Phương pháp tiếp cận dựa trên tính năng và dựa trên đồ thị để Sự cố phát hiện cuộc gọi spam

TS Ngọc C Lê và các đồng chí

Trường Toán ứng dụng và Tin học
Đại học Khoa học và Công Nghệ Hà Nội
1 Đại Cồ Việt, Hà Nội

Tóm tắt — Thư rác qua điện thoại Internet (SPIT) ngày càng trở nên nghiêm trọng và đã thu hút sự chú ý đáng kể từ các nhà cung cấp viễn thông do tác hại lớn của nó đối với tài chính và người dùng ' trải qua. Đã có một lời kêu gọi khẩn cấp về các cơ chế để ngăn chặn những kẻ tấn công SPIT gian lận và các hoạt động khó chịu.
Sau khi nghiên cứu kỹ lưỡng các công trình của các nhà nghiên cứu trước đây, có hai cách tiếp cận ấn tượng là dựa trên đặc điểm và dựa trên biểu đồ truyền cảm hứng cho chúng tôi để phát triển một cơ chế có thẩm quyền hơn. Với dựa trên tính năng, chúng tôi sẽ thiết lập một tập hợp 11 tính năng mà có sự phân biệt quan trọng trong việc phân phối giữa các SPITers và người dùng hợp pháp. Mặt khác, dựa trên đồ thị, dựa trên nền tảng của thuật toán CallRank trong [1], chúng tôi tiến hành sửa đổi theo cách chúng tôi tính trọng số của các cạnh liên kết hai người dùng. Triển khai trên tập dữ liệu được thu thập thông tin từ iCaller - một ứng dụng chặn cuộc gọi spam, chúng tôi sẽ đề xuất các kết quả có thể chấp nhận được để làm cơ sở cho các phương pháp luận của chúng tôi. Một số kết luận được cho là sẽ được đưa ra trong phần cuối cùng.

Điều khoản lập chỉ mục — Thư rác qua điện thoại Internet (SPIT) · Giao thức thoại qua Internet (VoIP) · Dựa trên tính năng · Dựa trên đồ thị · Phương pháp học tập có giám sát · ROC

I. GIỚI THIỆU

Trong những năm tới, cùng với sự phát triển nhanh chóng của công nghệ Internet để liên lạc bằng giọng nói, điện thoại VoIP đã có sự gia tăng đáng kể về số lượng thuê bao.
Tuy nhiên, điều này được kết hợp bởi Thư rác qua điện thoại Internet (SPIT), một hình thức lạm dụng và gian lận đang gia tăng tương ứng trên toàn thế giới. SPIT là một trong những mối đe dọa nghiêm trọng nhất trong VoIP do tổn thất tài chính nghiêm trọng và kích thích tinh thần do nó gây ra cho người dùng điện thoại. Hãy xem xét công nghệ pro tocol khởi tạo phiên (SIP), công nghệ được áp dụng rộng rãi nhất trong giao thức báo hiệu, như một minh họa, những kẻ tấn công có thể dễ dàng thực hiện các cuộc gọi rác đến hàng nghìn địa chỉ IP của người dùng với sự trợ giúp của chi phí thấp và thuận tiện trong việc sử dụng SIP.

Do sự cần thiết của việc ngăn chặn những kẻ gửi thư rác lạm dụng, trên nền tảng của kinh nghiệm trong quá trình bảo vệ SPAM, rất nhiều cách tiếp cận để giải quyết vấn đề đã được đề xuất. Chúng bao gồm các yếu tố đáng chú ý như dựa trên danh tiếng [2], dựa trên tần suất cuộc gọi [3], danh sách ngược động, lấy dấu vân tay [4], làm rõ các cuộc gọi đáng ngờ bằng capchas [5] và việc sử dụng một số thuật toán tinh gọn máy như máy vectơ hỗ trợ [6], [7] và phân cụm bán giám sát [8].
Sau khi xem xét kỹ lưỡng, chúng tôi đánh giá cao hai cách tiếp cận hiệu quả là dựa trên tính năng và dựa trên đồ thị để khám phá SPIT tiềm năng. Bằng một số sửa đổi của các công trình trước đây và sự đa dạng của các thuật toán học máy, chúng tôi sẽ thiết kế một hệ thống phát hiện SPIT thông qua các phương pháp tiếp cận dựa trên tính năng và dựa trên đồ thị.

Những đóng góp quan trọng của chúng tôi trong bài báo này có thể được liệt kê là theo sau

- Chúng tôi xem xét 11 tính năng nổi bật được sửa đổi cẩn thận và được chứng minh là đáng chú ý. • Chúng tôi triển khai tập dữ liệu của mình với 8 máy học khác nhau trong các thuật toán, tiết lộ các yếu tố cơ bản để đánh giá nhằm tìm ra thuật toán nào được coi là trạng thái của nghệ thuật.
- Chúng tôi cung cấp cái nhìn sâu sắc hơn về phương pháp tiếp cận dựa trên đồ thị bằng cách phát triển CallRank [1] với một số sửa đổi và cung cấp bằng chứng rằng phương pháp luận của chúng tôi là rất tốt hiệu quả.

