研讨纪要

# 研读论文

## 杨雨嫣

**SalGAN: visual saliency prediction with adversarial networks**

这篇文章提出了一种数据驱动的基于度量的显著性预测方法SalGAN，这个网络前部分使用深度神经网络根据输入图像的原始像素预测显著图；后半部分通过分类器网路区分显著图和真实结果，预测结果与真实结果无法区分时SalGAN有望生成与基本事实相似的显著性图。贡献点在于解决了选择最佳损失函数和不同显著性度量指标难对于显著图的含义不同的问题。

## 倪欣

阅读论文《Phishing Attacks on Modern Android》，该论文阐述了针对安卓系统的钓鱼攻击。近年来，通过移动设备访问社交网站等在线服务的人越来越多，流量也越来越大，本文研究发现安卓系统的移动密码管理器和Instant Apps存在漏洞，可能会被钓鱼攻击，对攻击场景进行了展示，并提出一种新的API来避免常见攻击。

主要工作：

1.对移动密码管理器及其依赖的三项核心技术进行了安全性分析，发现了设计和分析方面存在的漏洞。

2.展示了如何利用Instant Apps获得完全的UI控制，以及降低网络钓鱼攻击的门槛。

3.提出了端到端的网络钓鱼攻击。

4.提出了一个安全的API。

## 章雪琦

论文方向为机器学习与去匿名化相结合，去匿名化是指匿名数据和其他数据来源相互对照来重新识别匿名的数据来源。区别一个数据来源和另一个数据来源的任意信息都能用于去匿名化。

1. **Privacy-Enhanced Machine Learning with Functional Encryption**

这篇文章介绍了第一个用于功能加密的成熟的开源密码库。 它还介绍了如何使用功能加密来构建有效的增强隐私的机器学习模型，并提供了可应用于加密数据的三种预测服务的实现。 最后，本文讨论了使用同态加密来构建增强隐私的机器学习模型的另一种方法（使用同态）的优缺点。

功能函数与同态加密类似，它允许对加密数据进行计算。更准确地说，解密密钥的所有者可以学习加密数据的功能。通过控制人们可以从加密数据中获得的信息，可以将其用于各种分析或机器学习模型。

有一种方案为二次多元多项式的方案，该方案能够计算加密矢量上的二次多项式。 这可应用在机器学习模型，甚至是神经网络的基本版本。我们提供了一个机器学习项目（论文给出了代码），以演示如何在MNIST数据集上构建准确的神经网络分类器，以及如何使用FE在加密的数据集上应用分类器。这意味着，持有实体的FE密钥分类器可以对加密图像进行分类。

这篇文章虽然题目为使用功能加密的隐私增强型机器学习，但是全文8节只在第七节阐述了部分，前面大篇幅介绍实现了功能加密库、对其库的性能进行评估，比较了几种功能加密模式，并举例了两种将功能加密中的内积加密来实现对隐私数据的预测可行性。

## 4. 张敏

**(1) Deep Models Under the GAN: Information Leakage from Collaborative Deep Learning**

本文解决的主要问题:

使用GANs设计了一种针对协同深度学习的强大攻击。攻击的结果是，任何作为内部人员的用户都可以从受害者的设备中推断出敏感信息。攻击者只需运行协同学习算法并重建存储在受害者设备上的敏感信息。攻击者还能够影响学习过程并欺骗受害者发布更详细的信息。这种攻击在不影响服务操作者的情况下工作，甚至当模型参数通过差异隐私被模糊时也是如此.

创新点：

a. 设计了一个新的攻击 —— 基于GANs的分布式深度学习。GANs通常用于隐式密度估计，据我们所知，这是第一个恶意使用GANs的应用程序；

b. 本文的攻击比当前的信息提取机制更通用、更有效。特别是，本文的方法可以用来对付卷积神经网络（CNN），这是众所周知的模型反演攻击的困难；

c. 本文引入了协同学习中的欺骗概念，即对手欺骗受害者，使其在敏感数据上发布更准确的信息；

d. 本文设计的攻击在通过差异隐私模糊参数时也是有效的。这不是针对差异隐私的攻击，而是针对协同深度学习的攻击。并且在实践中证明的相关论文应用的差异私人训练（样例/记录级差异隐私）在本文提出的隐私概念下的协同学习环境中是无效的。

## 5. 赵愉悦

在Graph embedding上泛读相关文献, 包含四篇:

1. 【Word2Vec】Efficient Estimation of Word Representations in Vector Space，Google 2013；
2. 【DeepWalk】DeepWalk- Online Learning of Social Representations；
3. 【LINE】LINE - Large-scale Information Network Embedding，微软 2015；
4. 【Node2vec】Node2vec - Scalable Feature Learning for Networks，斯坦福 2016，

其中:

embedding: 表示一个映射关系. 通常的One-hot能够得到千万级别维度的稀疏矩阵(距离一样), 但是通过embedding, 可缩小到自定义维度大小, 并且直接计算向量相似度.

然后:

一般的word2vec, item2vec都是通过序列(swquence式), 学习、**单词**或者**商品组合**的真实含义.

Graph Embedding**图结构**, 包括节点图, 关系图, 可以生成隐含的信息.

DeepWalker -

1. **Random Walk** 负责对图进行采样: 随机选择初始点, 产生随机的节点序列路径.
2. **Skip-gram** 从关系（也就是采样的序列）中训练节点的 **Embedding** 向量。

网络节点的同质性（homophily）和结构性（structural equivalence）, 体现在embedding vector上就是向量之间的距离.

