李宏毅老师的

机器学习说出为什么[我知道]

还有机器也会产生错觉

终身学习：life long learning

机器也要终身学习，如果每个模型都学一种任务的话，那将存不下那么多模型。

学习如何学习，meta learning

Anomaly Detection:可以让机器知道我不知道这件事,只是找跟训练资料不太一样的东西，不一定是不好的东西

open set recognition

异常侦测：

Attack and Defense:

侦测带有恶意的程序，如何攻击一个model，如何防御一个model

攻击要做的事情：

original image + delta X = Attack Image

攻击分有目标的攻击和没有目标的攻击:Non-targeted Attack和Targeted Attack

当前的机器学习很流行，但是要能防御攻击，就是对方制造的一些要骗过machine的input,所以不仅要强还要能对付人类的恶意攻击

怎么找出那个特别的signal,

1.正常的训练是输出一个图片，跟正常的y\_true越接近越好，minimize crossentroy，这两个distribution的crossentroy；训练网络的参数。

2.Non-target Attack，找到另一张图片，放到network输出y‘，希望产生的输出跟y\_true越远越好，这个时候network的参数是固定好的，改的只是input。

3.target Attack：不仅仅希望找到一个input跟y\_true越远越好，跟y\_false越近越好。

4.Constraint:d(x0,x’) <= delta

–--L2-norm，就是每一维度上delta的平方和

----L-infinity max(delta x1,delta x2);指两个向量差距到底有多大

如果将一张图片的pixel改变一点点，视觉上感官不出来；还有一张是pixel改变大点，视觉上能感官出来，如果用L2-norm来衡量的话，L1-infinity两者来衡量。

Just like training a neural network but network parameter @ is replaced with input x’.

Gradient Descent

start from original image x0

For t = 1 to T

xt = xt-1 – theta delta L(xt-1)

How to attack:

x\* = arg min L(x’)

L2-norm: 如果是在圆圈的外面，半径外面

L-infinity

从x0是加的杂讯，如果将x0这个杂讯也有一定的分布，在一定的分布上会是一个某个类别的最高概率。

攻击方式就是不同的手法和限制。

x\* = arg min L(x’)------->不同的优化手法

d(x0,x’) <= sigma ----→ 不同的限制

Fast Gradient Sign Method (FGSM)

2种类型的防护：

Passive defense：

proactive

tranformer:就像变形金刚一样，用的是self-attention，RNN不容易被平行化。

Bert 就是unsupervised train的transformer(seq2seq model with ‘self-attention’)

如果是单向的RNN，输出b4的时候，输入a1,a2,a3;输出b3的时候，输入a1,a2;

如果是双向的RNN，输出b4的时候，输入a1,a2,a3,a4的时候；要输出b1-b4把输入都统统都看一遍。RNN经常用来处理输入是序列的情况。

RNN不容易并行化，因为要算出b4,必须把输入的a1，a2，a3，a4都看一下，接下来有用CNN替换RNN的情况，每个CNN的filter都考虑一个序列的窗口；

可以加多层的filter在第一层的输出上可以加一层。

AMR:自然语言领域的单根有向无环图，AMR 解析评测广泛采用的是一种称为 Sma

tch 的度量方法，它在对两个 AMR 图进行匹配

度计算时, 首先将每个 AMR 图转化成一个逻辑三元组(trile ) 的集合，其中每个三元组表示图中的一 个顶点或一条边；Smatch 方法计算两个三元组集合之间的匹配或重叠 程 度，度 量 指 标 也 分 为 准 确 率，、 召回率，F 1 值，搜 索 两 个 图 之 间 的 变 量；

AMR 解析算法

**基于图的解析方法**，JAMR 使用管道式的方法分阶段完成 AMR 解析的两个子任务 : 概念识别和关系识别。1）概念识别阶段的任务是将输入句子中的词片段映射到概念图片段。其中, 概念图片段 大 部 分 情况只含有单 个 的 标 注 概 念 节 点，但 它 有 时 也 可 能 是 一 个 含 有 多 个 概 念 节 点 和 标 注 边 的 子 图 结 构，JAMR 将概念识别任务看成是一个序列化标注和切分任务 , 使用半马尔科夫模型进行 概 念 识 别，并 设计了一个动态规划算法进行解码 。概念识别阶段的输出结果是概念图片段的序列。

2）关系识别是在第 1 阶段输出的概念图片段 序 列 中 , 通 过 添 加 带 标 注 的 有 向 边 而 构 造 一 个 AMR图。AMR 将关系识别看作是一个基于约束的组合优化问题，在给定约束集的条件下 , 搜索具 有 最 大 值

的子图。子图的分值则采用按边分解的模式和线性模型进行计算 , 然后采用拉格朗日松弛算法调整该生成图以满足 确 定 性 性 约 束，使得从图中每个顶点发出的有向边的关系标注类型；

**基于转换 的解析方法**