Phần còn lại của bài báo của chúng tôi được tổ chức như sau. Phần 2 được sử dụng để tóm tắt các công trình liên quan về cách tiếp cận dựa trên đặc điểm và cách tiếp cận dựa trên đồ thị. Trong phần 3, chúng tôi sẽ cung cấp chi tiết về dữ liệu được sử dụng cho đào tạo, phương pháp luận và các yếu tố đánh giá. Tiếp theo là phần 4 do chúng tôi thực hiện và một số đánh giá về kết quả so với các nhà nghiên cứu trước đây cũng như một số hạn chế còn tồn tại. Một số kết luận được đưa ra trong phần 5.

II. CÔNG TRÌNH LIÊN QUAN

A. Cơ chế phát hiện SPIT

SPIT trở thành mối nguy hiểm cao độ đối với khách hàng VoIP do chi phí cuộc gọi thoại đi xuống, trái ngược với

nền tảng hiện tại và sự thiếu vắng
và hệ thống hành chính. Nó ảnh hưởng đến sự tồn tại riêng tư của khách hàng và các thư từ của anh ta và biến thành một ổ hỗn loạn đối với họ. Mối đe dọa đó là nguồn cảm hứng cho các nhà nghiên cứu đưa ra nhiều cơ chế để xác định SPIT. Việc xem xét kỹ lưỡng cho phép chúng tôi tổng hợp với hai cơ chế nổi bật là Dựa trên hành vi và Xếp hạng cuộc gọi.

1) Dựa trên hành vi: Trong [9], Kusumoto et al. đưa vào ac
đếm một số mẫu cuộc gọi bao gồm tỷ lệ thực hiện cuộc gọi không thành công, thời lượng cuộc gọi trung bình, mối quan hệ người dùng. Sau quá trình tính toán các mẫu khác nhau đó, Naive Bayes được đưa vào kích hoạt để xác định những kẻ gửi thư rác. Bộ sưu tập của trước đây
tiêu chí được sử dụng có thể được liệt kê như sau.

- Tỷ lệ cuộc gọi đã trả lời [10], [11]
- Tỷ lệ cuộc gọi bị loại bỏ [7], [9], [10], [11] • Số lần thực hiện cuộc gọi [7], [8], [10], [11] • Thời lượng cuộc gọi [7], [8], [1], [9], [10], [11]

Cần lưu ý rằng mô hình thời lượng cuộc gọi được hầu hết các nhà nghiên cứu tận dụng. Kết hợp với các hành vi phổ biến từ những kẻ tấn công SPIT, chúng tôi phát triển một bộ tiêu chí có các chi tiết được đưa ra trong tiểu mục 3.2.

Để xác định xem một số điện thoại có phải là SPIT tiềm năng hay không, chúng tôi đã tiến hành thử nghiệm với 8 thuật toán học tập có giám sát được xem xét trong tiểu mục 2.2.

2) CallRank: Một cách nổi tiếng để tiếp cận vấn đề này là các tác giả tập trung vào điểm danh tiếng dựa trên nguyên tắc tính toán điểm danh tiếng. Khái niệm sử dụng điểm danh tiếng cũng nằm trong các công trình [12], [13], [14] ai
đã nỗ lực thiết lập một chuỗi tin cậy giữa người gọi và
callee. Chúng tôi đánh giá cao CallRank [1] vì nhóm tác giả đã tiếp cận vấn đề khá hiệu quả bằng phương pháp dựa trên đồ thị. Ý tưởng cơ bản trong [1] là xây dựng mạng xã hội
mối quan hệ và danh tiếng toàn cầu đối với người dùng. Dựa trên đó, chúng tôi tính toán điểm tác động của một khách hàng trong mạng. Chúng tôi kỳ vọng rằng những người dùng hợp pháp có thể có điểm ảnh hưởng cao hơn những kẻ tấn công SPIT, do đó có thể phát hiện ra những kẻ gửi thư rác.

B. Thuật toán học máy

Khi nói đến việc thực hiện các nhiệm vụ khó trong nhân tạo trong viễn thông, so với các thuật toán truyền thống, các thuật toán học máy có hiệu quả cao hơn về độ chính xác, hiệu suất thời gian. Công việc của chúng tôi đặt ưu tiên vào việc học tập tích cực. Trong học tập có giám sát, trên cơ sở các lớp học đã được xác định trước, một máy tiến hóa trong quá trình học tập và phân loại theo một mô hình phân loại. Hãy để chúng tôi mô tả một số thuật toán phân loại được sử dụng trong
khuôn khổ của chúng tôi. Các thuật toán đã chọn của chúng tôi là hồi quy Logistic, k-vùng lân cận gần nhất, Naive Bayes, Cây quyết định, Rừng ngẫu nhiên, Bagging, AdaBoost và XGBoost.

