Phương pháp tiếp cận máy học để ngăn chặn Cuộc gọi độc hại qua mạng điện thoại

Huichen Li1 Xiaojun Xu1 Chang Liu2 Teng Ren3 Kun Wu3 Xuezhi Cao1 Weinan Zhang1 Yong Yu1 Dawn Song2 1Shanghai Jiao Tong University 2University of California, Berkeley 3TouchPal Inc.

Tóm tắt — Các cuộc gọi độc hại, tức là các cuộc gọi lừa đảo và lừa đảo qua điện thoại, đã là một vấn đề thách thức lâu dài gây ra tổn thất tài chính hàng tỷ đô la hàng năm trên toàn thế giới. Công trình này trình bày giải pháp dựa trên máy học đầu tiên mà không dựa trên bất kỳ giả định cụ thể nào về cơ sở hạ tầng mạng điện thoại bên dưới.

Thách thức chính của vấn đề kéo dài hàng thập kỷ này là không rõ làm thế nào để xây dựng các tính năng hiệu quả mà không có quyền truy cập vào cơ sở hạ tầng của mạng điện thoại. Chúng tôi giải quyết vấn đề này bằng cách kết hợp một số đổi mới. Đầu tiên, chúng tôi phát triển giao diện người dùng TouchPal trên Ứng dụng di động để cho phép người dùng gắn thẻ các cuộc gọi độc hại. Điều này cho phép chúng tôi duy trì cơ sở dữ liệu nhật ký cuộc gọi quy mô lớn. Sau đó, chúng tôi thực hiện một nghiên cứu đo lường trong ba tháng nhật ký cuộc gọi, bao gồm 9 tỷ bản ghi. Chúng tôi thiết kế 29 tính năng dựa trên kết quả để các thuật toán máy học có thể được sử dụng để dự đoán các cuộc gọi độc hại. Chúng tôi đánh giá rộng rãi các phương pháp học máy hiện đại khác nhau bằng cách sử dụng các tính năng được đề xuất và kết quả cho thấy rằng phương pháp tốt nhất có thể giảm tới 90% các cuộc gọi độc hại không bị chặn trong khi vẫn duy trì độ chính xác trên 99,99% đối với lưu lượng cuộc gọi lành tính. Kết quả cũng cho thấy các mô hình hiệu quả để triển khai mà không phát sinh chi phí độ trễ đáng kể. Chúng tôi cũng tiến hành phân tích cắt bỏ, cho thấy rằng việc sử dụng 10 trong số 29 tính năng có thể đạt được hiệu suất tương đương với việc sử dụng tất cả các tính năng.

I. GIỚI THIỆU

Lừa đảo và lừa đảo qua mạng điện thoại đã gây ra thiệt hại tài chính hàng năm trị giá hàng tỷ đô la trên toàn thế giới [1] - [3]. Chúng tôi gọi chúng là những cuộc gọi độc hại.

Thật không may, vẫn chưa có giải pháp đơn giản và hiệu quả nào để ngăn chặn chúng [33]. Mặc dù các quốc gia, chẳng hạn như Hoa Kỳ , đã thành lập Cơ quan đăng ký quốc gia không gọi để giảm thiểu vấn đề này, nhưng vấn đề này nghiêm trọng hơn ở các quốc gia như Trung Quốc, nơi không có luật này.

Một trong những thách thức chính là thiếu sự hình thành hiệu quả để phát hiện chính xác các cuộc gọi độc hại. Khác với spam email truyền thống, các cuộc gọi độc hại thường yêu cầu phản hồi tức thì trước khi nội dung trong cuộc gọi được nghe.

Do đó, chỉ có thông tin tiêu đề mới có thể được sử dụng để ngăn chặn các cuộc gọi độc hại khiến người nhận mất thời gian, tiền bạc và năng suất. Các kỹ thuật ngăn chặn cuộc gọi độc hại trước đây chủ yếu tập trung vào Spam qua Internet Telephony (SPIT) và dựa vào thông tin phía máy chủ về người gọi để dự đoán các cuộc gọi độc hại [7], [11], [19], [21], [24], [31], [34], [36], [40].

Tuy nhiên, thông tin như vậy thường không có sẵn cho người dùng cuối trên các mạng điện thoại truyền thống.

Trong công việc này, chúng tôi tập trung vào việc ngăn chặn cuộc gọi độc hại mà không dựa vào bất kỳ cơ sở hạ tầng mạng điện thoại cơ bản cụ thể nào. Vì vậy, thách thức đầu tiên là làm thế nào để thu thập thông tin hiệu quả. Đóng góp đầu tiên của công việc này là thu thập thông tin về những người gọi độc hại nhằm xây dựng một cơ chế ngăn chặn hiệu quả, sử dụng giao diện người dùng TouchPal. Ý tưởng cơ bản là triển khai TouchPal như một chức năng của Ứng dụng dành cho thiết bị di động có một số lượng lớn người dùng. Sau đó, TouchPal cho phép người dùng gắn nhãn một cuộc gọi đã kết thúc là độc hại hay không và thực hiện cơ chế ngăn chặn danh sách đen dựa trên danh tiếng đơn giản dựa trên việc gắn thẻ của người dùng.

TouchPal cũng kịp thời gợi ý người dùng gắn nhãn các cuộc gọi đáng ngờ. Thiết kế này cũng làm tăng khối lượng nhật ký cuộc gọi được gắn thẻ ngoài việc người dùng tự nguyện gắn nhãn thông qua các dịch vụ báo cáo cuộc gọi độc hại, chẳng hạn như 12321. Làm như vậy, chúng tôi có thể thu thập một tập dữ liệu nhật ký cuộc gọi lớn mà không cần dựa vào bất kỳ cơ sở hạ tầng mạng điện thoại cụ thể nào.

Mặc dù cách tiếp cận danh sách đen đơn giản là hiệu quả, nó phải quan sát đủ các bản ghi cuộc gọi từ một số độc hại trước khi TouchPal có thể đưa vào danh sách đen số đó. Câu hỏi tiếp theo của chúng tôi là: làm thế nào chúng tôi có thể xây dựng một cơ chế hiệu quả để phát hiện sớm một số cuộc gọi độc hại mà không cần trả lời quá nhiều cuộc gọi được gọi từ số đó? Chúng tôi tìm kiếm một giải pháp học máy.

Trở ngại chính là thiết kế một tập hợp các tính năng hiệu quả. Để đạt được mục tiêu này, chúng tôi dựa vào nhật ký dữ liệu được thu thập từ TouchPal để phân tích thông tin nào hiệu quả được sử dụng làm tính năng. Trên thực tế, trong vài năm qua, TouchPal đã đạt hơn 56 triệu người dùng hoạt động hàng ngày và theo dõi hàng tỷ bản ghi cuộc gọi hàng tháng. Ngày nay, TouchPal duy trì cơ sở dữ liệu nhật ký cuộc gọi lớn nhất ở Trung Quốc liên quan đến cả khối lượng ID cuộc gọi và khối lượng thẻ cuộc gọi.

Sử dụng tập dữ liệu này, chúng tôi thực hiện một nghiên cứu đo lường quy mô lớn để hiểu thông tin nào hữu ích hơn để loại bỏ các cuộc gọi độc hại từ những cuộc gọi lành tính. Lưu ý rằng đã có một số công trình trước đây cung cấp các nghiên cứu về đo lường [15], [22]. Tuy nhiên, những công việc này chủ yếu tập trung vào các bản ghi cuộc gọi độc hại và không cung cấp thông tin chi tiết về việc các bản ghi cuộc gọi độc hại khác với các bản ghi lành tính như thế nào. Ngoài ra, họ tập trung vào các bản ghi cuộc gọi của Hoa Kỳ và không rõ liệu các kết luận tương tự có áp dụng cho hệ sinh thái cuộc gọi độc hại của Trung Quốc hay không. MobiPot [9] cung cấp một nghiên cứu để khắc phục hai vấn đề này; tuy nhiên, nghiên cứu chỉ dựa trên ít hơn 700 bản ghi cuộc gọi và chúng tôi cho thấy rằng các quan sát từ [9] không đủ mạnh khi chúng tôi tăng kích thước mẫu lên 7 bậc độ lớn.

Nghiên cứu của chúng tôi khắc phục tất cả những vấn đề này và làm sáng tỏ thiết kế tính năng, đây là vấn đề cốt lõi của công trình này. Kết quả của chúng tôi cho thấy (1) số lượng cuộc gọi độc hại của tỉnh nhạy cảm hơn với Tổng sản phẩm quốc nội của tỉnh

(GDP) so với các cuộc gọi lành tính; (2) các cuộc gọi ác ý có nhiều khả năng xảy ra trong một ngày làm việc và trong giờ làm việc hơn các cuộc gọi lành tính; và (3) khối lượng các cuộc gọi đến và đi từ một số là dấu hiệu để phân biệt các cuộc gọi độc hại với các cuộc gọi lành tính. Theo hiểu biết tốt nhất của chúng tôi, chúng tôi là người đầu tiên trình bày những phát hiện này liên quan đến việc phân biệt các cuộc gọi độc hại với các cuộc gọi lành tính.

Lấy cảm hứng từ nghiên cứu đo lường của chúng tôi, chúng tôi thiết kế 29 tính năng cho vấn đề dự đoán cuộc gọi độc hại để không chỉ bao gồm thông tin tĩnh về cuộc gọi hiện tại mà còn cả thông tin mở rộng bằng cách kiểm tra các bản ghi lịch sử về người gọi của cuộc gọi đến và bằng cách tham chiếu chéo nhiều bản ghi.
Chúng tôi đánh giá toàn diện hiệu quả của các tính năng này bằng cách sử dụng một số mô hình hiện đại. Để đạt được mục tiêu này, chúng tôi sử dụng cả điểm AUC tiêu chuẩn và thiết kế một chỉ số mới, dự đoán đầu tiên trung bình (AFP). AFP được thiết kế để đánh giá số lượng trung bình các cuộc gọi độc hại cần được quan sát trước khi một phương pháp tiếp cận có thể dự đoán đó là một người gọi độc hại, mà không ảnh hưởng đến lưu lượng cuộc gọi lành tính. Đánh giá của chúng tôi cho thấy rằng bằng cách sử dụng các tính năng được đề xuất của chúng tôi, một mô hình rừng ngẫu nhiên có thể đạt được điểm AUC ít nhất là 0,99; hơn nữa, nó làm giảm tới 90% các cuộc gọi độc hại được quan sát cần thiết trung bình từ cách tiếp cận danh sách đen, đồng thời đảm bảo rằng hơn 99,99% các cuộc gọi lành tính sẽ không bị chặn. Nói cách khác, mô hình rừng ngẫu nhiên tốt nhất sử dụng 29 tính năng được đề xuất của chúng tôi có thể giảm 90% các cuộc gọi độc hại không bị chặn.

Ngoài ra, đánh giá cho thấy rằng mô hình mạng nơ-ron có thể đạt được hiệu suất chính xác tương tự như rừng ngẫu nhiên tốt nhất, nhưng phải chịu chi phí độ trễ thấp dưới 1ms.
Điều này cho thấy rằng các mô hình trong đánh giá của chúng tôi có thể được triển khai hiệu quả trên cơ sở hạ tầng hiện tại để đạt được cả độ chính xác và hiệu quả cao.

Chúng tôi tiếp tục tiến hành nghiên cứu cắt bỏ để hiểu mức độ hiệu quả của từng tính năng được đề xuất và đánh giá của chúng tôi cho thấy chỉ cần 10 tính năng để đạt độ chính xác cao thay vì toàn bộ 29 tính năng.

Chúng tôi tóm tắt những đóng góp của chúng tôi như sau.

1) Chúng tôi phát triển giao diện người dùng TouchPal để theo dõi các cuộc gọi độc hại và cuộc gọi lành tính. Sử dụng cách tiếp cận này, TouchPal đã duy trì cơ sở dữ liệu nhật ký cuộc gọi lớn nhất ở Trung Quốc liên quan đến khối lượng ID cuộc gọi, tổng khối lượng bản ghi cuộc gọi và khối lượng cuộc gọi độc hại; 2) Chúng tôi thực hiện một nghiên cứu đo lường trên các bản ghi cuộc gọi quy mô lớn mà không có thông tin nhạy cảm của người dùng để rút ra thông tin chi tiết nhằm thiết kế các tính năng hiệu quả cho phương pháp tiếp cận ngăn chặn cuộc gọi độc hại dựa trên máy học; 3) Chúng tôi đề xuất 29 tính năng và đánh giá toàn diện 6 cách tiếp cận máy học hiện đại nhất. Kết quả cho thấy mô hình rừng ngẫu nhiên tốt nhất có thể đạt được AUC

điểm ít nhất là 0,99 và giảm tới 90% cuộc gọi độc hại không bị chặn so với cách tiếp cận danh sách đen, trong khi ít nhất 99,99% lưu lượng truy cập lành tính sẽ không bị chặn;

4) Chúng tôi đánh giá hiệu suất thời gian chạy của mô hình và cho thấy rằng một số mô hình hoạt động hiệu quả có độ trễ nhỏ trên không. Do đó, cách tiếp cận được đề xuất có thể được thực hiện một cách hiệu quả trên cơ sở cấu trúc hiện tại; 5) Để hiểu thêm về hiệu quả của các tính năng được đề xuất, chúng tôi tiến hành phân tích cắt bỏ. Chúng tôi nhận thấy rằng một số tính năng hữu ích hơn những tính năng khác và trong trường hợp cực đoan, việc sử dụng 10 tính năng hữu ích nhất hàng đầu có thể đạt được hiệu suất tương đương với việc sử dụng tất cả 29 tính năng.

II. TỔNG QUÁT

Trong phần này, chúng tôi trình bày tổng quan về vấn đề ngăn chặn cuộc gọi độc hại và giải pháp TouchPal. Trước tiên, chúng tôi sẽ xem xét ngắn gọn trạng thái cuộc gọi độc hại ở Trung Quốc, sau đó xác định vấn đề bằng cách cung cấp các yêu cầu về ngăn chặn cuộc gọi độc hại. Sau đó, chúng tôi sẽ giới thiệu tổng quan về giải pháp TouchPal, với những điểm nổi bật trong quá trình phát triển kỹ thuật của chúng tôi trong bài báo này.

A. Cuộc gọi ác ý ở Trung Quốc

Tình trạng luật pháp ở Trung Quốc chống lại các cuộc gọi ác ý đã quá sớm. Các dịch vụ, chẳng hạn như Đăng ký Không gọi Quốc gia ở Hoa Kỳ , chưa khả dụng ở Trung Quốc. Kênh chính do chính phủ Trung Quốc cung cấp là 12321, một dịch vụ chuyên báo cáo các cuộc gọi và tin nhắn SMS độc hại. Tuy nhiên, 12321 chủ yếu dựa vào báo cáo tình nguyện của người dùng; Ngoài ra, vẫn chưa rõ thông tin này cuối cùng được sử dụng như thế nào.

Hai chính sách được thực thi bởi chính phủ Trung Quốc, có thể có hiệu lực đối với các cuộc gọi ác ý. Đầu tiên, các nhà cung cấp dịch vụ viễn thông ở Trung Quốc được yêu cầu đăng ký danh tính thực với mọi số điện thoại. Thứ hai, một số quay số khác ở tỉnh khác sẽ phải trả phí đường dài. Hai chính sách này có thể làm tăng chi phí cho một người gọi độc hại. Tuy nhiên, bắt đầu từ ngày 1 tháng 9 năm 2017, phí đường dài đã bị hủy bỏ cùng với phí chuyển vùng [4].

B. Ngăn chặn cuộc gọi độc hại

Chúng tôi xem xét vấn đề để ngăn chặn các cuộc gọi độc hại từ phía thiết bị di động. Có nghĩa là, một trình ngăn chặn cuộc gọi độc hại được thực hiện trên điện thoại di động để cung cấp dịch vụ phát hiện liệu cuộc gọi đến có phải là cuộc gọi độc hại hay không. Như chúng tôi sẽ giải thích trong Phần III, chúng tôi chủ yếu coi các cuộc gọi quấy rối và lừa đảo là các cuộc gọi độc hại. Chúng tôi có các yêu cầu sau.

Nếu không có quyền truy cập vào cơ sở hạ tầng mạng điện thoại bên dưới. Chúng tôi yêu cầu giải pháp phải được triển khai trên thiết bị của người dùng cuối; do đó, nó không có quyền truy cập vào nhiều thông tin về người gọi mà chỉ có sẵn từ các máy chủ của các nhà cung cấp dịch vụ viễn thông. Điều này loại bỏ hầu hết các đề xuất phòng chống SPIT hiện có. Tuy nhiên, chúng tôi nhấn mạnh rằng yêu cầu này không ngăn cản một giải pháp sử dụng máy chủ để thu thập và lưu trữ thông tin được báo cáo từ thiết bị di động.

- Trọng lượng nhẹ cho người sử dụng. Cơ chế ngăn chặn không được phát sinh nhiều thao tác bổ sung cho người dùng cuối. Tốt nhất, người dùng nên nhận một cuộc gọi lành tính hoặc quay số như bình thường, và chỉ cần thao tác khác với điện thoại khi cuộc gọi đến được dự đoán là cuộc gọi ác ý.

Tính hiệu quả. Giải pháp không được ngăn người dùng nhận các cuộc gọi lành tính. Phần lớn các cuộc gọi lành tính (tức là, ≥ 99,99%) sẽ đi qua trình dò cuộc gọi độc hại.

Phát hiện sớm. Tốt nhất, một trình ngăn chặn cuộc gọi độc hại hiệu quả nên bắt đầu chặn tất cả các cuộc gọi độc hại từ một số điện thoại khi có ít cuộc gọi độc hại được thực hiện nhất có thể.

Hiệu quả. Trình ngăn chặn cuộc gọi độc hại phải chịu độ trễ thấp trên mặt điện thoại để phát hiện cuộc gọi đến có phải là cuộc gọi độc hại hay không. Lý tưởng nhất là chi phí độ trễ phải là 10ms hoặc thấp hơn.

C. Tổng quan về giải pháp

Bây giờ chúng tôi trình bày tổng quan về giải pháp của chúng tôi để sử dụng học máy để ngăn chặn cuộc gọi độc hại. Chúng tôi đã xây dựng TouchPal (Phần III) như một chức năng bổ sung trên một Ứng dụng dành cho thiết bị di động có hàng trăm triệu người dùng. TouchPal cung cấp chức năng cho phép người dùng gắn nhãn một cuộc gọi điện thoại là độc hại và sử dụng cơ chế ngăn chặn cuộc gọi độc hại trong danh sách đen dựa trên danh tiếng.

Tuy nhiên, cách tiếp cận danh sách đen yêu cầu cùng một num ber phải được gắn nhãn nhiều lần trước khi nó có thể bị đánh dấu là người gọi độc hại và bị chặn. Để giảm thiểu vấn đề này, chúng tôi phát triển một phương pháp học máy để dự đoán liệu một số điện thoại có phải là một người gọi độc hại hay không trước khi thực hiện quá nhiều cuộc gọi độc hại.

Thách thức chính là làm thế nào để thiết kế các tính năng hiệu quả mà không cần khai thác nội dung cuộc gọi của người dùng. Để đạt được mục đích này, TouchPal giữ nhật ký cuộc gọi chứa mỗi bản ghi cuộc gọi về người dùng TouchPal. Trong bản ghi cuộc gọi, chỉ những thông tin ít nhạy cảm hơn như thời lượng cuộc gọi và thời gian cuộc gọi được lưu trữ.

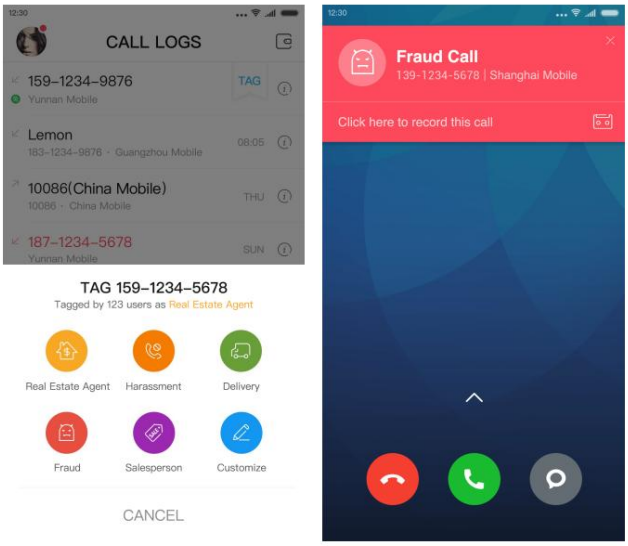
Mặc dù nhật ký cuộc gọi ẩn thông tin nhạy cảm, quy mô của nó cho phép chúng tôi thực hiện các quan sát quan trọng. Chúng tôi thực hiện một nghiên cứu đo lường quy mô lớn (Phần IV) bằng cách sử dụng dữ liệu trong khoảng thời gian ba tháng chứa 9 tỷ bản ghi cuộc gọi và kiểm tra thông tin nào hữu ích hơn để phân biệt những người gọi độc hại với những người gọi lành tính. Do đó, nghiên cứu của chúng tôi làm sáng tỏ việc thiết kế lựa chọn các đối tượng địa lý (Phần V).

Về mặt trực quan, bên cạnh tập hợp các tính năng cơ bản về cuộc gọi hiện tại, thông tin từ các bản ghi cuộc gọi lịch sử và thông tin từ nhiều bản ghi có thể hữu ích trong việc phát hiện các cuộc gọi độc hại. Chúng tôi đánh giá rộng rãi một số phương pháp học máy hiện đại (Phần VI), bao gồm rừng ngẫu nhiên, mạng nơ-ron, SVM và hồi quy logistic và chúng tôi nhận thấy rằng hầu hết các mô hình có thể đạt được hiệu suất cao. Đầu tiên, điểm AUC của hầu hết các mô hình có thể đạt 0,99 hoặc cao hơn.

Quan trọng hơn, chúng tôi thực thi độ chính xác của các cuộc gọi lành tính ít nhất là 0,99 và đánh giá số lượng cuộc gọi độc hại cần được quan sát trước khi số có thể được phát hiện. Đánh giá của chúng tôi cho thấy rằng một khu rừng ngẫu nhiên hoặc một mạng nơ-ron chỉ cần quan sát trung bình 2,5 cuộc gọi để phát hiện một người gọi độc hại; do đó, chúng tôi giảm tới 90% các cuộc gọi độc hại không bị chặn bằng cách sử dụng phương pháp danh sách đen hiện tại trong TouchPal.

III. TouchPal ĐỂ PHÒNG NGỪA CUỘC GỌI MALICIOUS

Trong phần này, chúng tôi giải thích đường dẫn của TouchPal để giúp người dùng ngăn chặn các cuộc gọi độc hại. TouchPal sử dụng



(a) Nhắc các thẻ gắn nhãn (b) Nhắc từ chối cuộc gọi lừa đảo

Hình 1: Giao diện người dùng TouchPal trên Android

một cách tiếp cận danh sách đen dựa trên danh tiếng để ngăn chặn các cuộc gọi độc hại. TouchPal cho phép người dùng gắn nhãn các cuộc gọi độc hại. Dựa trên những thông tin đó, TouchPal sẽ đánh dấu một số điện thoại là độc hại khi TouchPal tự tin làm như vậy. Chúng tôi giải thích chi tiết trong phần sau.

Bộ sưu tập thông tin. Giao diện người dùng cho TouchPal để cho phép người dùng gắn nhãn những người gọi độc hại được cung cấp trong Hình 1a. Khi cuộc gọi kết thúc, TouchPal sử dụng một số phương pháp phỏng đoán đơn giản để phát hiện các cuộc gọi đáng ngờ và nhắc người dùng bằng giao diện gắn nhãn. Làm như vậy có thể làm tăng cơ hội người dùng có thể gắn nhãn cuộc gọi, vì hầu hết người dùng không muốn chủ động gắn nhãn cuộc gọi. Heuristics được thiết kế để lời nhắc chỉ hiển thị cho người dùng khi cuộc gọi có nhiều khả năng được gắn nhãn, để giảm bớt gánh nặng cho người dùng TouchPal.

Đặc biệt, nếu một số không bao giờ được gắn thẻ (gần như chắc chắn là lành tính) hoặc đã bị đưa vào danh sách đen (gần như chắc chắn là độc hại), TouchPal sẽ không nhắc người dùng của nó; chỉ khi TouchPal không chắc chắn về một số, nó sẽ nhắc người dùng gắn thẻ. Ngay cả khi lời nhắc không hiển thị, TouchPal vẫn cung cấp một nút trong lịch sử cuộc gọi để gắn nhãn cuộc gọi độc hại.

Người dùng TouchPal có thể chọn một trong năm thẻ được xác định trước để gắn nhãn cuộc gọi. Thẻ thứ sáu cho phép người dùng tùy chỉnh các thẻ của họ. Năm thẻ tích hợp là Đại lý bất động sản, Quấy rối (dành cho thư rác), Giao hàng, Gian lận (dành cho lừa đảo) và Nhân viên bán hàng. Các danh mục này được tạo dựa trên cuộc khảo sát nội bộ về loại cuộc gọi mà người dùng TouchPal chủ yếu muốn chặn. TouchPal cũng cung cấp thông tin về thẻ được gắn nhãn thường xuyên nhất và tần suất của nó. Trong phân tích của mình, chúng tôi coi Quấy rối và Gian lận là thẻ cuộc gọi độc hại và những thẻ khác là những thẻ lành tính.

Danh sách đen dựa trên danh tiếng đơn giản. TouchPal sử dụng một chính sách tinh vi và thận trọng để gắn thẻ càng nhiều số điện thoại càng tốt, đồng thời giảm thiểu số lượng sai

Id người	Giải trình
dùng trường	ID người dùng TouchPal
Loại cuộc gọi	Một giá trị nhị phân cho biết liệu đây có phải là một hoặc một cuộc gọi đi
điện thoại khác	Số điện thoại ẩn danh trong cuộc gọi này khác hơn người dùng TouchPal
điện thoại khác_md5	Mã hóa duy nhất cho mỗi số điện thoại
ngày cuộc gọi	Dấu thời gian (tính bằng giây) khi bắt đầu cuộc gọi
thời lượng cuộc gọi	Số giây cuộc gọi kéo dài
gọi liên hệ	Một giá trị nhị phân cho biết nếu số kia là trong liên hệ của người dùng TouchPal
thẻ cuộc gọi	Thẻ của cuộc gọi

BẢNG I: Cấu trúc của các bản ghi nhật ký dữ liệu.

	Tháng 10	Tháng 11	Tháng 12	Tổng
Hồ sơ cuộc gọi	3.043	2.959	3.001	9.002
Hồ sơ cuộc gọi lành tính	3.017	2.933	2.979	8.929
Bản ghi cuộc gọi độc hại	26	25	22	73
Người gọi riêng biệt	256	248	248	447
Bé con riêng biệt	299	288	287	519
Người dùng TouchPal riêng biệt	24	24	24	35
Số cuộc gọi độc hại riêng biệt	0,6	0,5	0,5	0,8
Các số khác biệt	348	338	335	583

BẢNG II: Thống kê nhật ký dữ liệu (triệu) từ tháng 10 đến Tháng 12 năm 2016

các thẻ. Nhiều nguồn thông tin được sử dụng để gắn thẻ điện thoại con số. Một trong những nguồn chính là thẻ được cung cấp từ người dùng. Tuy nhiên, rất phổ biến là người dùng có thể gắn nhãn sai một số số điện thoại. TouchPal áp đặt một ngưỡng, có thể khác nhau từ 30 đến 100, để tự tin về thẻ của số điện thoại. Ví dụ: 10086, số dịch vụ của China Mobile, thường bị người dùng gắn nhãn là Quấy rối, và ngưỡng của nó do đó được đặt là rất cao. Thông tin khác nguồn cũng được sử dụng để xác nhận thẻ. Ví dụ, thực cơ quan bất động sản thường cung cấp số của họ trực tuyến, và TouchPal thu thập dữ liệu các trang web cho những con số đó để xác nhận tag. Lưu ý rằng ngưỡng được sử dụng trong TouchPal không phải là ngưỡng tĩnh, và có thể thay đổi từ số này sang số khác.

Lưu ý rằng danh tiếng từ cách tiếp cận danh sách đen này cũng phục vụ sự thật nền tảng để xây dựng máy học của chúng tôi các mô hình. Vì TouchPal chỉ cho phép người dùng gắn thẻ độc hại các cuộc gọi thay vì những cuộc gọi lành tính, rất khó để người dùng độc hại không dán nhãn miễn là có đủ người dùng lành tính gắn nhãn các số cuộc gọi độc hại là độc hại.

Ngăn chặn cuộc gọi độc hại. TouchPal cho phép người dùng cấu hình hành vi mặc định khi số điện thoại gọi đến cuộc gọi được đánh dấu bằng một thẻ. Ví dụ: người dùng có thể chọn kết thúc một cuộc gọi độc hại trực tiếp mà không có bất kỳ thông báo nào, hoặc chọn để nhắc người dùng. Hình 1b minh họa giao diện khi số điện thoại của một cuộc gọi đến được đánh dấu là Gian lận. Lưu ý rằng mặc dù TouchPal cung cấp chức năng ghi lại nội dung cuộc gọi, người dùng phải bật thủ công cho mỗi cuộc gọi; ngoài ra, bản ghi chỉ có sẵn trên thiết bị của người dùng và không bao giờ được tải lên máy chủ.

Việc ngăn chặn trong triển khai hiện tại là hoàn toàn dựa trên trên thẻ số điện thoại, có thể dễ dàng bị phá vỡ bằng cách sử dụng các kỹ thuật như giả mạo ID người gọi. Chúng tôi xem xét vấn đề này như một hướng quan trọng trong tương lai.

Các chức năng khác. TouchPal cũng cung cấp các chức năng khác như ngăn chặn tin nhắn SMS. Trong công việc này, chúng tôi tập trung về vấn đề ngăn chặn cuộc gọi độc hại.

IV. HIỂU CÁC CUỘC GỌI ĐÁNG YÊU Ở TRUNG QUỐC

Trong phần này, chúng tôi điều tra nhật ký cuộc gọi để có được thông tin chi tiết về hành vi của người gọi ác ý và làm sáng tỏ cách thiết kế thuật toán phát hiện cuộc gọi độc hại dựa trên máy học. Sau đây, trước tiên chúng tôi trình bày cấu trúc của lệnh gọi nhật ký và một số thống kê cơ bản. Sau đó, chúng tôi nghiên cứu sự phân phối của người dùng TouchPal, cuộc gọi độc hại và các số khác một số chiều: (1) tỉnh; (2) thời gian cuộc gọi; (3) liệu người gọi là người dùng TouchPal và / hoặc trong liên hệ của người gọi; (4) âm lượng cuộc gọi đến và cuộc gọi đi; và (5) tính chủ động. Vì phần lớn phân tích, chúng tôi sử dụng nhật ký cuộc gọi kéo dài ba các tháng từ tháng 10 đến tháng 12 năm 2016. Đối với sự sống động phân tích, chúng tôi sử dụng tất cả nhật ký cuộc gọi của cả năm 2016.

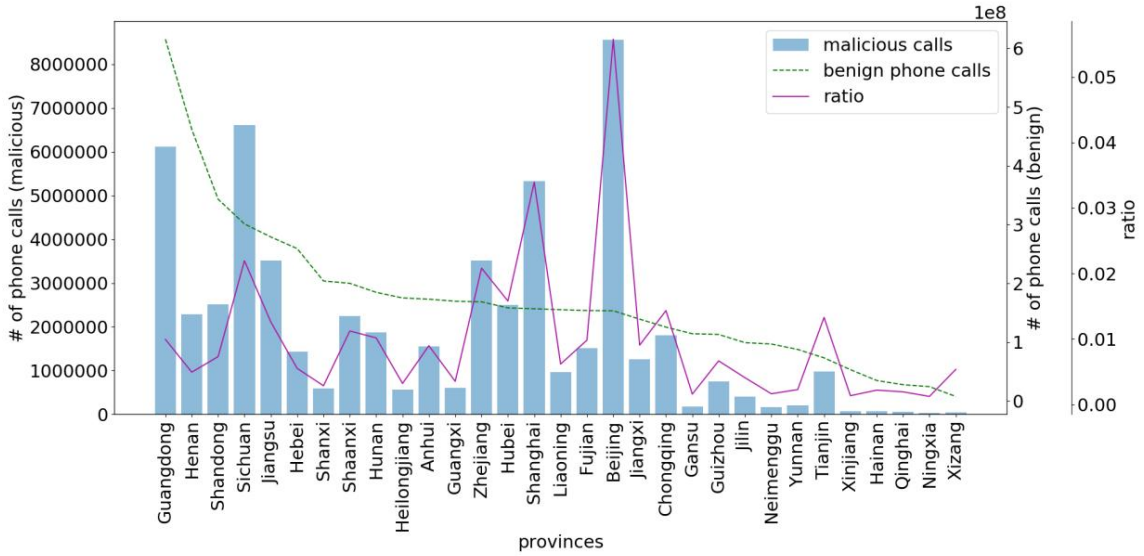
A. Mô tả nhật ký dữ liệu

Cấu trúc của mỗi bản ghi nhật ký dữ liệu được trình bày trong Bảng I. Khi người dùng TouchPal thực hiện hoặc nhận cuộc gọi, một bản ghi nhật ký sẽ được tạo. Trường loại cuộc gọi ghi lại liệu người dùng nhận cuộc gọi hoặc thực hiện cuộc gọi. Số khác trong cuộc gọi được ẩn danh bằng cách xóa tất cả các chữ số ngoại trừ một số chữ số đầu tiên tương tự như mã vùng trong các số của Hoa Kỳ . Những chữ số này chỉ chứa thông tin tỉnh về số lượng. Các MD5 muối của toàn bộ số cũng được ghi lại để nó có thể phân biệt giữa các số khác nhau cho phân tích. Khi làm như vậy, chúng tôi ngắt liên kết từ nhật ký dữ liệu đến số điện thoại thực tế để duy trì tình trạng ẩn danh. Lưu ý rằng ánh xạ từ ID người dùng đến MD5 của điện thoại có tính bảo mật cao. Chúng tôi tránh chạm vào ánh xạ này trong tất cả các phân tích của chúng tôi.

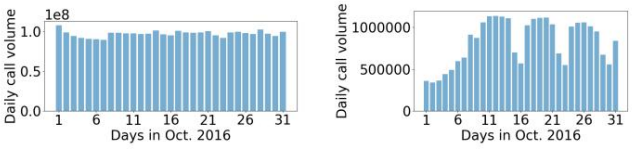
Hồ sơ chứa ba loại thông tin về gọi: (1) dấu thời gian, bao gồm cả ngày và thời gian; (2) thời gian tính bằng giây; và (3) liệu cái kia có trong danh bạ của người dùng TouchPal. Họ có thể được sử dụng để dự đoán các cuộc gọi độc hại. Mỗi bản ghi cuộc gọi cũng chứa trường thẻ cuộc gọi ghi âm cuộc gọi có thuộc về một trong 6 danh mục (tức là bình thường hoặc một trong năm thẻ). Trường này được sử dụng như sự thật cơ bản của cuộc gọi độc hại của chúng tôi nhiệm vụ dự đoán.

Trong toàn bộ công việc của mình, chúng tôi sử dụng hai bảng khác cung cấp thông tin thêm: (1) tỉnh mà mỗi người dùng TouchPal thuộc về; và (2) tập hợp tất cả các giá trị băm MD5 của TouchPal người dùng.

Chúng tôi tính toán các thống kê cơ bản của nhật ký dữ liệu từ tháng 10 đến tháng 12 năm 2016 và báo cáo chúng trong Bảng II. Lưu ý rằng mỗi cuộc gọi giữa hai người dùng TouchPal tạo ra hai bản ghi: một cho cuộc gọi đến và một cho cuộc gọi đi. Chúng ta có thể quan sát hơn 9 tỷ bản ghi cuộc gọi và hơn 500 triệu số điện thoại trong thời gian xem xét. Quy mô của tập dữ liệu đủ lớn để chúng tôi vẽ các sions conclu thú vị. Chúng tôi nhận thấy rằng các bản ghi cuộc gọi độc hại tương đối ít đối với tổng số hồ sơ, như mong đợi. Có hơn 73 triệu bản ghi cuộc gọi độc hại với khoảng 800.000

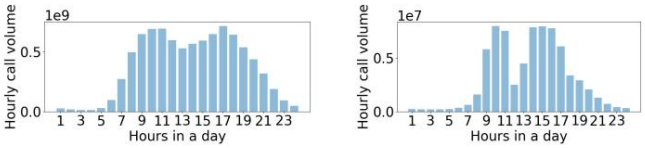


Hình 2: Biểu đồ của những người gọi đến từ các tỉnh khác nhau. Số lượng người gọi độc hại, người gọi lành tính và tỷ lệ của họ được tính toán cho mỗi tỉnh. Các tỉnh trên trục x được liệt kê theo thứ tự giảm dần trong tổng số lượng cuộc gọi của họ.



(a) cuộc gọi lành tính (b) cuộc gọi ác ý

Hình 3: Biểu đồ về số lượng cuộc gọi lành tính và cuộc gọi ác ý trong tháng 10 năm 2016. Chúng tôi bao gồm các bản ghi cuộc gọi từ tất cả các tỉnh trong kết quả.



(a) cuộc gọi lành tính (b) cuộc gọi ác ý

Hình 4: Biểu đồ phân bố hàng giờ cho cuộc gọi lành tính và cuộc gọi độc hại. Chúng tôi bao gồm các bản ghi cuộc gọi từ tất cả các tỉnh trong kết quả.

các số được sử dụng để thực hiện các cuộc gọi độc hại. Nói cách khác, trong số 100 cuộc gọi, sẽ có một cuộc gọi độc hại, cho thấy vấn đề cuộc gọi độc hại đang nghiêm trọng ở Trung Quốc.
Trung bình, mỗi số cuộc gọi độc hại thực hiện 91 cuộc gọi độc hại Vì vậy, việc xác định một số cuộc gọi độc hại sớm hơn (ví dụ: trước khi nó được sử dụng để thực hiện 10 cuộc gọi độc hại) có thể giúp ngăn chặn một lượng đáng kể các cuộc gọi độc hại.

Xác thực sự thật cơ bản. Trong nghiên cứu này, chúng tôi chủ yếu dựa vào việc gắn thẻ của người dùng để tính toán sự thật cơ bản. Để xác thực xem điều này có chính xác hay không, chúng tôi lấy mẫu ngẫu nhiên một tập hợp con các cuộc gọi độc hại và gọi lại cho họ. Chúng tôi quan sát thấy rằng hầu hết các con số không được trả lời mặc dù đã thử nhiều lần trong thời gian khác nhau. Chúng tôi xác nhận những con số như vậy là độc hại, chiếm phần lớn. Tuy nhiên, cũng có một phần nhỏ trong số những con số thực sự được trả lời. Chúng tôi thấy rằng chúng thuộc về số điện thoại cá nhân của các chuyên gia liên quan đến bán hàng (ví dụ: nhân viên ngân hàng). Lưu ý, mặc dù TouchPal cung cấp một thẻ cụ thể, hầu hết người dùng Trung Quốc vẫn coi các cuộc gọi lạnh từ các chuyên gia liên quan đến bán hàng là độc hại. Cho đến nay, chúng tôi không thể phân biệt những con số như vậy với các cuộc gọi độc hại khác trên quy mô lớn, nhưng chúng tôi coi chúng là công việc trong tương lai.

Nhận xét về đạo đức. Người dùng TouchPal phải đồng ý với Điều khoản sử dụng để truy cập vào toàn bộ chức năng của TouchPal.
TouchPal thông báo cho người dùng về việc thu thập dữ liệu thông qua Điều khoản sử dụng. Ngoài ra, người dùng TouchPal có tùy chọn chọn không tham gia để lịch sử cuộc gọi của họ sẽ không bị thu thập với chi phí là chức năng họ có thể sử dụng bị hạn chế. Nghiên cứu của chúng tôi chỉ liên quan đến những người dùng đã đồng ý với Điều khoản sử dụng.

B. Phân phối cuộc gọi trên các tỉnh thành khác nhau

Chúng tôi tính toán biểu đồ của số lượng bản ghi cuộc gọi độc hại và tất cả các bản ghi cuộc gọi cũng như tỷ lệ của chúng cho các tỉnh khác nhau và trình bày kết quả trong Hình 2. Trong hình, các tỉnh được liệt kê theo thứ tự giảm dần dựa trên tổng số cuộc gọi trong tỉnh. Chúng tôi quan sát thấy rằng sự phân bố rất lệch và một tỉnh có lượng cuộc gọi cao không có nghĩa là tỉnh đó cũng phải có lượng cuộc gọi độc hại cao.

Chúng tôi cũng thực hiện một số quan sát thú vị. Thứ nhất, số lượng người gọi độc hại có liên quan một phần đến Tổng sản phẩm quốc nội (GDP) cho mỗi tỉnh. Đối với kỳ thi, trong số 8 tỉnh hàng đầu có số lượng người gọi độc hại tối đa, 6 trong số đó cũng được xếp hạng top 6 dựa trên

trên GDP của họ vào năm 2016 [23], và hai thành phố còn lại là Bắc Kinh và Thượng Hải, tức là hai thành phố tự trị lớn nhất ở Trung Quốc. Điều này ngụ ý rằng các cuộc gọi độc hại trong một khu vực có thể liên quan đến các hoạt động kinh tế của khu vực đó. Thứ hai, tổng số lượng cuộc gọi có thể không được liên kết với GDP. Ví dụ, Sơn Tây được xếp hạng 7 dựa trên tổng số lượng cuộc gọi của họ, nhưng xếp hạng GDP của họ lại nằm trong top 10.

C. Phân phối cuộc gọi qua các ngày và giờ khác nhau

Trong phần này, chúng tôi nghiên cứu sự phân bố của các cuộc gọi độc hại và các cuộc gọi lành tính liên quan đến (1) ngày trong năm; (2) ngày trong tuần; và (3) giờ trong một ngày.

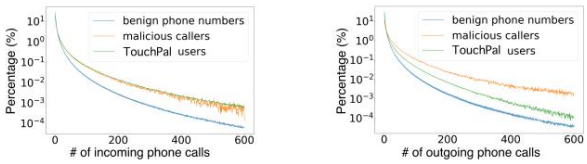
Chúng tôi vẽ biểu đồ của các cuộc gọi lành tính và các cuộc gọi độc hại cho mỗi ngày trong tháng 10 năm 2016 trong Hình 3. Chúng tôi quan sát thấy rằng số lượng bản ghi cuộc gọi độc hại trong thời gian từ ngày 1 tháng 10 đến ngày 7 tháng 10 nhỏ hơn đáng kể so với những ngày khác. Khoảng thời gian này trùng với thời điểm quan sát Ngày Quốc khánh Trung Quốc khi hầu hết công nhân đang đi nghỉ, và chúng tôi cho rằng quan sát này là do lý do. Ngoài ra, chúng tôi quan sát thấy rằng các bản ghi cuộc gọi độc hại được giảm đáng kể ba lần sau đó, tức là ngày 15-16, ngày 22-23 và ngày 29-30. Ba kỳ này đều là cuối tuần. Chúng tôi quan sát thấy hiện tượng tương tự cho tháng 11 và tháng 12 (xem Hình 12 và Hình 13 trong phần phụ lục). Do đó, chúng tôi kết luận rằng số lượng cuộc gọi độc hại có liên quan đến việc ngày của

cuộc gọi là một ngày làm việc.

Tuy nhiên, mặt khác, mối tương quan giữa khối lượng cuộc gọi lành tính và ngày làm việc không mạnh bằng khối lượng cuộc gọi độc hại. Chúng tôi quan sát thấy lượng cuộc gọi giảm trong khoảng thời gian từ ngày 2 tháng 10 đến ngày 7 tháng 10, nhưng số lượng cuộc gọi bình thường vào ngày 1 lớn hơn tất cả các ngày khác trong tháng 10.
Một lý do tiềm ẩn có thể là do quy ước xã hội Trung Quốc gọi lời chào khi bắt đầu xưng hô. Chúng tôi thiếu thông tin để phân tích thêm lý do đằng sau quan sát này, nhưng điều này có thể là lợi ích độc lập đối với một số ngành khoa học xã hội.

Chúng tôi tiếp tục phân tích mô hình hàng giờ của cuộc gọi lành tính và cuộc gọi độc hại. Biểu đồ được trình bày trong Hình 4.
Chúng tôi quan sát thấy hiện tượng tương tự: số lượng cuộc gọi độc hại trong giờ làm việc cao hơn đáng kể so với giờ ngoài giờ. Ngoài ra, các cuộc gọi ác ý từ trưa đến 1 giờ chiều, là thời gian ăn trưa điển hình, ít hơn so với các cuộc gọi từ 9 giờ sáng đến 1 giờ chiều - 5 giờ chiều. Những quan sát này cũng xác nhận rằng số lượng các cuộc gọi độc hại có liên quan nhiều hơn đến giờ làm việc so với các cuộc gọi lành tính.

Lưu ý rằng quan sát của chúng tôi rất khác so với quan sát được trình bày trong [9], cũng nhằm mục đích tìm hiểu các hành vi gọi điện ác ý của Trung Quốc. Chúng tôi cho rằng điều này là do chỉ có ít hơn 700 bản ghi cuộc gọi được thu thập trong [9], và do đó, kết quả trong [9] có thể không chắc chắn về mặt thống kê. Mặt khác, các quan sát tương tự đã được thực hiện dựa trên dữ liệu của Hoa Kỳ [15], mặc dù một số chi tiết khác nhau. Ví dụ, biểu đồ khối lượng cuộc gọi hàng giờ từ [15] trông giống với Hình 4a cho các cuộc gọi lành tính hơn là Hình 4b cho các cuộc gọi độc hại.



(a) Cuộc gọi đến (b) Cuộc gọi đi

Hình 5: Phân phối số điện thoại dựa trên khối lượng cuộc gọi đến và gọi đi của chúng.

Tất cả các quan sát trên cho thấy rằng thời gian cuộc gọi có thể là một chỉ báo hữu ích để phân biệt cuộc gọi độc hại với cuộc gọi lành tính những cái.

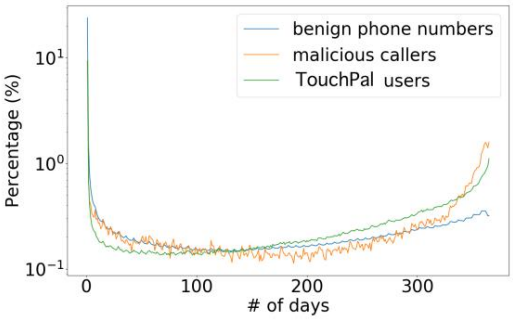
D. Người dùng TouchPal có thể không thực hiện các cuộc gọi độc hại, nhưng có thể lưu trữ những người gọi độc hại trong liên hệ của họ Bây

giờ chúng tôi điều tra xem liệu người dùng TouchPal có sử dụng số điện thoại đã đăng ký của họ để thực hiện các cuộc gọi độc hại hay không. Chúng tôi khuyến cáo rằng người dùng TouchPal có thể không sử dụng các số đã đăng ký của họ để thực hiện các cuộc gọi độc hại vì họ hiểu cơ chế cách TouchPal ngăn chặn các cuộc gọi độc hại. Chúng tôi quan sát rằng đây thực sự là trường hợp. Trong số 73 triệu bản ghi cuộc gọi ma licious, chúng tôi xác định được 103, 673 (tức là 0,14%) bản ghi có người gọi là người dùng TouchPal. Trong số các bản ghi này, chúng tôi tìm thấy 2, 541 người dùng TouchPal riêng biệt (tức là 0,007% trong số 35 triệu người dùng TouchPal hoặc 0,32% trong số 0,8 triệu người gọi độc hại). Lưu ý rằng việc ghi nhãn dữ liệu của chúng tôi không hoàn hảo và cũng có thể một số con số này bị dán nhãn sai. Do đó, chúng tôi kết luận rằng người dùng TouchPal không có khả năng sử dụng số điện thoại đã đăng ký của họ để thực hiện các cuộc gọi độc hại.

Chúng tôi phân tích thêm về việc liệu địa chỉ liên hệ của người dùng TouchPal có thể là robot gọi điện hay không. Về mặt khái niệm, chủ sở hữu của các số điện thoại được lưu trữ trong danh sách liên hệ của người dùng TouchPal có thể có mối quan hệ xã hội với người dùng và do đó không chắc họ là người gọi ác ý. Tuy nhiên, chúng tôi quan sát điều ngược lại. Trong số tất cả 73 triệu bản ghi cuộc gọi độc hại, chúng tôi quan sát thấy 9,9 triệu bản ghi trong số đó (tức là 13,56%) có người gọi nằm trong danh sách liên hệ của người gọi. Chúng tôi phỏng đoán lý do là tập dữ liệu độc hại của chúng tôi có thể chứa số lượng cá nhân của các chuyên gia liên quan đến bán hàng. Một số người dùng TouchPal chọn hợp tác kinh doanh với những chuyên gia này có thể giữ số của họ trong danh sách liên hệ, trong khi những người khác có thể dán nhãn các số này là độc hại.

E. Phân bổ số điện thoại dựa trên âm lượng cuộc gọi đến và âm lượng cuộc gọi đi

Đối với mỗi số điện thoại, chúng tôi có thể tính toán khối lượng cuộc gọi đến của nó. Đối với mỗi n, chúng ta có thể tính toán có bao nhiêu số cuộc gọi độc hại nhận được n cuộc gọi đến, sau đó tính tỷ lệ phần trăm của chúng trong số tất cả các số cuộc gọi độc hại.
Chúng tôi vẽ biểu đồ các giá trị phần trăm này bằng cách thay đổi n từ [1, 600] dưới dạng một đường trong Hình 5a. Tương tự, hai dòng khác được rút ra bằng cách xem xét các cuộc gọi đến của người dùng TouchPal và tất cả những người dùng lành tính. Từ hình này, chúng tôi quan sát thấy phân bố đuôi dài, phù hợp với các báo cáo trước đó bằng cách sử dụng



Hình 6: Sự phân bố thời gian hoạt động

tập dữ liệu nhỏ hơn [15]. Tuy nhiên, chúng tôi quan sát thấy đuôi của các số cuộc gọi độc hại và người dùng TouchPal cao hơn đáng kể so với các số điện thoại lành tính. Điều này cho thấy rằng có một tỷ lệ phần trăm người dùng cuộc gọi độc hại đang thực hiện nhiều cuộc gọi hơn những người dùng điện thoại lành tính. Do đó, chỉ số khối lượng cuộc gọi đến của một số trong nhật ký cuộc gọi có thể là một chỉ báo hiệu quả cho các cuộc gọi độc hại.

Lưu ý rằng không phải điển hình là một phần không đáng kể những người gọi độc hại nhận được một lượng lớn các cuộc gọi đến.
Chúng tôi phỏng đoán rằng điều này là do những người gọi độc hại được gắn nhãn của chúng tôi có thể chứa số cá nhân của các chuyên gia liên quan đến bán hàng như đã giải thích ở trên.

Chúng tôi tiếp tục tạo các biểu đồ tương tự cho dữ liệu cuộc gọi đi trong Hình 5b. Chúng tôi quan sát thấy hiện tượng tương tự: những người gọi độc hại có xu hướng thực hiện nhiều cuộc gọi hơn cả người dùng TouchPal và số điện thoại lành tính. Vì vậy, thông tin này có thể được tận dụng để dự đoán cuộc gọi độc hại.

F. Số điện thoại Phân phối thời gian hoạt động

Chúng tôi tính toán tổng số ngày mà một số được sử dụng tích cực bằng cách tính toán khoảng thời gian giữa cuộc gọi đầu tiên và cuộc gọi cuối cùng của nó trong nhật ký. Để phân tích có nhiều thông tin hơn, chúng tôi kết hợp nhật ký dữ liệu của toàn bộ năm 2016. Sau đó, chúng tôi tính toán phân phối theo cách tương tự như đối với phân tích lượng cuộc gọi đến và đi và vẽ biểu đồ kết quả trong Hình 6.

Chúng tôi quan sát thấy một đường cong hình "móng ngựa": có một sự sụt giảm mạnh ở đầu, sau đó lớn dần lên ở cuối. Chúng tôi cho rằng điều này là do người dùng có thể ngừng sử dụng một số sau một thời gian ngắn dùng thử hoặc tiếp tục sử dụng số đó trong thời gian dài. Hiện tượng này có ý nghĩa hơn đối với người dùng TouchPal, những người mà chúng tôi có thông tin đầy đủ trong nhật ký cuộc gọi.

Đáng ngạc nhiên là có một tỷ lệ phần trăm các số cuộc gọi độc hại được sử dụng trong suốt năm 2016 cao hơn nhiều so với các số lành tính hoặc người dùng TouchPal. Phần dưới cùng của biểu đồ phần trăm cho các số cuộc gọi độc hại xuất hiện trong khoảng 200 ngày và tỷ lệ phần trăm của nó cũng thấp hơn cả giá trị tương ứng cho người dùng TouchPal hoặc các số lành tính.

Hiện tượng này bằng cách nào đó mâu thuẫn với phân tích về những kẻ gửi thư rác trong các lĩnh vực khác, trong đó thời gian tồn tại của tài khoản thư rác là rất ngắn. Chúng tôi cũng phỏng đoán rằng lý do có thể do tập dữ liệu độc hại của chúng tôi không chỉ chứa các cuộc gọi spam / lừa đảo mà còn chứa các cuộc gọi liên quan đến bán hàng. Trong trường hợp này, các số liên quan đến bán hàng sẽ được sử dụng lâu hơn, vì chúng là một số số điện thoại cá nhân. Chúng tôi có kế hoạch điều tra chúng thêm trong tương lai.

V. PHÒNG NGỪA CUỘC GỌI MALICIOUS SỬ DỤNG MÁY HỌC

Trong phần này, chúng tôi trình bày thiết kế của chúng tôi về thuật toán dự đoán dựa trên học máy. Đặc biệt, chúng tôi cho rằng khi người dùng TouchPal A nhận được cuộc gọi từ B, TouchPal cần dự đoán liệu đây có phải là cuộc gọi độc hại hay không. Trong phần sau, trước tiên chúng ta sẽ thảo luận về thiết kế của các tính năng, sau đó trình bày các lựa chọn của chúng ta về các mô hình học máy. Vì người dùng TouchPal không có khả năng thực hiện các cuộc gọi độc hại, nên trong công việc này, chúng tôi tập trung vào những người dùng không phải TouchPal và để việc xử lý những người gọi độc hại TouchPal như một hướng đi trong tương lai.

A. Tính năng

Như chúng ta đã thảo luận trước đó, vấn đề thách thức nhất của việc phát hiện cuộc gọi độc hại là thiết kế một tập hợp các tính năng cung cấp thông tin. Tập hợp các tính năng mà chúng tôi sử dụng được liệt kê trong Bảng III.

Về mặt trực quan, tập hợp các tính năng cơ bản về một cuộc gọi bao gồm (1) liệu người gọi có trong danh bạ của người gọi hay không (đang liên lạc); (2) ngày và giờ của cuộc gọi (ngày và giờ trong tuần); và (3) người gọi có ở cùng tỉnh với người gọi hay không (cùng địa điểm). Tuy nhiên, bộ tính năng cơ bản này không cung cấp nhiều thông tin để mô hình phát hiện được chính xác. Do đó, chúng tôi mở rộng nó theo hai chiều trực giao để trích xuất thêm thông tin từ nhật ký cuộc gọi: thông tin lịch sử và thông tin tham khảo chéo. Chúng tôi giải thích chúng dưới đây.

Thông tin lịch sử. Các tính năng cơ bản chỉ sử dụng thông tin về bản ghi cuộc gọi hiện tại. Một chiều là mở rộng để xem xét tất cả các bản ghi có liên quan trong nhật ký cuộc gọi. Đặc biệt, chúng tôi truy xuất tất cả các bản ghi liên quan đến người gọi hiện tại, có thể là cuộc gọi đến hoặc cuộc gọi đi. Đối với mỗi bản ghi, chúng ta có thể tính toán một tập hợp các đối tượng địa lý và do đó thông tin lịch sử tạo thành một chuỗi các vectơ đối tượng địa lý.

Đối với các tính năng cơ bản, ngoài tập hợp các tính năng cơ bản được giải thích ở trên, có hai tính năng bổ sung có thể được tính cho mỗi bản ghi lịch sử: (1) kiểu cuộc gọi cho biết bản ghi là cuộc gọi đến hay cuộc gọi đi; và (2) thời lượng của cuộc gọi. Lưu ý rằng tính năng loại cuộc gọi bị loại trừ khỏi các tính năng tĩnh, vì trong cuộc gọi hiện tại, giá trị của nó luôn là "cuộc gọi đến". Thời lượng của cuộc gọi hiện tại không khả dụng trước khi cuộc gọi được trả lời.

Lưu ý rằng những người gọi khác nhau có thể có số lượng bản ghi lịch sử khác nhau, nhưng mô hình học máy thường lấy vectơ đặc trưng có độ dài cố định làm đầu vào. Chúng tôi sẽ giải thích cách tổng hợp tất cả các vectơ đặc trưng lịch sử khi thảo luận về các mô hình học máy cụ thể.

Thông tin tham khảo chéo. Đối với mỗi bản ghi cuộc gọi, cả bản ghi hiện tại hoặc bản ghi lịch sử, các tính năng tĩnh được xem xét cho đến nay chỉ được tính bằng cách sử dụng thông tin từ một bản ghi.

Loại	Tính năng	Lịch sử hiện tại		Giá trị	Sự mô tả
ψιι	đang liên lạc	✓	✓	Nhị phân	Số điện thoại khác có trong danh bạ của người dùng TouchPal hay không (trường nhật ký cuộc gọi)
	loại cuộc gọi		✓	Nhị phân	Cho dù cuộc gọi là cuộc gọi đến hay cuộc gọi đi (trường nhật ký cuộc gọi)
	thời lượng		✓	Số	Tổng số giây mà một cuộc gọi kéo dài (trường nhật ký cuộc gọi)
	ngày trong	✓	✓	Số	Ngày trong tuần khi cuộc gọi bắt đầu
	tuần giờ	✓	✓	Số	Giờ trong ngày khi cuộc gọi bắt đầu
	cùng một địa điểm	✓	✓	Nhị phân	Người gọi và người gọi có ở cùng tỉnh hay không
	người gọi vượt ra	✓	✓	Numeric	Có bao nhiêu cuộc gọi đi mà người gọi đã thực hiện trước khi ghi âm
	ngoài người gọi	✓	✓	Số	Có bao nhiêu cuộc gọi đến mà người gọi đã nhận được trước khi ghi âm
	người gọi ngoại lệ	✓	✓	Số	Người gọi đã gọi bao nhiêu số điện thoại khác nhau trước đây kỷ lục đang xem xét
	người gọi điện	✓	✓	Số	Có bao nhiêu số điện thoại khác nhau đã gọi cho người gọi trước đây kỷ lục đang xem xét
–	callee outs	✓	✓	Số	Tương tự như người gọi ra ngoài, người gọi trong, người gọi không đồng ý, người gọi không xác định, nhưng thống kê được tính dựa trên callee trong bản ghi được xem xét chứ không phải người gọi
	callee ins	✓	✓	Số	
	callee outdegree	✓	✓	Số	
	callee indgree n	✓	✓	Số	
	cuộc gọi là	✓		Numeric	n, trong đó bản ghi là cuộc gọi thứ n được thực hiện bởi callee trong bản ghi
	gọi lại	✓		Nhị phân	Cuộc gọi gần đây nhất của người gọi hiện tại có cùng số với cuộc gọi hiện tại hay không
	khoảng cách đến tiếp theo		✓	Số	Khoảng thời gian (tính bằng giây) giữa bản ghi được xem xét và bản ghi tiếp theo được thực hiện bởi cùng một callee

BẢNG III: Tất cả 29 tính năng đầu vào. Đối với loại giá trị, nhị phân cho biết rằng đối tượng nhận giá trị từ {0, 1}. Số cho biết đối tượng nhận một giá trị số nguyên. Giá trị này được sử dụng trực tiếp dưới dạng đối tượng một chiều hoặc được chuyển đổi thành vectơ mã hóa một nóng. Các vectơ cho tất cả các đối tượng được nối với nhau để tạo thành toàn bộ vectơ đối tượng đầu vào.

Ngoài ra, chúng tôi cũng xem xét các tính năng tham chiếu chéo mà được tính bằng cách truy cập nhiều bản ghi cuộc gọi trong nhật ký.

Đầu tiên, chúng tôi coi nhật ký cuộc gọi như một dòng hồ sơ và do đó, với bất kỳ điểm nào trong luồng, chúng tôi có thể chụp nhanh để chỉ chứa các bản ghi trước điểm. Đặc biệt, được bất kỳ bản ghi nào, chúng tôi có thể tính toán thông tin sau của mỗi người dùng dựa trên ảnh chụp nhanh của bản ghi: (1) có bao nhiêu cuộc gọi (trong); (2) có bao nhiêu cuộc gọi đi (outs); (3) có bao nhiêu số cuộc gọi đến duy nhất (không xác định); và (4) có bao nhiêu số cuộc gọi đi duy nhất (outdegree). Đối với mỗi bản ghi cuộc gọi (hiện tại hoặc lịch sử), chúng tôi có thể tính toán bốn tính năng cho cả người gọi và người gọi trong bản ghi.

Chúng tạo thành 8 loại tính năng tham chiếu chéo đầu tiên. Như chúng ta đã quan sát trong Phần IV, những thống kê này rất hữu ích để phân biệt những người gọi độc hại với những người gọi lành tính.

Thứ hai, chúng tôi coi rằng cuộc gọi hiện tại là cuộc gọi thứ n được người dùng TouchPal A nhận được từ cùng một người gọi B. Sau đó tính năng gọi n nhận giá trị n. Một cách trực quan, khi người dùng TouchPal A đã kết thúc nhiều cuộc gọi với người gọi B, ít có khả năng rằng B là một người gọi ác ý. Vì vậy, chúng tôi nghĩ rằng n cuộc gọi có thể là một tính năng hữu ích.

Thứ ba, chúng tôi xem xét liệu người gọi B vừa kết thúc cuộc gọi với cùng một người dùng TouchPal A trong cuộc gọi gần đây nhất của anh ấy. Cuộc gọi này có thể theo một trong hai hướng: từ A đến B hoặc ngược lại. Nếu đây là trường hợp này có thể là một cuộc gọi lại và do đó ít có khả năng là một cuộc gọi ác ý. Chúng tôi tính toán nó như một tính năng nhị phân được quay số lại.

Thứ tư, khoảng cách so với đối tượng địa lý tiếp theo sẽ xem xét đối với từng ghi từ callee B, khoảng cách (tính bằng giây) giữa bản ghi và bản tiếp theo. Trên thực tế, chúng tôi giả thuyết rằng mô hình cuộc gọi của người gọi độc hại có thể nổi bật hơn người dùng lành tính '. Do đó, có nhiều khả năng xác định các mẫu dựa trên về khoảng cách với các tính năng tiếp theo.

Nhận xét. Chúng tôi muốn nhận xét thêm về sự khác biệt giữa chiều tham chiếu chéo và chiều lịch sử mension. Một cách trực quan, các đối tượng địa lý lịch sử xây dựng một chuỗi các vectơ đặc trưng làm đầu vào; và mỗi tính năng tham chiếu chéo chỉ thêm một thứ nguyên bổ sung cho mỗi vectơ đặc điểm. Do đó, hai kích thước này là kích thước trực giao của đối tượng địa lý không gian.

Chúng tôi cũng nhận xét về tính mới của các tính năng của chúng tôi. Trước tác phẩm đã xem xét một số tính năng như thời lượng [41] và các loại cuộc gọi và thời gian [20]. Tuy nhiên, hầu hết các công trình đề xuất các tính năng đặc biệt dựa trên trực giác. Trong công việc của chúng tôi, ngược lại, các tính năng được thiết kế theo hướng dẫn của một nghiên cứu đo lường quy mô lớn, và đề xuất một cách có hệ thống theo hai khía cạnh chung đã thảo luận ở trên. Ngoài ra, một số tính năng, chẳng hạn như liên hệ, cùng một vị trí và tất cả các tính năng tham chiếu chéo đều mới đề xuất trong công việc này.

B. Mô hình học máy

Vấn đề dự đoán cuộc gọi độc hại là một tệp nhị phân tiêu chuẩn vấn đề phân loại: phân loại đầu vào thành tích cực (độc hại) hoặc tiêu cực (lành tính). Chúng tôi sử dụng một số mô hình học máy hiện đại nhất cho vấn đề này: mạng lưới; mô hình rừng ngẫu nhiên [10]; Máy vectơ hỗ trợ (SVM) các mô hình [13]; và các mô hình hồi quy logistic [14].

Chúng tôi muốn nhấn mạnh rằng trong khi các đối tượng địa lý phi lịch sử có thứ nguyên cố định, thứ nguyên lịch sử tạo thành một chuỗi đầu vào vectơ. Do đó, chúng ta cần chuyển chuỗi thành vectơ đầu vào có kích thước cố định. Đối với các cách tiếp cận đã đề cập ở trên, chúng ta chỉ cần lấy giá trị trung bình của tất cả các vectơ trong dãy. Tuy nhiên, ngoài ra, chúng tôi cũng có thể sử dụng một nơ-ron tái phát mạng [16], được thiết kế để tính toán một lần nhúng vectơ từ một chuỗi các vectơ đầu vào. Do không gian hạn chế, chúng tôi giải thích chi tiết mô hình trong phần phụ lục.

Chúng tôi cũng muốn lưu ý rằng tất cả các mô hình được xem xét sẽ phát ra một điểm p, cho biết xác suất của dự đoán là tích cực. Do đó, chúng ta có thể đặt ngưỡng mô hình τ để ,

mô hình sẽ dự đoán độc hại khi p ≥ τ và ngược lại ,
ngược lại. Bằng cách điều chỉnh τ một mô hình có thể tạo ra sự cân bằng giữa
độ chính xác và thu hồi của nó.

VI. SỰ ĐÁNH GIÁ

Trong phần này, chúng tôi đánh giá việc học máy khác nhau
các phương pháp tiếp cận về tính hiệu quả của chúng để ngăn chặn các số cuộc
gọi khó hiểu so với danh sách đen đơn giản
cách tiếp cận. Trong phần sau, trước tiên chúng tôi sẽ giải thích thử nghiệm
thành lập. Sau đó, chúng tôi sẽ kiểm tra hiệu suất của các mô hình khác nhau
(1) khi được đào tạo và kiểm tra bằng cách sử dụng dữ liệu từ cùng một tỉnh;
(2) khi được đào tạo ở tỉnh này và được kiểm tra ở tỉnh khác; và
(3) độ trễ tổng thể của chúng. Trong phần tiếp theo, chúng ta hiểu
hiệu quả của các tính năng khác nhau bằng phân tích cắt bỏ và
kiểm tra các tính năng được chọn bởi một mô hình hoạt động tốt,
đó là một khu rừng ngẫu nhiên.

Một thiết lập

Trong phần này, chúng tôi giải thích thiết lập thử nghiệm. Chúng tôi sẽ
bắt đầu với chi tiết triển khai của các mô hình khác nhau theo sau
bằng các số liệu khác nhau được sử dụng để đánh giá một mô hình. Cuối cùng, chúng tôi
trình bày các chi tiết về đào tạo và xây dựng bộ thử nghiệm.

Chi tiết thực hiện mô hình. Đối với mạng nơron vani,
SVM và hồi quy logistic, chúng tôi sử dụng tính năng tích hợp sẵn của chúng
từ sklearn [26]. Chúng tôi gọi chúng là NN, SVM và
LR tương ứng. Chúng tôi triển khai mô hình RNN dựa trên LSTM
trong Tensorflow [5]. Chúng tôi gọi nó là RNN. Cho ngẫu nhiên
mô hình rừng, chúng tôi sử dụng hai cách triển khai, một từ sklearn,
và cái khác từ XGBoost [12]. Chúng tôi gọi hai điều này là
RF và XGBoost tương ứng.

Các thước đo đánh giá. Chúng tôi sử dụng hai số liệu trong đánh giá của mình.
Một là điểm AUC, là thước đo tiêu chuẩn để đánh giá
hiệu suất của mô hình học máy. Lưu ý rằng chúng tôi thích AUC hơn
điểm số so với các chỉ số tiêu chuẩn khác như độ chính xác, thu hồi,
và điểm F-1, vì điểm AUC phù hợp hơn với dữ liệu
xiên (tức là, trong trường hợp của chúng tôi, các ví dụ phủ định nhiều hơn 100 lần so với
những cái tích cực).

Lưu ý rằng các thuộc tính mong muốn của chúng tôi về mô hình là: (1) hầu hết
trong số các cuộc gọi lành tính không nên được dự đoán là độc hại
cuộc gọi; và (2) một mô hình phải xác định một cuộc gọi độc hại mới
bằng cách quan sát số lượng cuộc gọi độc hại tối thiểu.
Để kiểm tra xem một mô hình có thể đạt được hai mục tiêu này tốt như thế nào
so với cách tiếp cận danh sách đen, chúng tôi thiết kế một
chỉ số, dự đoán đầu tiên ở ngưỡng nhãn M và độ chính xác
p, gọi tắt là FP @ (M, p). M là ngưỡng nhãn. Đó là một
số điện thoại được gắn nhãn là số cuộc gọi độc hại khi nó là
được người dùng TouchPal gắn nhãn ít nhất M lần. Theo đánh giá của chúng tôi,
chúng tôi coi một cài đặt đơn giản hóa mà mọi cuộc gọi độc hại sẽ
được gắn nhãn như vậy, và do đó, cách tiếp cận danh sách đen sẽ vượt qua lúc
ít nhất M cuộc gọi độc hại trước khi số đó có thể được ngăn chặn.

Chúng tôi cũng bắt buộc rằng mô hình có thể đạt được độ chính xác ≥ p
trên các cuộc gọi lành tính. Trên thực tế, chúng tôi luôn có thể tăng mô hình
ngưỡng τ để tăng độ chính xác của mô hình. Ví dụ,
trong trường hợp cực đoan, chúng ta luôn có thể đặt τ = + ∞, sao cho gần như
tất cả các cuộc gọi được dự đoán là các cuộc gọi lành tính để đạt được độ chính xác là

Người mẫu	Bắc Kinh	Tứ Xuyên	Quảng Đông	Thượng Hải	Chiết Giang	
RF	0,9985	0,9984	0,9978	0,9978	0,9981	
XGBoost	0,9979	0,9981	0,9972	0,9969	0,9977	
NN	0,9978	0,9972	0,9961	0,9966	0,9976	
RNN	0,9972	0,9962	0,9957	0,9965	0,9975	
SVM	0,9914	0,9927	0,9895	0,9892	0,9930	
LR	0,9846	0,9822	0,9770	0,9807	0,9848	

BẢNG IV: Điểm AUC của các kiểu máy khác nhau. Mỗi mô hình là
được đào tạo và kiểm tra sử dụng dữ liệu từ cùng một tỉnh. Các
dữ liệu đào tạo sử dụng hồ sơ từ tháng 10 đến tháng 11
2016 và dữ liệu thử nghiệm sử dụng các bản ghi vào tháng 12 năm 2016.

100%. Tuy nhiên, trong trường hợp này, τ được đặt quá cao để chụp được
cuộc gọi độc hại. Do đó, chúng ta định nghĩa τ (p) là τ nhỏ nhất nên
rằng độ chính xác của mô hình đối với các lệnh gọi lành tính ít nhất là p.

Đưa ra một mô hình có τ (p) và một số cuộc gọi độc hại, chúng tôi
quan tâm đến số lượng bản ghi cuộc gọi cần được quan sát
trước khi con số này có thể được dự đoán là độc hại. Giá trị này là
sau đó được định nghĩa là FP @ (M, p). Về mặt hình thức, đã cho một số, có
bản ghi cuộc gọi là R1, ..., Rn, FP @ (M, p) được định nghĩa là
i nhỏ nhất để mô hình với τ (p) dự đoán đầu vào
được tạo ra từ R1, ..., Ri là một cuộc gọi độc hại. Nếu tôi> M
hoặc không có i nào như vậy tồn tại, thì FP @ (M, p) được định nghĩa là
(M + 1). Với một tập hợp các số cuộc gọi độc hại, do đó chúng tôi có thể
xác định FP trung bình @ (M, p) (hay viết tắt là AFP @ (M, p)) thành
là giá trị trung bình của tất cả các giá trị FP @ (M, p) cho cuộc gọi độc hại
số trong tập hợp.

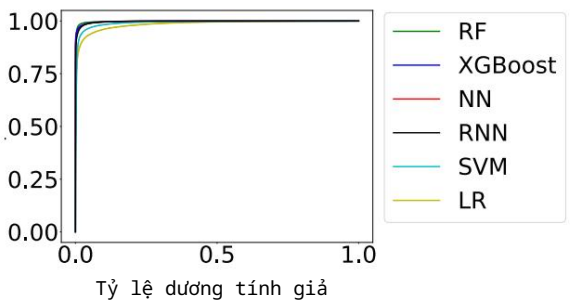
Chúng tôi muốn nhận xét rằng trong chỉ số FP (và AFP), chúng tôi
bao gồm tham số M cho âm thanh. Đó là, đối với một số
trường hợp, mô hình sẽ không bao giờ dự đoán một số cuộc gọi độc hại
như độc hại. Trong trường hợp này, giá trị FP của nó sẽ là + ∞ mà không có
cung cấp giá trị M và do đó là chỉ số AFP, sẽ
be + ∞, không phải là biểu thị. Chúng tôi giảm thiểu vấn đề này bằng cách bao gồm
tham số M trong chỉ số.

Xây dựng dữ liệu. Với một khoảng thời gian và một tỉnh,
chúng tôi xây dựng dữ liệu đào tạo bằng cách chọn tất cả
hồ sơ cuộc gọi từ tỉnh nhất định và trong thời gian nhất định
Giai đoạn. Vì có nhiều bản ghi cuộc gọi lành tính hơn
bản ghi cuộc gọi độc hại, chúng tôi lấy mẫu một tập hợp con của tất cả các cuộc gọi lành tính
chứa cùng một lượng số điện thoại lành tính như
độc hại để duy trì sự cân bằng của tích cực và tiêu cực
mẫu trong tập huấn luyện. Trong mỗi thử nghiệm của chúng tôi, chúng tôi lấy
mẫu lại nhóm đào tạo 5 lần và lấy trung bình kết quả của chúng thành
làm cho chúng trở nên mạnh mẽ đối với quy trình lấy mẫu.

Tương tự, tập kiểm tra được xây dựng theo cách tương tự. Khi nào
tính toán số liệu AFP @ p, tuy nhiên, chúng tôi sử dụng tất cả các
hồ sơ cuộc gọi từ tỉnh nhất định và trong khoảng thời gian nhất định
của thời gian. Điều này là do AFP @ p không nhạy cảm với tỷ lệ
của các mẫu dương tính và âm tính.

B. Thí nghiệm độ chính xác

Trong phần này, chúng tôi đánh giá việc học máy khác nhau
mô hình về khả năng tổng quát hóa của chúng. Đầu tiên chúng ta sẽ kiểm tra
khả năng tổng quát hóa theo thời gian, và sau đó kiểm tra
tính chung cho các tỉnh khác nhau. Chúng tôi trình bày các chi tiết
phía dưới.



Hình 7: Đường cong ROC của các mô hình khác nhau được đào tạo bằng cách sử dụng hồ sơ nhật ký cuộc gọi của Bắc Kinh trong tháng 10 và tháng 11 năm 2016 và được thử nghiệm trên hồ sơ nhật ký cuộc gọi của Bắc Kinh trong tháng 12 năm 2016

1) Khả năng tổng quát hóa theo thời gian: Theo trực giác, chúng tôi hy vọng một mô hình được đào tạo trên các bản ghi dữ liệu hiện tại có thể hoạt động tốt trong tương lai. Chúng tôi gọi đặc tính này là tính tổng quát hóa theo thời gian. Chúng tôi chọn top 5 tỉnh có số lượng cuộc gọi độc hại nhiều nhất. Đối với mỗi tỉnh, chúng tôi đào tạo một mô hình sử dụng dữ liệu từ tỉnh này trong tháng 10 và tháng 11, và thử nghiệm mô hình này bằng cách sử dụng dữ liệu từ tỉnh đó trong tháng 12.

Điểm AUC. Kết quả AUC được trình bày trong Bảng IV. Từ bảng, chúng ta có thể quan sát thấy hầu hết các phương pháp trên bất kỳ tỉnh nào đều có thể đạt điểm AUC từ 0,985 trở lên. Điều này cho thấy mô hình rất chính xác trong việc dự đoán các cuộc gọi độc hại.
Các mô hình khác nhau được liệt kê trong bảng theo thứ tự giảm dần hiệu suất của chúng, từ trên xuống dưới. Chúng ta có thể quan sát thấy rằng đối với mỗi tỉnh, thứ tự này giống hệt nhau.
RF (tức là từ việc triển khai sklearn) đạt được điểm AUC tốt nhất trên tất cả các tỉnh, và hiệu suất AUC của nó ít nhất là 0,9978. Ngoài ra, việc triển khai XGBoost có thể đạt được điểm AUC tương tự mặc dù thấp hơn một chút.

Hai phương pháp tiếp cận mạng nơ-ron tuân theo các phương pháp tiếp cận rừng ngẫu nhiên. Một số lý do tiềm ẩn có thể gây ra điều này: (1) công suất của mô hình không đủ lớn; (2) vấn đề có không gian đầu vào có kích thước thấp, mà cách tiếp cận mạng nơ-ron có thể không phải lúc nào cũng là tốt nhất; và (3) rừng ngẫu nhiên về cơ bản là một cách tiếp cận tổng hợp, trong khi chúng tôi không sử dụng nhóm cho các cách tiếp cận NN của mình. Chúng tôi để nguyên nhân để điều tra thêm. Đáng ngạc nhiên là hiệu suất của RNN không tốt bằng NN. Điều này có thể một phần là do hiệu suất thu được từ các tính năng lịch sử không lớn lắm.

Chúng ta sẽ xem xét kỹ hơn giả thuyết này trong Phần VII.

Hai cách tiếp cận truyền thống khác, tức là SVM và hồi quy logistic, không hiệu quả như các lựa chọn thay thế khác. Điều này là hợp lý, vì cả hai cách triển khai về cơ bản đều là bộ phân loại tuyến tính, có thể không đủ biểu cảm để xử lý vấn đề.

Đường cong ROC. Để hiểu rõ hơn về điểm AUC, chúng tôi vẽ đồ thị đường cong ROC cho mô hình được đào tạo và thử nghiệm dựa trên dữ liệu của Bắc Kinh trong Hình 7. Từ hình này, chúng ta có thể quan sát thấy các khu vực dưới đường cong gần như chiếm toàn bộ lô đất – - và do đó điểm AUC gần bằng 1. Điều này cho thấy rằng đối với hầu hết các kiểu máy, ngưỡng τ có thể được điều chỉnh thích hợp để đạt được thu hồi rất cao (nghĩa là giá trị trục y đạt đến 1) trong khi rất ít số lành tính được dự đoán là độc hại (nghĩa là giá trị trục x gần bằng 0). Do đó, đường cong ROC càng khẳng định tính hiệu quả của phương pháp tiếp cận của chúng tôi.

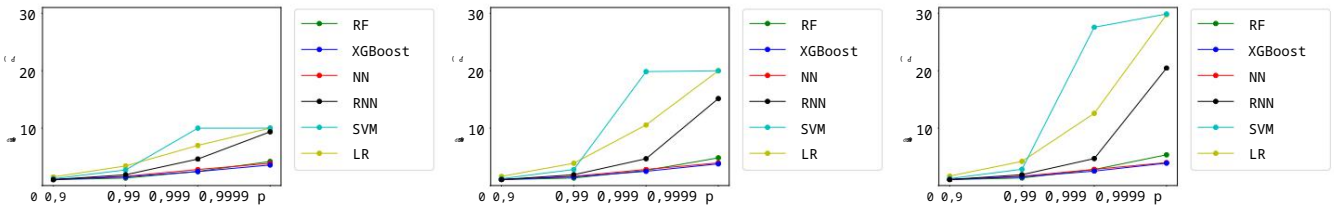
Dự đoán trung bình đầu tiên. Bây giờ chúng tôi trình bày kết quả bằng cách sử dụng số liệu AFP @ (M, p) cho M = 10, 20, 30. Kết quả được trình bày trong Hình 8. Từ hình này, chúng tôi quan sát thấy XG Boost và NN tốt hơn RF một chút và AFP Điểm @ (M, p) của cả ba mô hình này luôn dưới 5,5. Mặt khác, đối với ba mô hình khác, tức là RNN, SVM và LR, điểm AFP @ (M, p) của chúng gần với M, khi p được đặt lớn, tức là 99,99%. Lý do là các mô hình này rất khó đạt được độ chính xác cao trên dữ liệu lành tính để đáp ứng yêu cầu về độ chính xác, và do đó chúng có xu hướng gắn nhãn bất kỳ cuộc gọi nào là lành tính. Trong trường hợp này, các mô hình không hiệu quả trong việc dự đoán các cuộc gọi độc hại. Các mô hình rừng ngẫu nhiên và mạng nơ-ron không tuần hoàn không bị vấn đề này.

Bằng cách thay đổi M từ 10 đến 30, chúng tôi nhận thấy rằng điểm AFP @ (M, p) của mỗi trong số ba mô hình tốt nhất tăng nhẹ. Ví dụ, điểm của mẫu tốt nhất, XGBoost, tăng từ 3,57 lên 3,90. Điều này là do các mô hình này có thể dự đoán một số điện thoại là số cuộc gọi độc hại sớm hơn rất nhiều trước khi đạt đến ngưỡng M.

Chúng tôi xem xét tổng số cuộc gọi độc hại không bị chặn bởi phương pháp danh sách đen và phương pháp học máy tốt nhất của chúng tôi, XGBoost. Giả sử có N số cuộc gọi độc hại khác nhau, thì cách tiếp cận danh sách đen không thể ngăn chặn N · M cuộc gọi độc hại trước khi nó bắt đầu có hiệu quả; Mặt khác, sử dụng XGBoost, số tiền này là N · (AFP @ (M, p) - 1). Do đó, XGBoost có thể giảm số lượng cuộc gọi độc hại không bị chặn xuống 1 - Theo đánh giá của chúng tôi, XGBoost có thể đạt được tỷ lệ giảm cuộc gọi không bị chặn từ 75,3% (tức là M = 10) xuống 90,3% (tức là M = 30). $\frac{\text{AFP @ (M, p)}}{M}$ 1

Chúng tôi tiếp tục điều tra những dự đoán đầu tiên. Đặc biệt, chúng tôi đặt τ của mỗi mô hình để đạt được độ chính xác ≥ p. Sau đó, với mỗi n {1, ..., 30}, chúng ta xây dựng dữ liệu thử nghiệm bằng cách chỉ giữ lại n bản ghi lệnh gọi đầu tiên của một số. Trong bộ thử nghiệm này, chúng tôi có thể tính toán cuộc gọi độc hại dưới dạng tỷ lệ phần trăm các số cuộc gọi độc hại được dự đoán chính xác chỉ sử dụng n bản ghi cuộc gọi đầu tiên của chúng. Chúng tôi gọi chỉ số này là MR @ (n, p) và nó cung cấp thông tin chi tiết hơn AFP @ (M, p). Bằng cách đặt p = 99,99%, chúng tôi vẽ đồ thị đường cong MR @ (n, p) trong Hình 9.

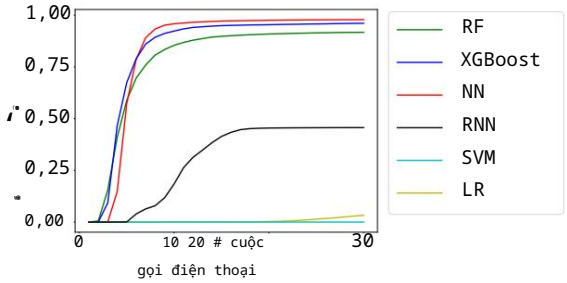
Chúng tôi quan sát thấy rằng XGBoost và NN có thể đạt đến mức gọi lại cuộc gọi độc hại cao hơn 80% bằng cách quan sát 6 cuộc gọi trở xuống, nhưng khó có kiểu máy nào đạt được mức gọi lại cuộc gọi độc hại cao hơn 92%. Bằng cách kiểm tra chặt chẽ các đường cong, chúng tôi nhận thấy rằng mức thu hồi của NN vẫn ở mức thấp, và sau đó NN vượt qua tất cả các cách tiếp cận khác sau n = 7. Mức thu hồi của XGBoost hầu như luôn là tốt nhất cho đến khi bị NN vượt qua. Điều này cho thấy NN cần quan sát nhiều bản ghi cuộc gọi hơn để đưa ra dự đoán hiệu quả, trong khi XGBoost yêu cầu ít hơn. Chúng tôi kết luận rằng các phương pháp học máy tốt nhất, XGBoost và NN, có thể nắm bắt hầu hết các cuộc gọi độc hại bằng cách quan sát ít bản ghi cuộc gọi hơn nhiều so với cách tiếp cận danh sách đen.



(a) AFP @ (10, p) (b) AFP @ (20, p) (c) AFP @ (30, p)

Hình 8: AFP @ (M, p) của các mô hình khác nhau được đào tạo bằng cách sử dụng hồ sơ nhật ký cuộc gọi của Bắc Kinh trong tháng 10 và tháng 11 năm 2016, và được thử nghiệm trên hồ sơ nhật ký cuộc gọi của Bắc Kinh trong tháng 12 năm 2016.

.



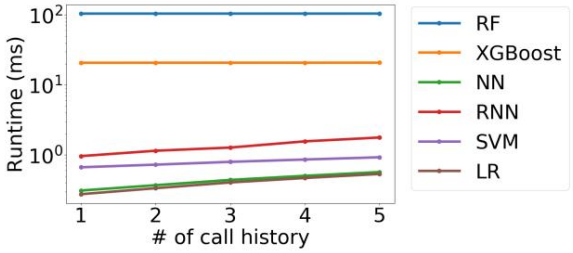
Hình 9: MR @ (n, 0,9999) của các mô hình khác nhau được đào tạo bằng cách sử dụng hồ sơ nhật ký cuộc gọi của Bắc Kinh trong tháng 10 và tháng 11 năm 2016 và được thử nghiệm trên hồ sơ nhật ký cuộc gọi của Bắc Kinh trong tháng 12 năm 2016

2) Khả năng tổng quát hóa đến các địa điểm khác: Trong phần này, chúng tôi đánh giá liệu mô hình được đào tạo trên dữ liệu từ tỉnh này có thể tổng quát hóa cho tỉnh khác hay không. Chúng tôi tiến hành thử nghiệm này với hai mục đích. Đầu tiên, vì tập hợp số điện thoại từ các tỉnh khác nhau hoàn toàn tách biệt với nhau, thử nghiệm này có thể cung cấp cho chúng ta những hiểu biết sâu hơn về việc liệu mô hình có thể tổng quát hóa thành các mô hình không nhìn thấy được hay không. Thứ hai, một số khu vực có thể có quá ít hồ sơ cuộc gọi để đào tạo một mô hình hiệu quả, đặc biệt là ở giai đoạn đầu phát triển kinh doanh trong khu vực đó. Trong trường hợp này, sử dụng một mô hình được đào tạo với dữ liệu từ một tỉnh khác là một giải pháp đầy hứa hẹn. Thử nghiệm này rất hữu ích để làm sáng tỏ những ứng dụng như vậy.

Chúng tôi đào tạo một mô hình sử dụng nhật ký cuộc gọi từ Bắc Kinh và đánh giá mô hình đó dựa trên dữ liệu từ các tỉnh khác. Chúng tôi quan sát thấy hiện tượng tương tự về hiệu suất của các mô hình khác nhau về cả điểm AUC và giá trị AFP @ (M, P). Do giới hạn về không gian, chúng tôi chuyển các chi tiết của thí nghiệm sang phần phụ lục. Chúng tôi cũng thử nghiệm với các mô hình được đào tạo sử dụng dữ liệu từ các tỉnh khác nhau. Các quan sát của chúng tôi là nhất quán trong tất cả các thử nghiệm. Do đó, chúng tôi kết luận rằng mô hình được đào tạo trên một tỉnh với khối lượng cuộc gọi lớn có thể tổng quát hóa thành các số không nhìn thấy từ tỉnh khác.

C. Đánh giá thời gian chạy

Trong tiểu mục này, chúng tôi đánh giá hiệu quả thời gian chạy cho các mô hình khác nhau. Các thử nghiệm được chạy trên một máy chủ được trang bị CPU Intel i7-6900K với 15 lõi chạy ở tốc độ 3,20GHz và bộ nhớ 96GB. Theo đánh giá của chúng tôi, tất cả dữ liệu được tải trước vào bộ nhớ để loại bỏ độ trễ I / O. Chúng tôi lặp lại từng



Hình 10: Thời gian chạy của các mô hình khác nhau

thử nghiệm 5 lần và tính kết quả trung bình của các lần chạy khác nhau.

Đối với mỗi mô hình, chúng tôi chuẩn bị một chuỗi các tính năng lịch sử và các tính năng cho bản ghi hiện tại. Chúng tôi tính toán độ trễ dự đoán mô hình từ khi xử lý chuỗi tính năng thô cho đến khi đưa ra dự đoán. Lưu ý rằng để truy xuất chuỗi tính năng thô, chúng tôi có thể tận dụng cơ sở hạ tầng lưu trữ khóa-giá trị hiện có được triển khai trong quy trình sản xuất TouchPal, thường mất từ 1ms đến 10ms để truy xuất và cập nhật một bản ghi. Do đó, một mô hình hiệu quả không nên gánh chịu một khoản chi phí đáng kể.

Đối với mỗi mô hình, chúng tôi cung cấp n = 1, ..., 5 bản ghi lịch sử để xây dựng các đầu vào nhằm kiểm tra tính hiệu quả của độ dài của các đối tượng địa lý lịch sử đối với độ trễ dự đoán của mô hình. Kết quả được trình bày trong Hình 10. Chúng ta có thể quan sát thấy rằng các mô hình rừng ngẫu nhiên có độ trễ thời gian chạy cao, từ 20ms đến hơn 100ms. Điều này là do sự phức tạp của mô hình.

Lưu ý rằng mỗi khu rừng ngẫu nhiên chứa 100 cây quyết định và mỗi cây quyết định có ba cấp độ. Việc tính toán liên quan đến quá trình dự đoán của một khu rừng ngẫu nhiên lớn hơn nhiều so với các mô hình khác.

Chúng tôi nhận thấy rằng đối với tất cả các mô hình khác, độ trễ thời gian chạy nhỏ hơn 2ms, đây là một mức chi phí hợp lý. Đặc biệt, chúng tôi thấy rằng mô hình NN, đạt được kết quả AUC và AFP @ (M, p) tốt thứ ba, cũng đạt được độ trễ dự đoán mô hình thấp thứ hai. Điều này phần lớn là do sự đơn giản của mô hình. Do đó, trong trường hợp khi cần phải cân bằng giữa độ chính xác và độ trễ thời gian chạy, mạng nơ-ron không lặp lại có thể là lựa chọn tốt nhất trong thực tế.

Hơn nữa, chúng tôi nhận thấy rằng khi độ dài của chuỗi đối tượng địa lý lịch sử tăng lên, thì độ trễ dự đoán của mô hình cho NN,

	RF	XGBoost	NN	RNN	SVM	LR		
Tất cả	0,9984		0,9979		0,9977	0,9974	0,9913	0,9846
-His	0,9978	-CR		0,9978	0,9961	0,9934	0,9890	0,9730
	0,9444	Cơ bản		0,9482	0,9524	0,9556	0,9350	0,9302
	0,9079			0,9112	0,9094	0,9084	0,9023	0,8976

BẢNG V: Kết quả phân tích cắt bỏ. "Tất cả" cho biết tất cả các tures fea được sử dụng; "-CR" cho biết loại trừ tất cả các tính năng tham chiếu chéo khỏi đầu vào; "-Nhà" loại trừ tất cả các tính năng lịch sử khỏi đầu vào; "Basic" tương đương với "-CR-His". Mỗi ô (i, j) chỉ ra điểm AUC của một mô hình tại cột j được huấn luyện với dữ liệu bằng cách sử dụng bộ tính năng đầu vào tương ứng với hàng i.

RNN, SVM, LR tăng nhẹ. Tuy nhiên, đối với việc triển khai RF và XGBoost, độ trễ dự đoán mô hình cho các độ dài chuỗi khác nhau không thể hiện sự khác biệt có thể quan sát được. Điều này là do độ dài trình tự chỉ xác định thời gian để xây dựng các tính năng đầu vào, trong khi độ trễ dự đoán mô hình cho các mô hình rừng ngẫu nhiên bị chi phối bởi thời gian dự đoán sau khi đầu vào đã được xử lý trước.

Lưu ý rằng XGBoost phải chịu chi phí là 20ms. Do đó, đối với việc sử dụng thực tế, cách tiếp cận mạng nơ-ron có thể phù hợp hơn vì nó đạt được hiệu quả tương đương, nhưng đòi hỏi thời gian suy luận ít hơn nhiều. Tuy nhiên, có thể tối ưu hóa nhiều hơn để đẩy nhanh hơn nữa việc cố vấn rừng ngẫu nhiên. Chúng tôi kết luận rằng hầu hết các mô hình học máy được đề xuất trong Phần V đều phải chịu một khoản chi phí hợp lý nhỏ để được triển khai trong quy trình sản xuất thực tế.

VII. HIỂU BIẾT TÁC DỤNG CỦA CÁC TÍNH NĂNG.

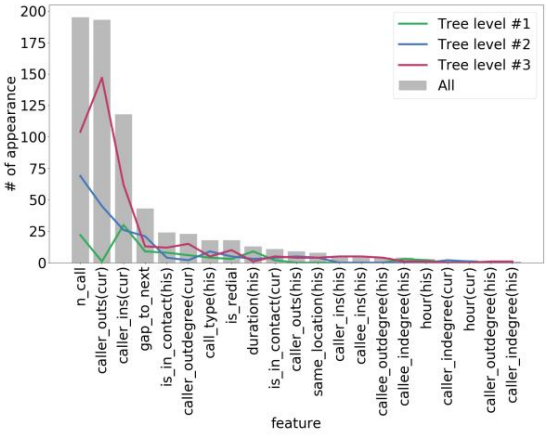
Trong phần này, chúng tôi phân tích hiệu quả của các tính năng được đề xuất. Trước tiên, chúng tôi sẽ thực hiện một nghiên cứu cắt bỏ xem việc thêm các tính năng lịch sử và / hoặc các tính năng tham khảo chéo có thực sự giúp cải thiện hiệu suất hay không. Sau đó, chúng tôi sẽ phân tích một trong những cây quyết định hiệu quả nhất để xem xét các tính năng nào là quan trọng trong quá trình ra quyết định.

A. Phân tích cắt bỏ

Trong phần này, chúng tôi trình bày phân tích cắt bỏ bằng cách loại bỏ các đối tượng địa lý lịch sử và / hoặc đối tượng địa lý tham khảo chéo. Chúng tôi tạo dữ liệu đào tạo bằng cách sử dụng nhật ký cuộc gọi tháng 10 và tháng 11 của Bắc Kinh và dữ liệu thử nghiệm bằng cách sử dụng nhật ký cuộc gọi tháng 12 của Bắc Kinh. Kết quả được trình bày trong Bảng V. Chúng tôi quan sát thấy rằng điểm AUC của bất kỳ mô hình nào chỉ sử dụng bộ tính năng cơ bản đều rất thấp, tức là khoảng 0,9. Vì vậy, việc sử dụng nhiều tính năng hơn từ nhật ký cuộc gọi là cần thiết để đạt được hiệu suất tốt hơn.

Chúng tôi cũng nhận thấy rằng việc thêm bất kỳ bộ tính năng nào sẽ cải thiện hiệu suất, mặc dù các cải tiến là khác nhau. Chúng tôi nhận thấy rằng việc thêm các tính năng tham khảo chéo (-His) vào các tính năng cơ bản có thể cải thiện 10% điểm AUC; Mặt khác, bằng cách thêm các tính năng lịch sử (-CR), mức cải thiện chỉ khoảng 4% - 5%.

Sử dụng tất cả các đối tượng địa lý là cách tiếp cận chính xác nhất, nhưng chúng tôi nhận thấy rằng sự cải thiện từ việc thêm các đối tượng địa lý lịch sử lên trên các đối tượng địa lý tham khảo chéo là rất nhỏ (ví dụ: <0,001). Khi chúng ta muốn đạt được hiệu quả tốt hơn mà không phải hy sinh



Hình 11: Biểu đồ của các đối tượng địa lý dựa trên tần suất của chúng được sử dụng trong khu rừng ngẫu nhiên hiệu quả nhất. Các tính năng được liệt kê từ trái sang phải theo thứ tự giảm dần tần số của chúng. Chúng tôi cũng hiển thị ba ô dựa trên tần suất xuất hiện của từng đối tượng ở các cấp độ khác nhau của các cây quyết định trong rừng.

độ chính xác, tuy nhiên, nó có thể hiệu quả hơn bằng cách chỉ sử dụng các tính năng tham khảo chéo.

B. Tìm hiểu các tính năng được sử dụng bởi một khu rừng ngẫu nhiên hoạt động tốt

Bây giờ chúng tôi phân tích xem tính năng nào trong số 29 tính năng được sử dụng nhiều hơn bởi một mô hình hoạt động tốt và kiểm tra tính năng nào là tiêu biểu nhất để diễn giải hiệu suất của các mô hình học máy. Để làm điều này, chúng tôi chọn một mô hình rừng ngẫu nhiên hiệu quả nhất được đào tạo bằng XGBoost. Nó chứa 100 cây quyết định với mỗi cây có 3 lớp. Chúng ta hình dung một cây quyết định trong phần phụ lục (xem Hình 15).

Mỗi cây quyết định sử dụng 7 tính năng để đạt được một lá (lưu ý các tính năng được sử dụng bởi các cấp thấp hơn có thể bị trùng lặp). Do đó, chúng tôi tính toán tần suất của từng tính năng đang được sử dụng trong số tất cả 100 cây quyết định. Chúng tôi vẽ biểu đồ dựa trên tần suất này trong Hình 11. Chúng tôi có thể quan sát sự phân bố đuôi dài.

Trong mô hình này, 3 tính năng hàng đầu, cụ thể là n cuộc gọi, người gọi ngoài cuộc gọi hiện tại và người gọi trong cuộc gọi hiện tại, được sử dụng thường xuyên hơn nhiều so với các tính năng khác. Đặc điểm lịch sử được sử dụng thường xuyên nhất trong khu rừng ngẫu nhiên này là liên hệ và nó được sử dụng bởi ít hơn 25 cây quyết định. Quan sát này phù hợp với phân tích cắt bỏ của chúng tôi.

Lưu ý rằng các tính năng ở cấp độ cao hơn sẽ được kiểm tra thường xuyên hơn các tính năng ở cấp độ thấp hơn. Do đó, chúng tôi phân tích sự phân bố tính năng ở các cấp độ khác nhau. Chúng tôi vẽ ba dòng tương ứng với ba mức trong Hình 11. Chúng tôi quan sát thấy rằng tổng tần suất của một đối tượng địa lý thường được căn chỉnh với tần số của mỗi mức. Ví dụ, tính năng được sử dụng thường xuyên nhất trong mô hình, cuộc gọi n, thường được sử dụng ở cả ba cấp độ. Tuy nhiên, có một vài trường hợp ngoại lệ. Ví dụ: tính năng thường xuyên thứ hai, người gọi ngoài cuộc gọi hiện tại, là

	RF XGBoost NN RNN SVM LR						
Tất cả các	0,9984	0,9979	0,9977 0,9974 0,9913		0,9846 0,9967 0,9803		
Trên cùng-10	0,9965	0,9979	0,9905 0,9784				

BẢNG VI: Điểm AUC của các mô hình khác nhau bằng cách sử dụng tất cả các loại máy bay và chỉ 10 tính năng hàng đầu từ Hình 11.

được sử dụng nhiều hơn ở cấp thứ ba, nhưng ít hơn ở cấp cao nhất. Những trường hợp ngoại lệ như vậy là rất ít.

Chúng tôi quan sát thấy rằng chỉ có 21 trong số 29 đối tượng địa lý (tức là 70%) được sử dụng trong mô hình rừng ngẫu nhiên. Nếu chúng ta cắt đuôi dài bằng cách chỉ sử dụng các tính năng được sử dụng hơn 10 lần, thì chỉ 10 tính năng hàng đầu là cần thiết. Bây giờ chúng tôi kiểm tra xem liệu 10 tính năng hàng đầu được sử dụng trong rừng ngẫu nhiên này có đủ dấu hiệu cho vấn đề dự đoán cuộc gọi độc hại hay không. Đặc biệt, chúng tôi sử dụng thiết lập tương tự như phân tích cắt bỏ của chúng tôi và đánh giá hiệu suất của các mô hình khác nhau chỉ bằng 10 tính năng này.

Kết quả được trình bày trong Bảng VI. Chúng ta có thể quan sát thấy rằng, hiệu suất của việc triển khai XGBoost không hề thay đổi. Điều này đặc biệt bởi vì 10 tính năng hàng đầu được chọn để tạo hương vị cho việc triển khai XGBoost. Chúng tôi quan sát thấy rằng điểm AUC của tất cả các mô hình khác giảm từ 0,001 đến 0,01. Sự xuống cấp như vậy là không đáng kể và điểm AUC của 3 phương pháp tiếp cận tốt nhất vẫn trên 0,9965. Do đó, chúng tôi kết luận rằng bằng cách phân tích một mô hình rừng ngẫu nhiên, chúng tôi có thể tìm ra các đặc điểm tiêu biểu nhất để giải thích hiệu suất của các mô hình học máy.

VIII. THẢO LUẬN

Một hạn chế trong công việc của chúng tôi là nó không thể xử lý hiệu quả việc giả mạo người gọi. Đây là kết quả vì chúng tôi đã tập trung vào các phương pháp tiếp cận danh sách đen để chặn các cuộc gọi độc hại dựa trên các số. Chúng tôi coi việc giảm thiểu vấn đề này là một hướng đi quan trọng trong tương lai.

Ngoài ra, như đã đề cập trước đó, hệ thống của chúng tôi hiện không thể phân biệt rất rõ giữa những người gọi lừa đảo hoặc spam và những người gọi liên quan đến bán hàng. Như đã trình bày trong nghiên cứu của chúng tôi, thời gian hoạt động và liệu một số được lưu trong liên hệ của người dùng TouchPal có thể được sử dụng làm các tính năng để phân biệt như vậy hay không. Chúng tôi có kế hoạch điều tra các vấn đề liên quan trong tương lai.

Tuy nhiên, cách tiếp cận của chúng tôi không tuân theo những kẻ tấn công có thể muốn đưa vào danh sách trắng một số cuộc gọi độc hại cụ thể.
Như đã giải thích trong Phần III, TouchPal sử dụng cơ chế danh sách đen và do đó, miễn là có đủ người dùng lành tính gắn thẻ một số cuộc gọi độc hại cụ thể, nó sẽ được gắn nhãn là độc hại bất kể những kẻ tấn công có nỗ lực như thế nào. Do đó, cách tiếp cận của chúng tôi không bị những kẻ tấn công đầu độc dữ liệu, những kẻ cố gắng thao túng dữ liệu đào tạo để làm cho mô hình dự đoán một số cuộc gọi độc hại là lành tính.

Tuy nhiên, cách tiếp cận máy học của chúng tôi vẫn có thể bị hai loại kẻ tấn công máy học. Đầu tiên, mặc dù những kẻ tấn công không thể liệt kê các số điện thoại độc hại, chúng có thể sử dụng các cuộc tấn công độc hại để liệt kê đen các số lành tính bằng cách thiết lập một trang trại để gắn thẻ các số lành tính là độc hại. Thứ hai, phương pháp tiếp cận máy học được đề xuất của chúng tôi có thể dễ bị những kẻ tấn công trốn tránh, những người thao túng dữ liệu thử nghiệm trong quá trình thời gian phục vụ mô hình. Đặc biệt, có một số tính năng, chẳng hạn như gap to next, có thể bị kẻ tấn công cố tình thao túng để bắt chước hành vi của người gọi lành tính. Chúng tôi xem xét việc giảm thiểu những vấn đề này trong tương lai.

IX. CÔNG TRÌNH LIÊN QUAN

A. Các kỹ thuật phát hiện cuộc gọi độc hại hiện có Đã có

nhiều nghiên cứu trước đây thảo luận về việc phát hiện cuộc gọi độc hại, chẳng hạn như danh sách trắng / đen [17], [25], [35], [39] và danh tiếng miền của người gọi [25], [32]. Công việc của chúng tôi sử dụng các phương pháp học máy để đạt được giải pháp chính xác.

Các công trình hiện có cũng thiết kế phương pháp tiếp cận phát hiện cuộc gọi mali dựa trên máy học, dựa trên hành vi của người gọi [30], [35], hành vi của người nhận [20], [39], kết nối xã hội [6], [8], [21], [27], [28], và phản hồi của khách hàng [17], [18], [32], [37] - [39], [41]. Tuy nhiên, tất cả các công việc này giả định một máy chủ trong mạng điện thoại có thể cung cấp thêm thông tin về người gọi. Ngược lại, công việc của chúng tôi là phương pháp tiếp cận dựa trên máy học đầu tiên mà không dựa trên bất kỳ giả định nào về các mạng điện thoại bên dưới.

B. Phân tích cuộc gọi độc hại qua điện thoại

Đã có một loạt các nghiên cứu có hệ thống trong phân tích cuộc gọi độc hại. Ví dụ: [15] xây dựng một honeypot với 39.696 số điện thoại bị bỏ rơi vì chủ cũ nhận được quá nhiều cuộc gọi không mong muốn. Các cuộc gọi đến các số điện thoại này được coi là cuộc gọi độc hại và được phân tích. Đối với các cuộc gọi lừa đảo có mục tiêu hơn, chẳng hạn như lừa đảo hỗ trợ kỹ thuật, công việc hiện tại [22] thực hiện một nghiên cứu có hệ thống về cả các trò lừa đảo và các trung tâm cuộc gọi đằng sau chúng. Những công việc này thường yêu cầu ghi âm và phân tích nội dung thoại của các cuộc gọi đến, điều này có thể phá vỡ quyền riêng tư của người dùng. Phân tích của chúng tôi hoàn toàn không đụng đến nội dung cuộc gọi của người dùng và do đó loại bỏ quyền riêng tư này

mối quan tâm.

Một số công trình mô tả hệ sinh thái thư rác qua điện thoại và đánh giá mức độ chuyên nghiệp cao cho các kỹ thuật hiện có [29], [33]. Các công trình này nêu bật các yêu cầu về thiết kế các phương pháp tiếp cận ngăn chặn cuộc gọi độc hại hiệu quả, trong khi công việc của chúng tôi cung cấp một giải pháp cụ thể.

X. KẾT LUẬN

Trong công trình này, chúng tôi trình bày giải pháp dựa trên máy học đầu tiên mà không dựa trên bất kỳ giả định cụ thể nào về cơ sở hạ tầng mạng điện thoại bên dưới. Chúng tôi đề xuất một số kỹ thuật để đạt được mục tiêu. Đầu tiên, chúng tôi thiết kế giao diện người dùng TouchPal như một thành phần của Ứng dụng dành cho thiết bị di động để cho phép người dùng điện thoại gắn nhãn các cuộc gọi độc hại. Sau đó, chúng tôi tiến hành một nghiên cứu đo lường quy mô lớn trong ba tháng nhật ký cuộc gọi, bao gồm 9 tỷ bản ghi và thiết kế các tính năng dựa trên kết quả. Chúng tôi đánh giá rộng rãi các phương pháp học máy hiện đại khác nhau bằng cách sử dụng 29 tính năng được đề xuất và kết quả cho thấy rằng phương pháp tốt nhất có thể giảm tới 90% các cuộc gọi độc hại không thể phát hiện trong khi vẫn duy trì độ chính xác hơn 99,99% so với lưu lượng cuộc gọi lành tính. Kết quả cũng cho thấy các mô hình có thể đưa ra các dự đoán một cách hiệu quả và do đó có thể được triển khai trên thực tế.

NHÌN NHẬN

Chúng tôi cảm ơn những người đánh giá ẩn danh và người chăn cừu Matt Fredrikson của chúng tôi vì những ý kiến đóng góp quý báu của họ để cải thiện bài báo. Chúng tôi cảm ơn Xiaojing Liao về những cuộc thảo luận hữu ích. Công việc này được hỗ trợ một phần bởi FORCES (Founda tions Of Resilient CybEr-Physical Systems), tổ chức nhận được hỗ trợ từ National Science Foundation (số giải thưởng NSF là CNS-1238959, CNS-1238962, CNS-1239054, CNS 1239166), DARPA dưới sự tài trợ không. FA8750-17-2-0091, Berke ley Deep Drive và Trung tâm An ninh mạng dài hạn, NSFC (61632017, 61772333) và Chương trình Thuyền buồm Thượng Hải (17YF1428200).

Mọi ý kiến, phát hiện và kết luận hoặc khuyến nghị được trình bày trong tài liệu này là của (các) tác giả và không nhất thiết phản ánh quan điểm của National Science Foundation.

NGƯỜI GIỚI THIỆU

[1] "Đây bao nhiêu điện thoại năm lừa đảo Giá cả ameri lon ngoái ... " https://www.marketwatch.com/story/ heres-how-much-phone-scams-cost-americans-last-year-2017-04-19, truy cập: 2017-11-29.

[2] "Một trong những công ty lớn nhất châu Âu mất 40 triệu euro trong vụ lừa đảo trực tuyến," https://news.slashdot.org/story/16/08/31/2336207/ one-of-europes- lớn-công-ty-phát-cao-mô-hình-chống-thư-rác-trong-các-hệ-thống-voip quy mô lớn," trong Dis accessed: 2017-11-29. Hệ thống máy tính cống hiến, 2008. cả khi họ không trả lời Hội nghị về. IEEE, 2008, trang 85–92. gọi, "http://www.independent.co.uk/life-style/gadgets-and-tech/news/[26] F. Pedregosa, G. Varoquaux, A. ICDCS'08. Quốc tế thứ 28 [3] "Lừa đảo qua điện thoại khiến nạn nhân thiệt hại 300 ngay chí-khi-họ-không-trả lời-cuộc gọi-a7003486. Gramfort, V. Michel, B. Thirion, điện thoại-lừa đảo-chi phí-nạn nhân-300-thậm

truy cập: 2017-11-29. plas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot và E. Duch [4] "Bắt sóng tự viễn thông lớn của Trung Quốc để hủy bỏ dịch vụ chuyển vùng và cuộc gọi đường dài kể từ ngày 1 tháng 9 (bằng tiếng Trung Quốc), "http: // Nghiên cứu Học tập, tập 12, trang 2825–2830, 2011. www.gov.cn/xinwen/2017-09/01/content O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot và E. Duch html, trong nước," Scikit-learn: Machine learning trong Python," Tạp chí Phí máy dịch vụ hút và vấn đề thư rác, "11-29. trong Kỷ yếu hội thảo bảo mật VoIP lần thứ 2, 2005 5222006.htm, truy cập: 2017- [27] Y. Rebahi và D. Sisalem, "Các nhà cung cấp

[5] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, GS Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu và X. Zheng, "TensorFlow: Học máy quy mô lớn trên các hệ thống không đồng nhất," 2015, phần mềm có sẵn từ tensorflow.org. [Trực tuyến]. Có tại: https://www.tensorflow.org/

[6] MA Azad và R. Morla, "Phát hiện vết nhổ nhiều tầng khi chuyển tiếp qua lại," trong Phần mềm, Viễn thông và Mạng Máy tính (SoftCOM), Hội nghị Quốc tế lần thứ 19 năm 2011 về. IEEE, 2011, trang 1-9.

[7] — — , "Người đại diện: Phát hiện cuộc gọi không mong muốn với sức mạnh xã hội của người gọi," Máy tính & Bảo mật, tập. 39, trang 219–236, 2013.

[8] — — , "Người đại diện: Phát hiện cuộc gọi không mong muốn với sức mạnh xã hội của người gọi," Máy tính & Bảo mật, tập. 39, trang 219–236, 2013.

[9] M. Balduzzi, P. Gupta, L. Gu, D. Gao và M. Ahamad, "Mobipot: Hiểu được các mối đe dọa từ điện thoại di động với thẻ mật ong," trong Kỷ yếu của Hội nghị ACM lần thứ 11 về An ninh Máy tính và Truyền thông Châu Á . ACM, 2016, trang 723–734.

[10] L. Breiman, "Rừng ngẫu nhiên", Máy học, tập. 45, không. 1, pp. 5–32, 2001.

[11] N. Chaisamran, T. Okuda, G. Blanc và S. Yamaguchi, "Phát hiện spam voip dựa trên niềm tin dựa trên thời lượng cuộc gọi và mối quan hệ giữa con người," trong Ứng dụng và Internet (SAINT), 2011 IEEE / IPSJ 11 Hội nghị chuyên đề liên quốc gia Bật. IEEE, 2011, trang 451–456.

[12] T. Chen và C. Guestrin, "Xgboost: Một hệ thống thúc đẩy cây có thể mở rộng," trong Kỷ yếu của Hội nghị Quốc tế ACM SIGKDD 22Nd về Khám phá Kiến thức và Khai thác Dữ liệu. ACM, 2016, trang 785–794.

[13] C. Cortes và V. Vapnik, "Máy vectơ hỗ trợ," Máy học, tập. 20, không. 3, trang 273–297, 1995.

[14] DA Freedman, Các mô hình thống kê: lý thuyết và thực hành. cambridge báo chí đại học, 2009.

[15] P. Gupta, B. Srinivasan, V. Balasubramaniyan và M. Ahamad, "Phoneypot: Hiểu biết theo hướng dữ liệu về các mối đe dọa từ điện thoại". trong NDSS, 2015.

[16] S. Hochreiter và J. Schmidhuber, "Trí nhớ ngắn hạn dài", Neural Tính toán.

[17] N. Jiang, Y. Jin, A. Skudlark, W.-L. Hsu, G. Jacobson, S. Prakasam, và Z.-L. Zhang, "Cô lập và phân tích các hoạt động gian lận trong một mạng di động lớn thông qua phân tích đồ thị cuộc gọi thoại," trong Kỷ yếu của hội nghị quốc tế lần thứ 10 về các hệ thống, ứng dụng và dịch vụ di động. ACM, 2012, trang 253–266.

[18] N. Jiang, Y. Jin, A. Skudlark, và Z.-L. Zhang, "Greystar: Phát hiện nhanh chóng và chính xác các số spam sms trong các mạng di động lớn bằng cách sử dụng không gian điện thoại màu xám." trong Hội nghị chuyên đề về bảo mật USENIX, 2013, trang 1–16.

[19] P. Kolan và R. Dantu, "Bảo vệ kỹ thuật - xã hội chống lại việc nhập thư rác bằng giọng nói," Giao dịch ACM trên các hệ thống tự trị và thích ứng (TAAS), vol. 2, không. 1, tr. 2 năm 2007.

[20] P. Kolan, R. Dantu, và JW Cangussu, "Mức độ phiền toái của cuộc gọi thoại," Giao dịch ACM trên Máy tính, Truyền thông và Ứng dụng Đa phương tiện (TOMM), vol. 5, không. 1, tr. 6 năm 2008.

[21] A. Leontjeva, M. Goldszmidt, Y. Xie, F. Yu và M. Abadi, "Phân loại người dùng skype bí mật ban đầu thông qua máy học," trong Kỷ yếu Hội thảo ACM 2013 về Trí tuệ nhân tạo và Bảo mật, ser. AISec '13, 2013.

[22] N. Miramirkhani, O. Starov và N. Nikiforakis, "Quay số một để lừa đảo: Phân tích và phát hiện các trò gian lận hỗ trợ kỹ thuật," trong NDSS, 2017.

[23] NB của Thống kê Trung Quốc, "Dữ liệu quốc gia - khu vực - hàng quý bởi tỉnh -2016, "2016.

[24] P. Patankar, G. Nam, G. Kesidis, và CR Das, "Khám phá các mô hình chống thư rác trong hệ thống voip quy mô lớn," trong Hệ thống máy tính phân tán, 2008. ICDCS'08. Hội nghị quốc tế lần thứ 28 về. IEEE, 2008, trang 85–92.

[28] Y. Rebahi, D. Sisalem, và T. MageDanz, "Phát hiện thư rác," trong Viễn thông Kỹ thuật số, 2006. ICDT'06. Hội nghị quốc tế về. IEEE, 2006, trang 68–68.

[29] M. Sahin, A. Francillon, P. Gupta và M. Ahamad, "Sok: Gian lận trong mạng điện thoại. "

[30] C. Sorge và J. Seedorf, "Một hệ thống uy tín cấp nhà cung cấp để đánh giá chất lượng của các thuật toán giảm thiểu khạc nhổ," trong Communications, 2009. ICC'09. Hội nghị quốc tế IEEE về. IEEE, 2009, trang 1–6.

[31] Y. Soupionis và D. Gritzalis, "Aspf: Khung dựa trên chính sách chống phỉ báng thích ứng," trong Tính khả dụng, Độ tin cậy và Bảo mật (ARES), Hội nghị Quốc tế lần thứ sáu năm 2011 về. IEEE, 2011, trang 153–160.

[32] K. Srivastava và HG Schulzrinne, "Ngăn chặn thư rác dựa trên nền tảng nhâm nhi các phiên và tin nhắn tức thì, "2004.

[33] H. Tu, A. Doupe, Z. Zhao, và G.-J. Ahn, "Sok: Mọi người đều ghét cuộc gọi tự động: Một cuộc khảo sát về các kỹ thuật chống spam điện thoại," Hội nghị chuyên đề IEEE về Bảo mật và Quyền riêng tư (SP) 2016, trang 320–338, 2016.

[34] F. Wang, Y. Mo, và B. Huang, "P2p-avs: Tính năng lọc thư rác voip hợp tác dựa trên P2p," Hội nghị Mạng và Truyền thông Không dây IEEE 2007, trang 3547–3552, 2007.

[35] — — , "P2p-avs: Bộ lọc thư rác voip hợp tác dựa trên P2p," trong Hội nghị Mạng và Truyền thông Không dây, 2007. WCNC 2007. IEEE. IEEE, 2007, trang 3547–3552.

[36] F. Wang, FR Wang, B. Huang, và LT Yang, "Advs: một mô hình dựa trên danh tiếng về việc lọc nhổ qua mạng p2p-voip," Tạp chí Siêu máy tính, trang 1–18, 2013.

[37] — — , "Khuyến cáo: một mô hình dựa trên danh tiếng về việc lọc phần mềm qua mạng p2p-voip," Tạp chí Siêu máy tính, trang 1–18, 2013.

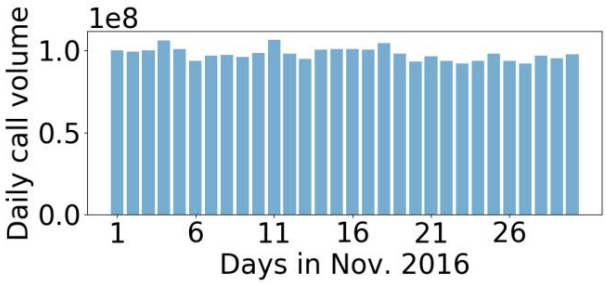
[38] Y.-S. Wu, S. Bagchi, N. Singh và R. Wita, "Phát hiện spam trong các cuộc gọi qua ip thoại thông qua phân nhóm bán giám sát" trong Dependable Sys tems & Networks, 2009. DSN'09. Hội nghị quốc tế IEEE / IFIP về. IEEE, 2009, trang 307–316.

[39] G. Zhang và S. Fischer-Hubner, "Phát hiện gần như trùng lặp trong hộp thư thoại bằng cách sử dụng hàm băm". trong ISC. Springer, 2011, trang 152–167.

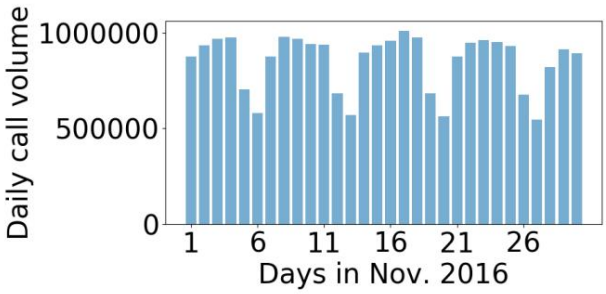
[40] R. Zhang và A. Gurtov, “Lọc thư rác bằng giọng nói hợp tác dựa trên danh tiếng,” trong Ứng dụng Hệ thống Cơ sở dữ liệu và Chuyên gia, 2009. DEXA'09. Hội thảo quốc tế lần thứ 20 về. IEEE, 2009, trang 33–37. [41] – – , “Lọc thư rác bằng giọng nói hợp tác dựa trên danh tiếng,” trong Ứng dụng Hệ thống Cơ sở dữ liệu và Chuyên gia, 2009. DEXA'09. Hội thảo quốc tế lần thứ 20 về. IEEE, 2009, trang 33–37.

RUỘT THỪA

Phân bổ lưu lượng cuộc gọi hàng ngày trong tháng 11 và tháng 12 năm 2016 lần lượt được trình bày trong Hình 12 và Hình 13.



(a) Cuộc gọi bình thường



(b) Cuộc gọi độc hại

Hình 12: Biểu đồ các cuộc gọi bình thường và cuộc gọi độc hại vào tháng 11 năm 2016.

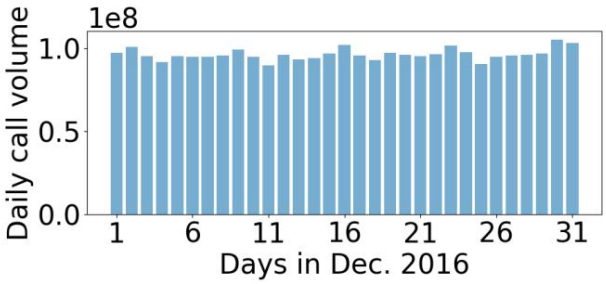
Giải thích mô hình mạng nơron.

Chúng tôi bắt đầu với trường hợp đơn giản khi các đối tượng địa lý lịch sử không được sử dụng. Chúng tôi sử dụng mạng chuyển tiếp hai lớp. Lưu ý rằng tất cả các đối tượng địa lý số được mã hóa dưới dạng vectơ một nóng. Nghĩa là, đối với một đối tượng nhận giá trị v, giả sử phạm vi của đối tượng là [0, max], thì một vectơ nóng của nó là vectơ chiều (tối đa + 1), trong đó thứ nguyên (v + 1) là 1 và tất cả các thứ nguyên khác là 0. Một mã hóa nóng tính năng số đầu vào là một thực tế phổ biến khi sử dụng mạng thần kinh.

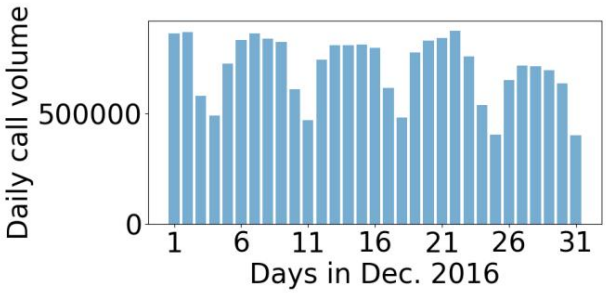
Các trạng thái ẩn chứa 20 tế bào thần kinh. Đầu ra, là một vectơ hai chiều, được kết nối với toán tử softmax để tính toán dự đoán cuối cùng. Về mặt hình thức, dự đoán có thể được viết là

$$p = \text{softmax}\ W1 \times \text{ReLU}\ (W2 \times x)$$

trong đó W1 là ma trận 2 × 20, W2 là ma trận 20 × n, ReLU là hàm chỉnh lưu tiêu chuẩn và n là kích thước đặc tính đầu vào. p là một vectơ hai chiều, trong đó p1 chỉ ra



(a) Cuộc gọi bình thường



(b) Cuộc gọi độc hại

Hình 13: Biểu đồ các cuộc gọi bình thường và cuộc gọi độc hại vào tháng 12 năm 2016.

xác suất cuộc gọi đến là cuộc gọi độc hại và p0 + p1 = 1 tương ứng với thuộc tính của softmax. Bằng cách đặt ngưỡng mô hình τ mô hình học máy có thể dự đoán cuộc gọi đến không phải là cuộc gọi độc hại nếu p1 ≥ τ. Bằng cách điều chỉnh ngưỡng mô hình, một mô hình được đào tạo có thể tạo ra sự cân bằng giữa độ chính xác và khả năng thu hồi.

Để tính đến các đối tượng địa lý lịch sử, một cách dễ hiểu là coi các đối tượng địa lý của mỗi bản ghi như một vectơ và tính giá trị trung bình của các vectơ đối tượng địa lý cho tất cả các bản ghi lịch sử dưới dạng một vectơ đối tượng địa lý lịch sử có độ dài cố định. Vectơ đặc trưng lịch sử này sau đó được nối với vectơ đặc trưng cho cuộc gọi hiện tại, vectơ này trở thành đầu vào cho mạng nơ-ron. Chúng tôi gọi cách tiếp cận này là cách tiếp cận NN vani.

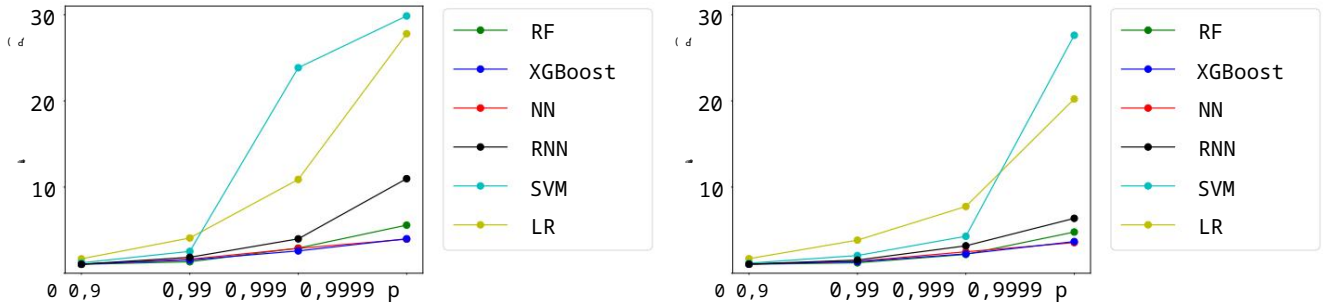
Tuy nhiên, lấy giá trị trung bình có thể không phải là cách hiệu quả nhất để tận dụng thông tin từ nhật ký cuộc gọi. Chúng ta có thể coi các bản ghi cuộc gọi lịch sử từ callee là một chuỗi có độ dài thay đổi. Do đó, chúng ta có thể sử dụng mạng nơ-ron tái phát (RNN) [16] để chuyển đổi chuỗi thành một nhúng có độ dài cố định. Đặc biệt, chúng tôi sử dụng một LSTM [16] với kích thước trạng thái ẩn 16 để tính toán việc nhúng, sau đó được nối với vectơ đặc trưng cho cuộc gọi hiện tại. Các tính năng kết hợp sau đó được đưa vào mạng nơ-ron ở trên để đưa ra dự đoán. Chúng tôi gọi cách tiếp cận này là một cách tiếp cận dựa trên RNN.

Giải thích về máy học mạng phi thần kinh al gorithm.

Mặc dù các phương pháp tiếp cận mạng nơ-ron đã đạt được những tiến bộ đáng kể để xử lý dữ liệu chiều cao, một số

		RF XGBoost NN RNN SVM LR					
Các tỉnh lớn	Quảng Đông 0,9979 0,9970 0,9969 0,9961 0,9893 0,9776 Thượng Hải 0,9979 0,9969						
	0,9969 0,9961 0,9892 0,9793 Tứ Xuyên 0,9987 0,9983 0,9982 0,9978 0,9926 0,9866						
	Chiết giang	0,9984	0,9978 0,9976 0,9972 0,9922 0,9847				
Các tỉnh nhỏ	Cát Lâm	0,9973 0,9964 0,9961 0,9955 0,9874 0,9713					
	Quý Châu 0,9987 0,9982 0,9981 0,9979 0,9941 0,9865						
	An Huy	0,9986 0,9979 0,9978 0,9975 0,9927 0,9845					

BẢNG VII: Bảng này trình bày điểm AUC của các mẫu xe khác nhau khi được huấn luyện trên nhật ký cuộc gọi của Bắc Kinh từ tháng 10 đến tháng 11 và được kiểm tra trên nhật ký cuộc gọi của các tỉnh khác nhau vào tháng 12. "Các tỉnh lớn" chỉ ra rằng top 5 tỉnh có số lượng cuộc gọi độc hại lớn nhất; "Các tỉnh nhỏ" cho biết ba tỉnh có ít cuộc gọi ác ý nhất.



(a) Dữ liệu thử nghiệm: Chiết Giang

(b) Dữ liệu thử nghiệm: Quý Châu

Hình 14: AFP @ (30, p) kết quả cho các mô hình khác nhau được đào tạo bằng cách sử dụng nhật ký cuộc gọi của Bắc Kinh.

các phương pháp tiếp cận mạng phi nơron vẫn hiệu quả hơn trong việc hạn chế các đầu vào có chiều thấp. Đặc biệt, đầu vào của chúng tôi chỉ bao gồm 29 tính năng, và do đó chúng tôi muốn kiểm tra xem liệu các phương pháp tiếp cận mạng phi thần kinh này có hiệu quả hơn các phương pháp mạng thần kinh hay không. Đặc biệt, chúng tôi quan tâm đến (1) mô hình rừng ngẫu nhiên [10]; (2) Hỗ trợ Vector Machine (SVM) mod els [13]; và (3) mô hình hồi quy logistic [14]. Chúng tôi giải thích ngắn gọn các mô hình này dưới đây.

Mô hình rừng ngẫu nhiên. Rừng ngẫu nhiên là một tập hợp các cây quyết định. Mỗi cây quyết định là một cây mà mỗi nút bên trong gắn nhãn một tính năng và một ngưỡng. Khi đưa ra dự đoán, mô hình cây quyết định đi ngang cây từ gốc đến lá và xác định di chuyển sang trái hoặc phải tùy thuộc vào việc giá trị của đối tượng đầu vào có nhãn trên nút có nhỏ hơn ngưỡng hay không. Mỗi lá được liên kết với một giá trị thực và giá trị trên lá ở cuối quá trình truyền tải được trả về dưới dạng đầu ra của nó. Mô hình rừng ngẫu nhiên đưa ra quyết định bằng cách lấy trung bình tất cả các giá trị được tính toán từ mỗi cây quyết định trong rừng để nhận được giá trị cuối cùng là p. Một lần nữa, quyết định có thể được thực hiện bằng cách đặt ngưỡng τ theo cách tương tự như cách tiếp cận trước đó đã

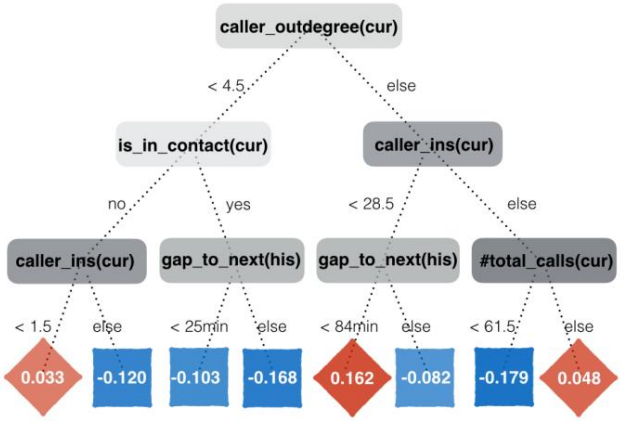
Các mô hình SVM. Mô hình SVM được thiết kế để xử lý vấn đề khi dữ liệu huấn luyện không thể phân tách tuyến tính. Cụ thể, nó sử dụng một ánh xạ φ do người dùng chỉ định để ánh xạ đối tượng đầu vào vào một không gian chiều cao, và sau đó đào tạo một mô hình y = w · φ (x) + b, sao cho mặt phẳng quyết định được xác định bởi w và b tối đa hóa lợi nhuận, đồng thời cho phép một số dữ liệu đào tạo bị phân loại sai. Thông thường, φ được cung cấp dưới dạng một hàm nhân κ, sao cho κ (x, x0) = φ (x) · φ (Dự đoán x cũng0Các được thực hiện bằng cách sử dụng trực tiếp hàm κ. Trong trường hợp của chúng ta, chúng ta sử dụng hàm nhân tuyến tính để đào tạo mô hình SVM. Lưu ý rằng SVM cũng phát ra một giá trị thực, có thể được sử dụng để dự đoán.

Các mô hình logistic. Mô hình logistic có thể được coi là một mạng nơron một lớp: p = σ (wx + b), trong đó σ là hàm sigmoid. Mô hình này được áp dụng phổ biến trong các ứng dụng công nghiệp do tính đơn giản và hiệu quả của nó.
Tuy nhiên, nó có thể không hiệu quả bằng các lựa chọn thay thế khác. Đầu ra p có thể được sử dụng để đưa ra dự đoán.

Lưu ý rằng tất cả các mô hình này sử dụng đầu vào có độ dài cố định. Do đó, chúng tôi sử dụng cùng một phương pháp trong phương pháp vani NN để tính toán một lần nhúng lịch sử cho tất cả các bản ghi lịch sử.

Kết quả đánh giá chi tiết về khả năng tổng quát hóa đến các vị trí khác.

Đặc biệt, chúng tôi xây dựng dữ liệu đào tạo bằng nhật ký cuộc gọi tháng 10 và tháng 11 năm 2016 của Bắc Kinh. Chúng tôi chọn 7 tỉnh khác, 4 tỉnh có lưu lượng cuộc gọi lớn và 3 tỉnh nhỏ, và sử dụng nhật ký cuộc gọi tháng 12 năm 2016 của họ để xây dựng 7 bộ thử nghiệm tương ứng. Chúng tôi đào tạo mô hình bằng cách sử dụng cùng một bộ đào tạo và đánh giá chúng trên 7 bộ thử nghiệm khác nhau tương ứng. Kết quả AUC được báo cáo trong Bảng VII. Chúng tôi quan sát thấy rằng hiệu suất của các mô hình khác nhau phù hợp với



Hình 15: Một cây quyết định ví dụ trong mô hình rừng ngẫu nhiên tốt nhất được đào tạo bằng XGBoost. Mỗi nút bên trong chỉ ra một tính năng cần được kiểm tra. Một trong những cạnh đang đi xuống của nó được gắn nhãn với điều kiện kiểm tra, trong khi cạnh kia được gắn nhãn "khác". Mỗi nút lá được liên kết với một giá trị.

các thí nghiệm. Mỗi mô hình được đào tạo bằng cách sử dụng dữ liệu của Bắc Kinh có thể đạt được hiệu suất AUC tương đương với mô hình được đào tạo bằng chính tỉnh thử nghiệm, điều này cho thấy rằng mô hình thực sự có thể tổng quát hóa thành các con số không nhìn thấy từ một vị trí khác.

Bằng cách xem xét kỹ, điều thú vị là chúng tôi nhận thấy rằng khi mô hình được đào tạo sử dụng dữ liệu từ Bắc Kinh, điểm AUC của nó ở một tỉnh khác thậm chí còn cao hơn một chút so với mô hình được đào tạo bằng dữ liệu từ chính tỉnh được thử nghiệm. Ví dụ: điểm AUC của RF ở Quảng Đông là 0,9979 khi được đào tạo sử dụng dữ liệu của Bắc Kinh, trong khi giá trị là 0,9978 khi mô hình được đào tạo bằng dữ liệu của Quảng Đông. Vì Bắc Kinh có số lượng bản ghi cuộc gọi độc hại lớn nhất, điều này cho thấy rằng một tập hợp đào tạo lớn hơn có thể giúp cải thiện hiệu suất.

Chúng tôi trình bày kết quả AFP @ (M, p) cho M = 30 và các bộ thử nghiệm được xây dựng bằng bản ghi nhật ký cuộc gọi từ Chiết Giang (khối lượng cuộc gọi lớn) và từ Quý Châu (khối lượng cuộc gọi nhỏ) trong Hình 14. Chúng tôi thực hiện các quan sát tương tự như các thí nghiệm trước đó : (1) cấp bậc của các mô hình khác nhau đối với AFP @ (M, p) nói chung là nhất quán với các quan sát trước đó; (2) mô hình rừng ngẫu nhiên và giá trị AFP @ p của mô hình NN đều dưới 5.

Hình dung cây quyết định. Trong Hình 15, chúng ta hình dung một cây quyết định trong khu rừng ngẫu nhiên được đào tạo bằng XGBoost trên dữ liệu của tháng 10 và tháng 11 ở Quảng Châu.