Head之后维度会小？二是它是怎么样被建模的？为什么每一个子空间是独立的，都能学出来？而且学出来的东西都是不一样的？直觉是，其借用了EM算法大体的一个思路。把每一个相互独立的特征空间，特征跟特征之间是有比较强的耦合的，是一个联合概率分布，一个联合概率分布就等价于一个子空间。一个联合概率分布是一个特征集。“高斯混合模型”。每一个特征集都是一个耦合比较紧密的子空间。在学习的过程中多个Head综合在一起会有一个互斥的作用。在一个Head编码了这块信息之后，其他Head就会被挤掉，就不用编码这一块信息。“狄利克雷过程”。

后面是说具体实现。首先说了一下Query、Key和Value三个矩阵是怎么来的。上次由于没提前准备文字材料，直接口头描述比较抽象。我接下来会整理一份Multi-Head Attention和Transformer的材料出来，提前发给老师。

提到概念：张量分解。 后面又讨论Transformer的Encode、Decode的问题。我自己Transformer解码的地方还没太搞清楚。

后面又提了我顺着pre-trained语言模型这一条线看的一些内容。从Word2Vec——ELMo——GPT——再到BERT。

最后我提了关于COMET里面的一个疑问，到底是输入s和r的拼接，o作为标签，还是sro都拼接到一起作为输入，这样的话标签是什么？后来讨论之后发现是理解插图理解的有偏差，图示的意思是数据格式的编码不同，和他模型的输入没关系。

提到：自编码器——Auto Encoder（回去补）

这周主要是学Transformer的细节以及了解Pre-trained模型的思想。老师希望先把目光放在思路上，对问题有一个整体的把握，之后再下去关注这些实现的细节。先关注这个思路有没有人做过研究。因此这周回去是理解自监督、看看自监督在NLP里面有什么应用，以及和我们的问题（snowball算法，关系抽取）如何结合到一起。用语言模型生成出一些新的relation，再把这些新生成的relation放到语言模型里tune一tune，然后再生成新的relation。