III. PHƯƠNG PHÁP NGHIÊN CỨU

A. Dataset - Ứng dụng iCaller

Chúng tôi thu thập dữ liệu từ iCaller trong khoảng thời gian từ năm 2018 đến năm 2021. iCaller là một ứng dụng đang được phát triển bởi Grooo International Jsc. Mục đích đằng sau sự phát triển của ứng dụng này là để ngăn chặn các cuộc gọi có liên quan đến gian lận, cho vay, nợ, quảng cáo, real_estate và các hoạt động kích động khác. Hiện tại, nó đang hoạt động dựa trên các báo cáo trực tiếp từ người dùng trên toàn thế giới. Cụ thể hơn, giả sử người dùng A vừa nhận được một cuộc gọi làm phiền từ người dùng B, người đã quảng cáo về quyền lợi bảo hiểm của họ. Sau cuộc gọi đó, A

cho rằng B là người gửi thư rác, sau đó A đánh dấu số điện thoại của B là người gửi thư rác quảng cáo trong giao diện người dùng ứng dụng một cách dễ dàng. Kể từ đó, bất cứ khi nào B bắt đầu cuộc gọi đến A, ứng dụng sẽ đẩy một cảnh báo về danh tính của B trên điện thoại của A và hỏi liệu A có muốn trả lời cuộc gọi đó hay không.
Để loại bỏ tất cả các hạn chế, chúng tôi có ý định
đó là áp dụng các thuật toán học máy trong hệ thống để kiểm tra tự động. Hơn nữa, phương pháp luận của chúng tôi có khả năng tăng độ chính xác trong phân loại vì trên thực tế, rất có thể người dùng đánh dấu nhầm người gọi là người gửi thư rác. Chúng tôi cũng nhấn mạnh vào việc làm rõ danh tính của người dùng sẽ giúp những người nhận cuộc gọi quyết định có trả lời cuộc gọi hay không. Hơn nữa, iCaller không có

tiếp cận cơ sở hạ tầng viễn thông của viễn thông
nhà cung cấp, điều này làm cho các phương pháp SPIT trước đây trở nên bất lực trong việc sử dụng thực tế. Tuy nhiên, chúng tôi có một giải pháp sử dụng máy chủ để thu thập chéo giữa các số điện thoại trong mạng, giúp chúng tôi thu thập và lưu trữ dữ liệu cần thiết để thực hiện các phương pháp mà chúng tôi sẽ đề xuất trong bài báo này.

Như tôi đã đề cập ở trên, iCaller cho phép người dùng đánh dấu la
tin vào số điện thoại dựa trên 6 loại người gửi thư rác được báo cáo là report_advertising, report_loan, report_debt, report_cheat, report_real_estate, report_other và trong trường hợp xác nhận một người dùng hợp pháp, report_not_spam được đưa vào sử dụng (xem hình 2). Hãy để chúng tôi tiết lộ mô tả tập dữ liệu của chúng tôi được trích xuất từ cơ sở dữ liệu của iCaller. Đối với mỗi cuộc gọi tương ứng với mỗi bản ghi, chúng tôi thu thập dữ liệu về cuộc gọi đó, loại người dùng là thành viên, người gửi thư rác hoặc không được gắn nhãn. iCaller không có
tiếp cận cơ sở hạ tầng viễn thông của viễn thông
nhà cung cấp, điều này làm cho các phương pháp SPIT trước đây trở nên bất lực trong việc sử dụng thực tế. Tuy nhiên, chúng tôi có một giải pháp sử dụng máy chủ

	member_phone	phone	type	time	in_contact	duration
299058	8.456921e+10	84939619767	1	11/28/2020 6:47	0	49
689753	8.433392e+10	84989230242	3	1/1/2021 9:45	1	36
476633	8.489882e+10	8.429E+11	1	12/5/2020 11:00	0	8
650656	8.498912e+10	84898424194	1	1/3/2021 11:57	1	56
512472	8.498333e+10	84981596541	3	12/13/2020 9:48	1	0
450309	8.497851e+10	84963059108	1	12/18/2020 14:31	1	15
595627	8.492794e+10	84922627759	1	12/26/2020 17:51	1	28
288596	8.498764e+10	84982830201	2	11/30/2020 13:43	1	14
449939	8.434348e+10	84963087963	1	11/17/2020 5:40	1	7
805678	8.493353e+10	84939289988	2	1/2/2021 9:03	0	51

Hình 1: Mẫu tập dữ liệu

	phone	from_contact	report_advertising	report_loan	report_debt	report_cheat	report_real_estate	report_other
36552	84975503535	1	0	0	0	0	4	1
93735	84900	1	1	1	0	3	5	2
93757	84965303997	1	0	0	0	0	0	7
93776	84394955554	0	0	2	2	0	1	2
93849	8418001090	1	1	0	1	0	0	1
...	...	...	...	...	...	...	...	...
914974	84931141036	0	3	0	0	0	0	0
924098	84877283378	0	1	0	2	0	0	0
928861	84932710282	1	0	0	2	1	0	0
928908	84932779367	1	0	2	2	2	2	0
1039300	84899909039	1	4	0	0	0	0	0

Hình 2: Mẫu tập dữ liệu

để thu thập chéo giữa các số điện thoại trong mạng,
giúp chúng tôi thu thập và lưu trữ dữ liệu cần thiết để triển khai
các phương pháp chúng tôi sẽ đề xuất trong bài báo này. Nhớ lại rằng A gọi
B. Thông tin đó có thể được giải thích như trong bảng I.

B. Dựa trên tính năng

1) Tìm hiểu các tính năng: Chúng tôi sẽ giải thích 5 tiêu chí để
phát hiện SPIT được nghiên cứu bởi ba nhà viễn thông lớn nhất
nhà cung cấp ở việt nam

- Tần suất cuộc gọi: số lượng cuộc gọi bắt đầu từ một cuộc gọi đã thuê
  điện thoại trong một khoảng thời gian. Ví dụ: tối thiểu 200
  các cuộc gọi đi mỗi ngày từ 8 giờ sáng đến 6 giờ chiều.
- Tỷ lệ cuộc gọi có thời lượng ngắn: Ví dụ:
  hơn 80% cuộc gọi có thời lượng dưới 25 giây trong
  khoảng thời gian.
- Tỷ lệ cuộc gọi có khoảng thời gian ngắn giữa các cuộc gọi:
  Đó là tỷ lệ các cuộc gọi có khoảng thời gian ngắn giữa
  hai cuộc gọi liên tiếp trên tổng số cuộc gọi. Vì
  ví dụ, hơn 50% cuộc gọi có ít hơn 20
  giây giữa các cuộc gọi.
- Tỷ lệ cuộc gọi đến các thuê bao không liên quan: Tỷ lệ cuộc gọi
  đến các số không liên quan (chưa từng gọi trước đây, không liên lạc
  danh sách) trên tổng số cuộc gọi đi. Thí dụ:
  90% các số được gọi là khác nhau, không lặp lại.
- Đặc điểm hành vi: Số điện thoại chủ yếu là
  được sử dụng cho cuộc gọi đi, không nhận và gửi SMS.

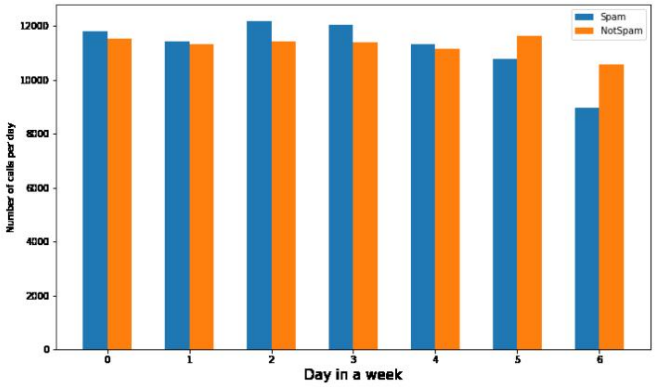
Cột	Giá trị	Nghĩa
điện thoại		số điện thoại
from_contact	0	số điện thoại đó là mới người sử dụng
	1	số điện thoại đó đã được sử dụng từ liên hệ của một thành viên danh sách
report_loan	0	không phải Loan Spam
	1	Loan báo Spam
report_advertising	0	không phải là Spam quảng cáo
	1	Spam quảng cáo được báo cáo
report_debt	0	không phải Thư rác nợ
	1	Thư rác Nợ được báo cáo
report_cheat	0	không gian lận Spam
	1	Đã báo cáo Spam gian lận
report_real_estate	0	không phải Real_estate Spam
	1	Real_estate Spam được báo cáo
report_other	0	không có loại thư rác nào khác
	1	loại thư rác khác được báo cáo
thời gian		khi cuộc gọi xảy ra
in_contact	0	Số của B nằm trong số của A danh sách liên hệ
	1	Số của B không bằng số A danh sách liên hệ
khoảng thời gian		thời lượng của cuộc gọi
loại hình	1	đã nhận thành công cuộc gọi của thành viên
	2	cuộc gọi bắt đầu thành công của thành viên
	3	cuộc gọi đã nhận không thành công thành viên
	4	cuộc gọi bắt đầu không thành công thành viên

BẢNG I: Trích xuất thông tin về cuộc gọi và người dùng

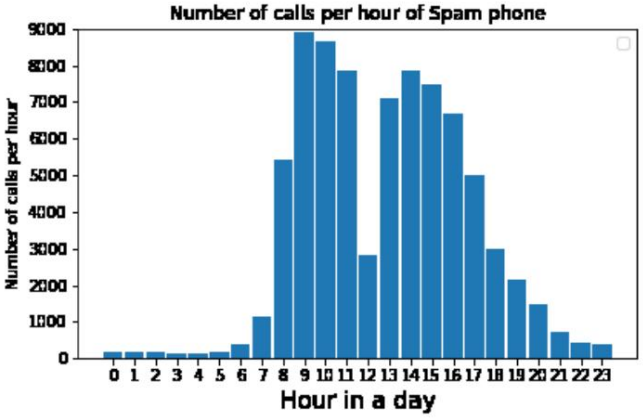
Tuy nhiên, đối với các lĩnh vực cụ thể, cần có một số điều chỉnh
ments cho phù hợp với đặc điểm của khu vực đó. Kết hợp
với các tiêu chí như vậy được sử dụng bởi các nhà nghiên cứu cũ, chúng tôi phát triển một bộ
các tiêu chí được chứng minh là hữu ích sau khi
kiểm tra lại. Ở đây, chúng tôi sẽ hiển thị các biểu đồ để cung cấp thêm kiến thức
về các tính năng đã chọn của chúng tôi.

So với các nghiên cứu khác, phân phối thời gian của chúng tôi
Theo ngày, không có sự khác biệt nào nổi bật giữa người dùng Spam và người
dùng hợp pháp (xem hình 3). Kết quả là chúng tôi đã thử

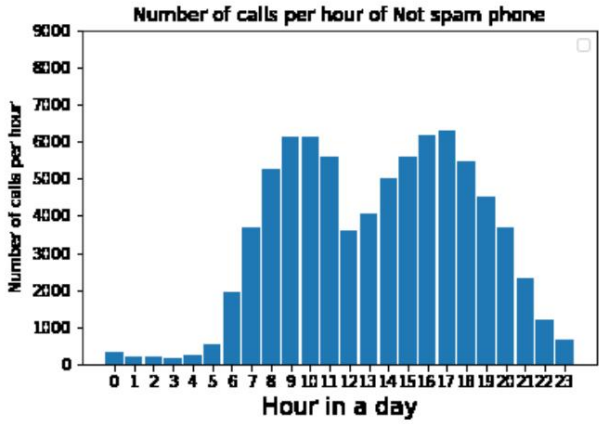
để tìm ra một phân phối thời gian khác và đưa ra tính năng liên quan đến số lượng cuộc gọi mỗi giờ.



Hình 3: Số lượng cuộc gọi mỗi ngày giữa Spam và notSpam



(a) Số lượng cuộc gọi mỗi giờ của Người gửi thư rác

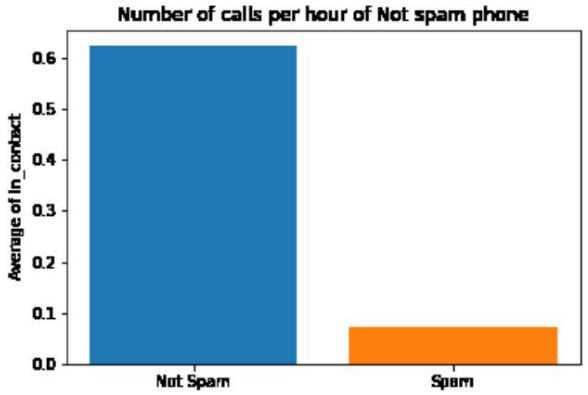


(b) Số lượng cuộc gọi mỗi giờ của người dùng hợp pháp

Hình 4: in_hour

Hình 4 cho thấy sự khác biệt đáng chú ý trong việc phân phối tính năng "in_hour" thể hiện số lượng cuộc gọi của những người gửi thư rác trong giờ làm việc cao hơn so với các cuộc gọi của các thành viên. Tại Việt Nam, hầu hết các công ty đều có lịch làm việc bắt đầu từ khoảng 7 giờ sáng đến 5 giờ chiều. Tuy nhiên, có thể nhận thấy rằng có sự chênh lệch đáng kể về số lượng cuộc gọi giữa người gửi thư rác và thành viên trong khoảng thời gian từ 5 giờ chiều - 6 giờ chiều. Đây có lẽ là chính sách làm thêm giờ của một số doanh nghiệp. Do đó, chúng tôi quyết định chọn phạm vi của tính năng "in_hour" từ 7 giờ sáng đến 6 giờ chiều

Chúng tôi thực sự tin rằng hầu hết những người gửi thư rác đều sử dụng số điện thoại của họ cho công việc, ngoài ra, các số được lưu trữ trong danh bạ của thành viên có thể có mối quan hệ với thành viên đó và do đó không có khả năng là người gửi thư rác, hãy cân nhắc bằng cách gán biến nhị phân [0, 1], trong đó 0 có nghĩa là số này không có trong danh bạ của thành viên, 1 thì ngược lại.
Tuy nhiên, chúng tôi nhận thấy rằng có một số số SPIT tồn tại trong danh bạ của thành viên, chúng tôi phỏng đoán rằng đã có trường hợp thành viên thực hiện hành động thêm liên hệ đó với mục đích ghi nhớ những người gửi thư rác, hoặc đó là quan hệ giữa người bán - người mua. Để làm cho việc quan sát có ý nghĩa hơn, chúng tôi đã tính trung bình trên tổng số cuộc gọi.
Chúng tôi cũng quan sát thấy rằng tỷ lệ số điện thoại có trong danh bạ của các thành viên hợp pháp cao hơn rõ rệt so với số điện thoại của những người gửi thư rác (xem hình), đây là một tiêu chí tốt trong mô hình phân loại.



Hình 5: Số điện thoại trong danh bạ của thành viên và người gửi thư rác

Bộ gồm 11 tính năng được lấy mẫu trong hình 6 và trưng bày trong bảng II.

call_to	call_in	call_to_miss	call_in_miss	duration_call_to	duration_call_in
2.708050	3.688879	2.708050	3.332205	6.364751	7.257708
0.693147	0.000000	0.000000	0.000000	2.846133	0.000000
6.819970	4.982439	7.075262	4.447482	10.374554	8.600068
3.367296	3.091042	3.135494	3.044522	6.695799	6.364751
3.238244	1.441367	2.852481	1.945910	7.252362	6.192523
3.688879	3.610918	3.401197	2.564949	7.471363	6.916715
0.000000	0.000000	0.693147	0.000000	0.000000	0.000000
1.032810	0.000000	1.742602	0.000000	0.580658	0.000000
0.000000	2.197225	0.693147	3.761200	0.000000	3.178054
3.076744	0.000000	2.629382	0.672943	7.912739	0.000000

(một)

avg_duration_call_to	avg_duration_call_in	avg_in_contact	in_hour	avg_success
3.077967	3.112840	0.512951	4.343805	0.376686
2.846039	0.000000	0.000000	0.693147	0.693147
2.798686	3.207819	0.051535	7.508758	0.358812
2.842579	2.717757	0.171272	3.931826	0.416394
3.570678	3.979767	0.032387	3.700606	0.428670
3.289278	3.091040	0.144250	4.394449	0.539715
0.000000	0.000000	0.000000	0.000000	0.000000
0.098667	0.000000	0.000000	0.112489	0.065802
0.000000	0.378436	0.000000	3.951244	0.000000
4.416036	0.000000	0.000000	3.568280	0.511333

(b)

Hình 6: Tập dữ liệu mẫu gồm 11 tính năng

Tính năng	Nghĩa
gọi tới	số lượng cuộc gọi do khách hàng bắt đầu
call_in	số lượng cuộc gọi mà khách hàng nhận được
call_to_miss	số cuộc gọi nhỡ do khách hàng bắt đầu
call_in_miss	số cuộc gọi nhỡ khách hàng nhận được
thời lượng_call_to	tổng thời lượng do khách hàng bắt đầu
thời lượng_call_in	tổng thời lượng khách hàng nhận được
avg_duration_call_to	lượng thời lượng trung bình do khách hàng bắt đầu
avg_duration_call_in	lượng thời lượng trung bình mà khách hàng nhận được
avg_incontact	số điện thoại trung bình được lưu trữ trong danh bạ của thành viên
trong giờ	tỷ lệ cuộc gọi được thực hiện trong giờ làm việc (7 giờ sáng - 6 giờ tối)
avg_success	tỷ lệ cuộc gọi thành công

BẢNG II: 11 tính năng

2) Thuật toán và số liệu đánh giá: Phát hiện SPIT là một vấn đề phân loại với một tham số nhị phân trong đó 0 có nghĩa là notspam và 1 có nghĩa là spam. Để phân loại, chúng tôi sử dụng 11 tính năng ở trên và 8 thuật toán phân loại cụ thể là: Hồi quy logistic, k-lân cận gần nhất, Naive Bayes, Cây quyết định, Rừng ngẫu nhiên, Đóng túi, AdaBoost và XGBoost. Với XGBoost, thủ thư được tích hợp sẵn trong xgboost được sử dụng, trong khi chúng tôi sử dụng thư viện sklearn cho các thuật toán khác.

Để ước tính hiệu quả của mô hình, chúng tôi sử dụng k-fold kỹ thuật xác nhận chéo, thường được sử dụng để so sánh và chọn mô hình tốt nhất cho một vấn đề. Kỹ thuật này dễ hiểu, dễ thực hiện và tạo ra sự tự tin đáng tin cậy hơn khoảng thời gian hơn các phương pháp khác.

