技术报告

社交软件网络中服务商对于流量检测、分类、管控的需求越来越大，于是对于社交软件的指纹识别要求也更高了。所以根据微信用户以及Telegram用户会做出的基本行为进行了用户行为数据的抓取。并且通过流量分段操作进行数据分流，再利用随机森林对样本进行训练预测，准确率高达98%，

具体来说该方法包含数据预处理模块，数据分割模块，训练测试模块三部分。

·数据预处理模块：通过编写JS脚本文件，对目前主流的两个社交软件：微信和Telegram软件进行了17种用户行为数据的采集，并且通过载入日志文件来对数据进行分流。

·数据分割模块：将数据预处理模块的输出矩阵进行分割处理，方便后面的训练测试。

·训练测试模块：采用的是默认参数的随机森林模型，相较于浅层的神经网络模型，随机森林能提供更好的准确率和效率。最后实现了微信的发送文本、图片、朋友圈、视频通话、红包、位置、视频文件、语音、语音通话、读公众号、转账和打开小程序，Telegram的发送文本、图片、语音、视频和发送文件这17种用户行为的分类预测。