AI商用产品体验–报告

一. 基本信息

- 体验平台：
- 体验时间：
- 体验身份：

拆分需求	（1）目标用户画像和实际用户画像 （2）功能的实际使用场景，流程和优化 （3）数据报表，下一步的运营策略
个人体验习惯	* 优先评估技术实现难度 * 优先以技术角度思考用户体验的原因 * 优先寻找具有替代性的同类Github开源项目
体验频率	
体验行为	
体验中的评估和反馈	
可能的认知偏见	

- 体验的产品

名称	版本

二. 体验内容

其他

直播 2020.05.26 – 不需要真实数据的模型窃取方法

主讲人： 旷视成都研究院实习生–周鸣一

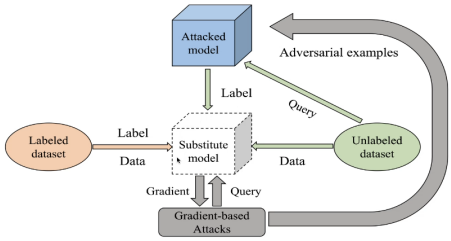
- 业务内容

一级类别	二级类别	应用	tricks
对抗样本	基于梯度攻击		因为无法获得目标模型梯度，所以需要数据集训练替身

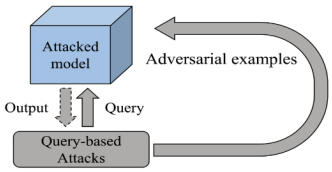
			模型进行梯度计算并迁移 一般的防护：通过保护数据以防止替身模型被获取
	基于查询的攻击		一般的防护：限制单个输入样本的被查询次数

攻击方式流程示意

基于梯度的攻击

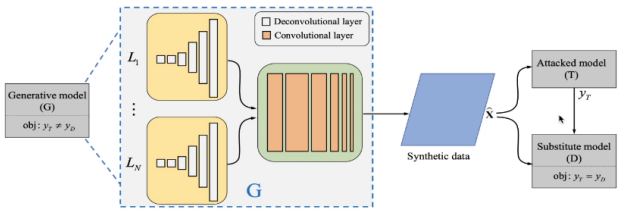


基于查询的攻击



论文中攻击方案的设计

不需要真实数据的模型窃取



the objective of G: generate samples $\widehat{\mathbf{X}} = G(\mathbf{X})$ and let $y_D(\widehat{\mathbf{X}}) \neq y_T(\widehat{\mathbf{X}})$
the objective of D: guarantee $y_D(\widehat{\mathbf{X}}) = y_T(\widehat{\mathbf{X}})$

这样的基本模型结构容易造成GAN的模型塌缩

DaST

The loss for controlling the label of synthetic samples is formulated as:

$$\mathcal{L}_C = \text{CE}(T(G(\mathbf{z}, n)), n). \quad (1)$$

Then we use D to replace the T in Eq. (1):

$$\mathcal{L}_C = \text{CE}(D(G(\mathbf{z}, n)), n). \quad (2)$$

The loss for G is

$$\mathcal{L}_G = e^{-d(T,D)} + \alpha \mathcal{L}_C, \quad (3)$$

The loss for D is

$$\mathcal{L}_D = d(T(\widehat{\mathbf{X}}), D(\widehat{\mathbf{X}})). \quad (4)$$

不需要真实数据的模型窃取方法

Algorithm Mini-batch stochastic gradient descent training of the proposed method DaST.

$\#$ acc denotes the accuracy of D. att denotes the attack success rate for the attacks generated by D.
1 : **While** iteration $< \delta$ **or** acc, att do not increace
2 : Generate m examples$\{\widehat{\mathbf{X}}^{(1)}, \ldots, \widehat{\mathbf{X}}^{(m)}\}$by G.
3 : Update the substitute model :
4 : $\mathcal{L}_D = d(T(\widehat{\mathbf{X}}), D(\widehat{\mathbf{X}}))$.
5 : Update the generative model :
6 : $\mathcal{L}_G = e^{-d(T,D)} + \alpha \mathcal{L}_C$.
7 : **end for**

Attack Scenario: attackers can probe the output labels of the attacked model (DaST-L). attackers can probe the output probability of the attacked model (DaST-P).

- 遇到的业务难题

（1）根据不同目标设计攻击模型输入信息的多少

（2）方法中多分枝结构用于帮助提高生成器控制类别的能力，以改善模型坍缩，副作用是对于大规模数据集耗费资源多，训练时间显著增长
（3）训练D模型时候需要大量查询目标模型，虽然可以减少实际部署查询次数

- 其他材料

Github：https://github.com/zhoumingyi/DaST
PPT：
录音：
视频回放：

NLP

直播 2020.04.01 – BERT在美团搜索业务中的应用

主讲人： NLP部算法专家–王金刚

- 业务内容

一级类别	二级类别	应用	tricks
单句分类	情感分析	垃圾评论识别和过滤 细粒度情感分析（比如每句话点评中的精选评论，点击评论标签完成评论召回）	用联合训练（考虑apect之间的关系，中间加attetion学权重）减轻不同aspect分布不均匀
	query意图识别	准确的query流量划分	
	推荐理由场景划分	和query、用户相关的个性化推荐理由的推送或召回理由	
句间关系	query改写语义一致性检测	对同意义，一词多义的query改写后是否和原意义一致（+人工审核）	
	query成分分析（NER序列标注）	对用户的随意搜索（关键词堆砌）做核心成分分析，做个二召	直接softmax看似整体准确率高，但是容易出现标签跳变，整个识别不完整，可用CRF规避

PS：
（1）完成分类后，还需要由业务方设计展示策略，从而完成类似“低星好评，高星差评”的问题的解决
（2）BERT使用性价比综合考虑显卡资源和FineTune后作为Baseline可以节省足够时间

- 遇到的业务难题
  - 业务方提供数据–train模型–交付，业务周期长人手少 – 搭建平台解决
  - 希望寻找模型效果好的原因 – 搭建平台解决
- 其他材料

PPT：E:\PM之路\日常积累\讲座资料\2020.04.01–BERT在美团搜索业务中的应用.pdf
录音：E:\PM之路\日常积累\讲座资料\2020.04.01–BERT在美团搜索业务中的应用.wav
视频回放：https://www.bilibili.com/video/BV1vC4y147px?from=search&seid=19738871039698078 97

图像处理

直播 2020.04.15 – 行为动作定位的算法流程介绍与分享

大纲：视界编码，proposal生成，proposal评价，模型ensemble

- 业务内容（主要针对视频中人的行为）

一级类别	二级类别	应用	tricks
目标识别 目标检测	视界编码–video representation – 模型 C3D Convolution Two–stream		– 模型增加降维层（卷积）和分类层（MLP）会work – encoding可以在attention、local global上挖掘trick，现在做的最好用的还是光流+做two stream（不太重视速度的话）
	proposal生成 – 模型 spliding windows archor based(SSAD) boundary based(BSN, SSN) combinations(BMN[confidence scroes+archor], DBG[proposal–level probabilities+anchor boxes], CTAP[spliding windows+TAG–temporal actioness grouping+complementary filter]) relative–aware pyramid network(RAM[我们的工作，包括Temporal contex distilling, Mutli–granularity proposal generation, Archor baxes selection])	短视频、游戏平台 – 对用户感兴趣的片段裁剪供用户预览 – QuickView – 视频部分片段和文字联动（搜索，评论，广告）– Focus on a section	– 用snippet windows采集视频clip，16帧采集效果比较好 – Archor baxes太少可能会对小proposal漏检 – DBG测试action detection任务表现不是很好 – 其中BSN（？不确定）处理速度相对较慢，其他模型一般速度可以达到200ms/次，可以用多clip等方式加速code
	proposal评价 – 模型 confidence score regression(BSN[extend boundary regions, BMN\DBG[pre–defined simple mask]) offset and action regression Reranking&Boundary Adjustment[我们的工作，包括Proposal Evaluation Module,		– contex的加入很有用（比如CTAP模型） – 和其他模型，比如BSN进行Boundary微调会work

	Boundary Adjustment Scheme]		
	模型ensemble – 根据特征、根据模型 confidence score regression complementary filters Reranking & Boundary Refinement		– 模型差异比较大，ensemble 效果相对较好

- 遇到的业务难题
  - **高质量体验** – ?
  - 准确分类 – combination和ensemble
- 其他材料
  PPT：E:\PM之路\日常积累\讲座资料\2020.04.15–云从数据–行为动作定位的算法流程介绍与分享.pdf

视频回放：https://b23.tv/BV1VA411b7G5

公众号文章 2020.04.08 – 整个世界都是你的绿幕：这个视频抠图换背景的方法着实真假难辨

- 来源：CVPR 2020论文
- 内容

评价指标	实验数据集	对比的深度蒙版算法	数据集上对比结果	已知BUG/限制条件	潜在应用场景	是否有教程
MSE	Adobe Dataset	BM: Bayesian Matting CAM: Context–Aware Matting IM: Index Matting LFM: Late Fusion Matting	Our: 1.72(Additional inputs: B) 1.73(Additional inputs: B') BM: 2.53(Additional inputs: Trimap–10, B) 2.86(Additional inputs: Trimap–20, B) 4.02(Additional inputs: Trimap–20, B') CAM: 3.67(Additional inputs: Trimap–10) 4.72(Additional inputs: Trimap–20) IM: 1.92(Additional inputs: Trimap–10) 2.36(Additional inputs: Trimap–20)	**限制条件：** （1）除了原始图像/视频之外，研究者还要求拍摄者多拍一张不带人物的背景图 **BUG：** （1）尤其是在摄像机拍摄的场景下，但手持拍摄的视频中，由于非平面背景导致的视差，还是会出现一些蒙版错误	云旅游 视频会议	是
SAD			Our:			

			0.97(Additional inputs: B) 0.99(Additional inputs: B') BM: 1.33(Additional inputs: Trimap–10, B) 1.13(Additional inputs: Trimap–20, B) 2.26(Additional inputs: Trimap–20, B') CAM: 4.50(Additional inputs: Trimap–10) 4.49(Additional inputs: Trimap–20) IM: 1.61(Additional inputs: Trimap–10) 1.10(Additional inputs: Trimap–20)			
主观指标–相对提升	10 个真实世界视频（手持相机）		BM: 52.9%muchbetter 41.4%better 5.7%similar 0%worse 0%much worse CAM: 30.8%muchbetter 42.5%better 22.5%similar 4.2%worse 0%much worse AM: 26.7%muchbetter 55.0%better 15.0%similar 2.5%worse 0.8%much worse LFM: 72%muchbetter 20%better 4%similar 3%worse 1%much worse			
	10 个真实世界视频（固定相机）		BM: 61%muchbetter 31%better 3%similar 4%worse 1%much worse			

			CAM: 43.3%muchbetter 37.5&#xfffd;tter 5%similar 4.2%worse 10%much worse AM: 33.3%muchbetter 47.5&#xfffd;tter 5.9%similar 7.5%worse 5.8%much worse LFM: 65.7%muchbetter 27.1&#xfffd;tter 4.3%similar 0%worse 2.9%much worse			

- 其他材料

Github：https://github.com/senguptaumd/Background-Matting