Đối với quá trình đánh giá, chúng tôi sẽ tính đến độ chính xác, độ chính xác, thu hồi và AUC. Trong bài báo này, chúng tôi chỉ có thể giải thích ngắn gọn về các yếu tố đánh giá này. Accu racy chỉ định một phép đo cho mức độ gần với giá trị được chấp nhận hoặc giá trị thực, trong khi độ chính xác có nghĩa là các phép đo gần như thế nào của cùng một mục là với nhau. Độ chính xác tự chủ từ độ chính xác. Điều đó báo hiệu độ chính xác cao có thể không dẫn đến độ chính xác cao, trong một số trường hợp, thậm chí độ chính xác thấp. Và, độ chính xác cao với độ chính xác thấp cũng có thể xảy ra. Như một

kết quả là quan sát khoa học có chất lượng được đánh giá cao hơn là điểm F1, là trung bình hài hòa của độ chính xác và nhớ lại với giả định rằng giá trị này không bằng 0. Chúng tôi cũng nhấn mạnh vào ROC - một đường cong biểu thị hiệu suất của một mô hình phân loại. các ngưỡng phân loại. AUC (Khu vực
dưới đường cong) cũng có thể được sử dụng như một hệ số hiệu quả.

C. Dựa trên đồ thị

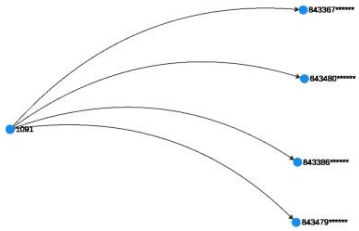
Dữ liệu cần thiết cũng được thu thập từ cơ sở dữ liệu của ứng dụng iCaller. Với dựa trên biểu đồ, chúng tôi hướng đến việc tính toán điểm danh tiếng cho mỗi người dùng, sau đó đánh giá để xác định xem người dùng đó có phải là người gửi thư rác hay không là đơn giản. Phương pháp của chúng tôi bao gồm hai bước cơ bản như sau

- Đầu tiên, chúng tôi xây dựng một biểu đồ đại diện cho một mạng có mười bảy người dùng trong đó mỗi người dùng được đại diện bởi một nút và mối quan hệ gọi giữa những người dùng được minh họa bằng một vòng cung. Mỗi cung có trọng số bằng tổng thời lượng giữa những người dùng.
- Sau đó, với sự trợ giúp của Eigentrust, bằng cách tính toán giá trị trung tâm làm chỉ số danh tiếng, chúng tôi đi đến kết luận rằng nhóm người dùng đáng tin cậy có điểm số cao hơn nhiều so với nhóm người dùng spam.

Hãy để chúng tôi giải thích lý do tại sao chúng tôi xây dựng cơ chế xoay quanh thời lượng cuộc gọi. Phương pháp luận của chúng tôi được lấy cảm hứng từ một nhận thức đơn giản rằng một khách hàng hợp pháp thường có một số lượng lớn các cuộc gọi kéo dài trong khoảng thời gian dài. Mặt khác, một người gửi thư rác / nhân viên bán hàng có khả năng cố gắng tiếp cận nhiều cá nhân nhất có thể bằng cách thực hiện một số lượng lớn các cuộc gọi ngắn vừa phải. Người gửi thư rác có thể thường xuyên không nhận được cuộc gọi đến hoặc một số lượng rất nhỏ cuộc gọi. Sự khác biệt trong các mẫu cuộc gọi là, đối với người gửi thư rác, mẫu cuộc gọi, ở một mức độ lớn, là một chiều trong khi nó là hai chiều đối với các khách hàng hợp pháp. Chúng tôi khai thác sự khác biệt này trong các mẫu cuộc gọi và sử dụng thời lượng cuộc gọi để thực hiện các công nhận đáng kể như xác nhận của một
mức độ tin cậy đã hiểu.

Cho đồ thị G = (V, A) đại diện cho mạng xã hội giữa những người sử dụng điện thoại trong đó V là tập đỉnh, mỗi v trong V đại diện cho một thành viên. Đương nhiên, A được gọi là tập các cung.
Nhớ lại rằng (u, v) A nếu và chỉ khi u thực hiện cuộc gọi đến v. Để minh họa, trong hình 7, số điện thoại 1091 đã từng thực hiện cuộc gọi đến bốn số điện thoại khác và các cuộc gọi đó đã được trả lời, do đó tồn tại các cung liên kết số điện thoại 1091 đến bốn số đó. Biểu thị wuv là trọng số của cung (u, v), sau đó chúng tôi đưa ra tỷ lệ tin cậy cục bộ chuẩn hóa có công thức
như sau

$$cuv = \frac{duv}{P\ j\ duj}$$



Hình 7: Mạng người dùng hợp pháp và gửi thư rác

với duv là tổng thời lượng cuộc gọi mà người dùng bạn đã thực hiện cho người dùng v. cuv đóng vai trò chính trong phương pháp tính trọng số vòng cung của chúng tôi. Điều đáng chú ý là khi nói đến một người dùng hoàn toàn mới, họ chưa thực hiện cuộc gọi, do đó, tổng thời lượng cuộc gọi do họ thực hiện là 0. Để tránh trường hợp chia cho không , nếu tồn tại một người dùng mà tôi chưa thực hiện bất kỳ cuộc gọi nào cho bất kỳ người dùng nào gần đó, chúng tôi cần trọng số ban đầu trong 1 cij = $\frac{}{|V|}$ đó | V | là số đỉnh.

Sau khi xây dựng biểu đồ đại diện cho mạng lưới giữa những người dùng, chúng tôi sử dụng Eigentrust, đây là thước đo trọng tâm để đánh giá uy tín của người dùng trên mạng xã hội. Điều này dựa trên giả định rằng mức độ hợp pháp của mỗi người dùng được xác định bởi những người dùng lân cận trong mạng.

Giả sử rằng chúng ta có m đồng nghiệp trong mạng xã hội.
Chúng tôi sẽ hiển thị một số biểu tượng đóng góp vào thuật toán Eigentrust.

Biểu tượng	Sự mô tả
t	Véc tơ của điểm danh tiếng
C	Giá trị tin cậy địa phương
e	phân phối ban đầu

Vectơ phân phối ban đầu e có đơn vị 1 chuẩn và
thành phần được xác định bởi ei = $\frac{1}{m}$. Do đó, để tính toán giá trị danh tiếng của người dùng được biểu thị bằng , chúng tôi tính toán vectơ - t phân phối tĩnh - t bằng cách giải phương trình sau

$$- \quad t = C^{T\ n} \quad - \ e$$

với n lớn là số lần lặp. Thuật toán được minh họa rõ ràng bằng thủ tục sau:

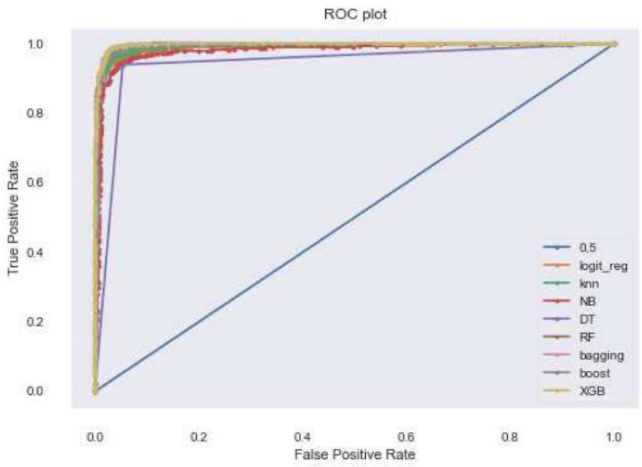
IV. THÍ NGHIỆM VÀ THẢO LUẬN

A. Dựa trên tính năng

Chúng tôi sử dụng các thuật toán học máy khác nhau để dự đoán SPITters trên tập dữ liệu gồm 4150 người gửi spam và 4150 thành viên hợp pháp. Sau đó, chúng tôi sử dụng kỹ thuật gấp k để so sánh và chọn mô hình có kết quả tốt nhất. Chúng tôi chia ngẫu nhiên tập dữ liệu thành 10 phần và huấn luyện trên 10 lần. Ở mỗi lần thử, chúng tôi sẽ chọn 1 phần làm dữ liệu xác nhận và phần còn lại làm dữ liệu tàu. Điều đó giúp chúng tôi đánh giá khác biệt và chính xác hơn.

	classifiers_name	Accuracy	F1 Score	Recall	Precision	AUC
7	XGBClassifier	0.975904	0.976038	0.979167	0.972930	0.996980
4	RandomForest	0.975502	0.975610	0.977564	0.973663	0.996540
5	BaggingClassifier	0.975502	0.975551	0.975160	0.975942	0.995786
6	AdaBoostClassifier	0.969076	0.969089	0.967147	0.971038	0.993122
0	LogisticRegression	0.952209	0.952533	0.956731	0.948372	0.987313
1	KNN	0.961446	0.961446	0.959135	0.963768	0.985764
2	Naive_bayes	0.932932	0.929447	0.881410	0.983021	0.978825
3	DecisionTree	0.942972	0.942880	0.939103	0.946688	0.942981

Hình 8: Kết quả của 8 thuật toán khác nhau



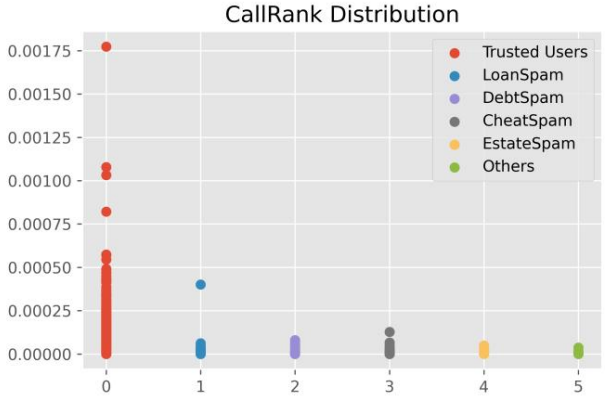
Hình 9: Triển lãm 8 thuật toán của ROC

Có thể dễ dàng nhận thấy trong hình 8 và 9 rằng XGB là hiện đại nhất, trong khi Rừng Ngