PriMa:一种有效的隐私保护机制的社交网络

prima访问规则是根据用户对其个人资料数据的隐私偏好、数据的敏感性以及向其他用户披露此类数据的风险而生成的。

基于 Nlp 的防止社交网络帖子隐私泄露的解决方案

开发了一个基于浏览器的插件。所提出的方法通过考虑引理在句子中的作用的机制来模拟句子的结构。

在信息发布时评估其敏感性的技术可以通过警告用户披露某一信息会带来风险来加强隐私保护。

通过利用用户经常用来泄露私人数据的循环自然语言模式，自动拦截社交网络帖子中传递的敏感信息。

文献中大量研究了机器学习拦截自然语言中的隐私泄露[6，17，22，22]。尽管机器学习在这种情况下被证明是有效的，但是如果(I)特征选择不当，其性能可能会下 降;(ii)训练集不适合测试集;以及(iii)没有使用正确的机器学习算法。

[4]中的研究人员列出了经常出现在推文中的隐私信息 类别。我们利用这种分类法，以便对社交网络帖子中披露的敏感信息进行分类。

确定了一组社交网络用户经常用来揭示敏感数 据的自然语言模式。对于每个识别出的模式，我们定义了NLP 试探法(旨在自动识别它)。

• RQ1:提出的方法在多大程度上能有效识别社交网络帖子中的敏感数据?

• RQ2:在检测社交网络帖子中的敏感数据方面，所提出的方法是否比机器学习方法更有效?

方法在机器学习方面展示了 额外的优势:它更精确，并且不依赖于特定的训练集或

特定特征的选择。 由于该方法的唯一限制是启发集的不完全性。

TABOO:使用递归神经网络检测非结构化敏感信息

通过学习文本文档的语义和句法结构来检测 文本文档中复杂的敏感信息。我们的方法基于自然语言处理方法 进行释义检测，并使用递归神经网络为句子结构的语义成分分配 敏感度分数。

一个实体(如公司名称)在一个上下文中可能是敏感的，而在另 一个上下文中可能是不敏感的，或者敏感信息甚至不能仅 由单个名称或术语捕获。

我们观察到，敏感信息可能嵌入在文本的语义中，即使文 本只包含“非敏感”单词。

我们的禁忌系统提取句子的组成子结构来学习敏感度 分数。它建立在成功的自然语言处理方法的基础上，包括 释义、情感分析和图像句子排序[6]，[7]。

在禁忌中，递归神经网络(RNN)以结构化的方式处理句 子。

禁忌输入文档或者通过扫描输入目录加载，或者通过 禁忌图形用户界面加载，用户可以直接加载或编写文档。 然后对每个文档进行 1)句子拆分，2)语法分析和 3) RNN 分 析。1)使用自然语言工具包 NLTK [8]，将文档拆分成句子进 行进一步处理。2)使用斯坦福 NLP 解析器以反映句子结构 的解析树的形式提取句子的子结构[9]。3)使用由伊尔索和 卡迪开发的深度递归神经网络在语法树上训练模型(或使用 现有模型进行检测[10]。

一种自动评估和感知隐私披露的工具

[7]中，作者 提出了一个自动系统，将推特上的帖子分成几个固定的类别。 [1]中的工作分析了与使用脸书社交网络相关的多个隐私问题， 但并不能帮助用户防止信息泄露。其他作品，如[6]，试图提 高隐私意识，每当一些敏感信息可能在网络导航过程中被披 露时，都会提示用户，但其行动范围仅限于基本的、结构化 的数据，如电子邮件、姓名、社会地位。

我们测试了多种机器学习算法，目标是能够评估组成社交帖子 的文本是否披露了敏感信息。表现最好的监督学习方法是 SVM，其结果如表 1 所示，其次是规则归纳，见表 2，以及朴 素贝叶斯，见表 3。

Private Detect：在大型社交网络中检测隐私信息和集体隐私行为

基于学习的方法“隐私侦探”是我们的主要贡献，原因有三:

1. 隐私侦探检测广泛的隐私类别。之前的工作集中在某 些类型的隐私上，如位置隐私、医疗隐私或受影响下 的写作。
2. 隐私检测采用一种基于学习的方法，而以前的方法侧 重于基于关键词和正则表达式的检测。
3. 隐私受到社会的影响，用户和她的朋友的隐私得分之 间的正相关就证明了这一点。

刘等。[19]根据隐私信息的敏感性和可见性，提出一个计算 在线社交网络中用户隐私得分的框架。本研究中的隐私得分表 明了用户因参与网络而带来的潜在风险。

Chow 等人。[8]设计一个文本修订助手，检测文本中的敏感 信息，并给出净化句子的建议。他们的方法包括在互联网上查 询检测和推荐。

在推特上已经有许多关于主题建模[18]、命名实体识别[25]、 情感分析[6]以及规范微文本[30]的研究，尽管没有特别关注推文。

通过适当的预处理，朴素贝叶斯可以优 于支持向量机(SVM)。在我们的实验中，增强的朴素贝叶斯明 显优于序列最小优化[24]，一种 SVM。

文本预处理---特征抽出---特征归一化---主题比率---

用 Twitter 单词聚类中的聚类关键词来代替。我 们使用了推特自然语言处理项目[22]中的 1000 个分层推特 单词聚类，这些单词是由布朗从 5600 万条超过 21.7 万个单 词的英语推特中聚类[7]而成的。我们手动检查了聚类，并 选择了一个描述聚类中单词的关键词。如果时间线中的任何 单词出现在聚类中，我们用聚类关键字替换该单词。

在将单词转换为聚类关键词后，我们删除了非 ASCII 字符， 以减少非英语语言和象形字符。用户句柄(例如@johnsmith) 被替换为单词 he，URL 被替换为关键字 URL，拼写错误根据 英语词典进行了更正。这些文本预处理步骤如图 3 所示。

表 3 显示了属于私人或中立类别的 6 个主题。我们从每个主 题中提取了前 20 个术语，以更好地评估主题的内容。