# Idea king

1：大于百分之90正确率的样本不进行惩罚，标签平滑的思想是削弱0-1防止过拟合。(低于百分之90往90偏)

1. 蒸馏学习中的特征迁移策略（attention transfer）（希望有人能写个keras形式的注意力层。）
2. 概率图与多任务的结合。事件检测有引入了大量的节点，与特征。
3. 联想图片。。补全图片，感觉和语义分割有些关系。
4. 变分推断计算期望的导数时可以

从频率的角度，叠加的大小可以是概率的大小，只需要对应放大的步长。

因为在足够大的样本下个数大致成比例，如果能

6尽可能考虑原有数据就可以塑造的问题（缺省。。）.而不是额外的特征。除非可以推断隐变量。

7神网能gibbs采样吗？

8神网非显示的对概率建模可以做到吗？（输入随机时可做。）

输入确定时，输出如何非显式的随机？

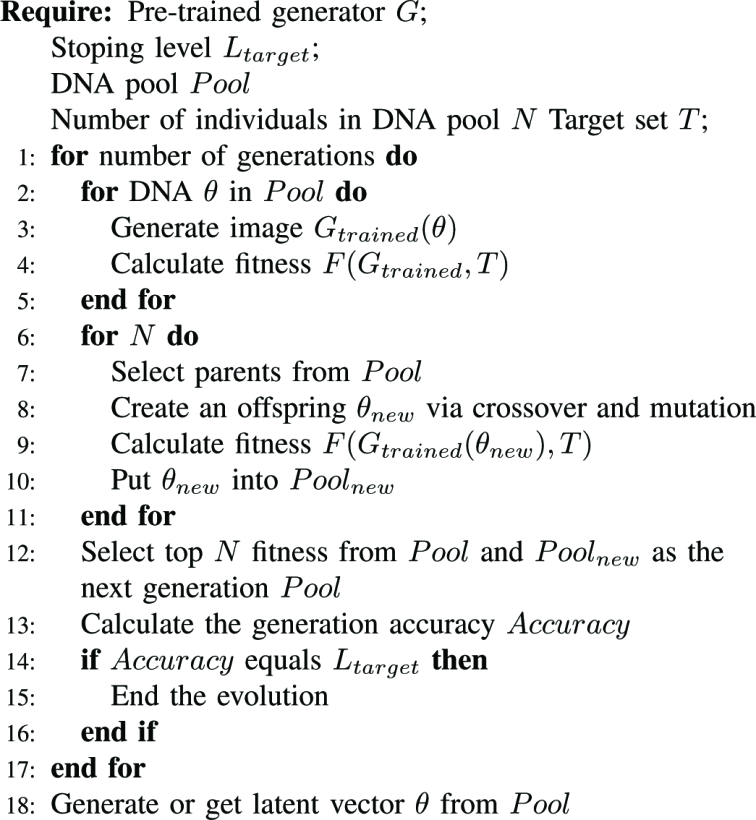
1. 非自编码式的推断生成网络。
2. 多任务的图模型，或者考虑关系的图模型，显变量用规则的方法生成。
3. 转换图结构才能构建大差异，以推断是否有关系。
4. 输入敏感翻译。大多数翻译会尝试通用模型，而部分时候用户可能会有一些自己想定义的东西，尤其是论文里。
5. 常识，多任务。‘
6. Numgcn,图神经网络与bert
7. 看gagan有感而言：也许可以放到文字领域去。

李明希瑞，他们有很多篇高等级会议的论文都是关于生成对抗网络的，而文字领域的gan网络目前还不成熟。说来最近大型语言预训练网络出来了这么多也许文本生成有更高的基线了。

文本生成现在最关心的是对话系统。其次是摘要生成。这些的输入都是明确定义的变量，而非图片生成中的隐变量。

不过类别要求式的还是有迹可循。

生成类的任务都相对取巧一点。



16：神经元尽可能多，万能近似。

17：属性混淆的可能性（人鼻子，狗鼻子）用gcn去解决。

18：gan的分类器往往生成时没用上，其实是可以用上的。只输出真实图片。

19之前一直觉得直接假设分布表达能力太有限了。。（线性）组合式的也许更好。

1. 网页翻译的排版太诟病了。。
2. Cca，多变量情况下：如像素一大堆的图，数据库涉及很多属性，映射X和Y，原始是拿knn做，在映射空间相对连续，最近零，等值相加求平均，欧氏距离加权，自己设的加权，等值效果最好。。。（1）想法：权重靠学习。（2）编码方式，数据库查询计划的编码。粗浅编码=》映射后的编码作为x，用cca学到的映射（核矩阵形式）编码
3. Plan structure tree用tree lstm
4. Wikisql目前完全和同一个数据库相关，无法迁移（尤其是如果两个表的结构不同但属性相似时）。Gcn编码，语言层，多表。多表关联。可迁移性。语言层，seq2sql检测in和where。迁移学习增加商用性。模糊查询基本做不到（模糊度），列。。假执行一次看对不对，再encode，错误词修正。强化学习（听说rl做nlp是大忌）
5. Emotion-cause pair
6. 既然有核回归那么就有核感知机，那么就有核网络。只是核网络中间层的辅助输出得格外推。理论上来讲只要和距离（内积），线性变换有关就可以往核方法上靠。不过一层的是已有的方法，比如径向基网络。
7. 分形网络，玄过头了。。
8. 群（加法），域（乘法）（向量域，或者说向量空间），环（除法）（矩阵环）。向量空间，希尔伯特空间（内积乘完后是不封闭的。会投到R里去。）。说到底有很多运算对于搭建结构的神经网络是封闭的，可以用群论的知识把神经网络结构的故事讲得更好。
9. Electra与t5已经公开了，正常来说该有人上模型跑结果。。
10. 乱序模型本应该有种更优美的方式。。（xlnet），自回归族中都更关注整个句子被生成的概率（即句首下x1，x0x1下x2的概率）而不直接关注这个句子合理的概率。（）
11. 我惊了。Electra与用预训练模型生成事件等几篇论文，都表明了bert在生成上还是有一定生命力。。
12. 即使0-1编码的词向量也具有2^100可能性，脉冲神经网络词向量表征也许意外的不错。。就是不知道怎么训练才合适，，梯度训练再近似归01？？

不过很大程度上这是因为词向量只考虑了内积意义上的相似性，，如果多加约束与应用也许可以限制死。

1. Elmo对付生僻字词翻译有益。
2. Zero-shot learning，零样本学习和许多建模知识有关。需要足够多的类别描述才能（人工或学习。）。
3. online learning，目前没看到关注nlp的online learning的

元技术核心：gan，gcn，迁移学习（zsl与预训练）。

Gan+ZSL的话会很难，但是有迹可循。ZSL在事件检测的应用还很少。

Gcn（gat）+gan。

辅技术：多语言（跨语言向量），多任务，多粒度，注意力（监督注意力）。

不靠谱技术：复数网络。VAE

它拿的高斯分布（拉普拉斯近似）近似的迪利克雷先验。gumbel分布可能更适合。。。或者二者混合。

如果zero shot用的隐藏vae如何

多个gan，但可能泛化的会太厉害。但如果成对可以gcn。

段落级别的构图？

文档级别的构图。

也能和ZSL结合学习事件之间的相互关系。

Kbp需要人做。Bert zsl需要人做。Bert的显注意力也需要人做。无论是拿来监督还是分类。

HMEAE，jrnn看不懂，

Zsl基本依赖于手工设计的特征。

如果gat似乎更适合针对每个事件。

监督注意力

Bert也能显示的监督注意力。：前面分别有监督arg注意力，与trigger注意力，与有event的句子的注意力。

注意力权重会一直乘上去。即最后的注意力矩阵。

即[1，1，1，1，1]=》[]

ERINE得有人去踩。包括实体标注与如何输入。

Attention转风格，转生成。高斯过程生成。基于注意力的生成。如果我们单纯去做事件

调参数会炸。。

用pooling把被分开的词或长度超过1的实体解决。

注意力是条路

写起来。替代transformer中的结构。

零样本学习，外加vae

Attention抽象。