COMPA: Detecting Compromised Accounts on Social Networks

简介

作者：Manuel Egele, Gianluca Stringhini, Christopher Kruegel and Giovanni Vigna. NDSS 2013.

动机

现有的识别方法无法区分虚假账户与恶意账户，依赖于消息中的URL特征，并且准确率一般。

目标

识别社交网络上被盗的恶意账户。

核心思想

对正常用户的常规活动进行建模，检测异常行为。具体思想为：找到一组相似的消息，并且这些用户的行为都与正常行为不符，这两个步骤可以打乱。

贡献

- 首个提出了检测受损账号的方法
- 使用了一组新颖的特征建模
- 准确率高

方法

消息特征建模

- 活跃小时数
- 消息源（即发送的应用程序）
- 语言
- 话题
- URL链接
- 用户交互（@someone）
- 地理位置信息

总消息数量小于10的用户将不会被建模。数据以键值对组的形式存储。

异常信息识别

value大于阈值的行为被认为是正常的，计0，小于阈值的行为将其比重f作为得分 $1-f$，不同模型之间的权重通过序列最小优化算法（SMO）获得。

相似消息分组

- 内容相似性，使用n-gram算法
- URL相似性

受损账号识别

- 可疑用户组的确定

  当用户组内的可疑消息超过了th条时，会判断为组内所有用户都受到了威胁。

$$th = max(0.1, kn + d), where\ k = -0.005, d = 0.82$$

- 批量应用程序的区分

  需要排除部分批量发送消息应用程序的影响，计算每个应用程序发送消息的平均编辑距离比率，以0.35为分界线。

实验

Network & Similarity Measure	Twitter Text		Twitter URL		Facebook Text	
	Groups	Accounts	Groups	Accounts	Groups	Accounts
Total Number	374,920		14,548		48,586	
# Compromised	9,362	343,229	1,236	54,907	671	11,499
False Positives	4% (377)	3.6% (12,382)	5.8% (72)	3.8% (2,141)	3.3% (22)	3.6% (412)
# Bulk Applications	12,347		1,569		N/A	N/A
# Compromised Bulk Applications	1,647	178,557	251	8,254	N/A	N/A
False Positives	8.9% (146)	2.7% (4,854)	14.7% (37)	13.3% (1,101)	N/A	N/A
# Client Applications	362,573		12,979		N/A	N/A
# Compromised Client Applications	7,715	164.672	985	46,653	N/A	N/A
False Positives	3.0% (231)	4.6% (7,528)	3.5% (35)	2.2% (1,040)	N/A	N/A

Table 2. Evaluation Results for the Text (Twitter and Facebook) and URL (Twitter) Similarity measure

其他

这篇文章只用到了SMO（还是现成的weka）用于计算权重，后续的研究可以考虑跟进。