LINE:

1. **First-order proximity（1 阶相似度）**：若节点之间存在直连边，则边的权重即为两个顶点的相似度，若不存在直连边，则 1 阶相似度为0。
2. **Second-order proximity（2 阶相似度）：** 描述相邻节点相似度关系。 即: 令p表示顶点 u 与所有其他顶点间的1阶相似度，则 u 与 v 的2阶相似度可以通过 p\_u 和 p\_v 的相似度表示。若u与v之间不存在相同的邻居顶点，则2阶相似度为0。

*有效避免DeepWalk导致node embedding 聚集的情况.*

Node2Vec:

1. **BFS:** 首先走相邻节点 -> 相同step节点表示`结构`上相似性；
2. **DFS:** 首先走一根肠子 -> 相同step节点表示`距离`上同质性。

**t -> V, 在节点V:**

* 参数 **p** 为 “返回参数”（return parameter），**p**越大[>max(q,1)], 返回原节点的概率越小, 倾向于走DFS.
* 参数 **q** 为 “进出参数”（in-out parameter），**q**越大, 和**t**的远离节点权重越小, 倾向于走靠近t的相邻节点, 即BFS

*通过修改参数 p 和 q，就能控制 Random Walk 采样过程，是更倾向 DFS 还是 BFS。*

可用于发掘**不同特征**, 保留不同特征信息.

## 6. 杨英光

**(1) Manipulating Machine Learning:**

**Poisoning Attacks and Countermeasures for Regression Learning**

1) 当前论文引出的问题

当前越来越多的应用使用机器学习进行决策，易受数据影响的产生不同模型（脆弱性一），通过生成的数据更新模型（脆弱性二），其具体样例有：黑客通过操纵数据集或插入毒化数据（安全行业提交毒化指标，医疗行业插入毒化病例数据）。可以通过数据推断隐私（与该研究无关）

2) 提出的攻击方案

在训练过程中，不断更新毒化样本，来使得线性回归的损失最大化，从而改变线性回归模型的参数。

使用与训练集同分布的数据，并从训练集中计算均值和协方差，在根据计算出的均值和协方差数据，选择预测变量Y的边界值，来使损失最大化，从而改变参数。

3)提出的防御方案

TRIM不同于其他方法仅仅删除训练集中的异常值。TRIM迭代的预测回归参数，同时每次迭代中训练有着最低残差的子数据集，本质上使用了一个修剪函数计算不同子集在每次迭代的残差。

希望能够区分所有的毒化样本，用剩下的N个样本训练回归模型，试图寻找一组相对于回归模型有最低残差的训练点（观察值与估计值的差）

## 7. 唐玮

**TIRESIAS: Predicting Security EventsThrough Deep Learning**

论文的核心目标是通过RNN，构建一个可以预测攻击者在实施攻击时可能采用的具体攻击步骤（主要是当前有限步骤之后的下一个最优可能采取的步骤）的模型。

传统的检测方法只关注于二元结果，即是否会被攻击，不能指出攻击者下一步的目的与行为，同时传统的检测方法通常需要大量的标记数据来构建模型。

构建这样一个模型的主要难点在于：

* 即使是同样的攻击，其具体的执行模式并不相同，即事件发生的(执行的)顺序并不相同，也即事件 ei 与 ej 并不会有固定的前后关系;
* 终端会观测到与此次攻击不相关的其他安全事件，即攻击事件 e3, e4 之间会观测到与攻击事件无关的其他安全事件 e22 等;
* 终端可能会在同一时间观察到多个攻击活动，也就是说一个事件序列里面包含了多个攻击行为的序列。

通过 stacking RNNs 在每个 time-step 来创建一个多层的前馈神经网络被证实是切实高效好用的，但其对新数据的通用性不够(比如 stacking mechanism 为当前训练数据选择和调整时需要进行严格的评估，但这可能不能很好的适应新数据)，所以与其堆叠 RNNs，不如在 RNN cell 里构建一个更复杂的 memory structeres，在保证单层 RNN 网络训练效效率的同时保持 temporal memories。具体的做法是基于 Rocki 的 recurrent memory array，通过修改 LSTM 架构实现。

# 本周计划

1. **杨雨嫣**

* 想看一下这周论文的代码，有可能的话可以复现一下
* 总结这两周的论文做出ppt分享

1. **倪欣**

* 选择一篇论文精读，完成课堂作业。
* 根据课程安排继续学习机器学习。

1. **章雪琦**

* 读一篇A类论文；
* 完成对机器学习课程的学习。

1. **张敏**

* 精读一篇论文；
* 继续学习深度学习的课程。

1. **赵愉悦**

* 继续CS229的学习, 搭配阅读深度学习—花书.
* 精读一篇对抗样本的论文。
* 泛读一篇深度学习在网络信道中应用的论文

1. **杨英光**

* 完成课程安排的有关机器学习编程的作业；
* 继续学习深度学习相关知识
* 精读一篇论文。

1. **唐玮**

* 根据课程安排学习深度学习；
* 深入阅读一篇论文；
* 完成课程作业。

# 三、遇到的问题

* 没有服务器可能很难复现代码的训练过程
* 希望能够购买一台投影仪放置在老师办公室，这样在组会时可以方便地通过PPT来讲解论文。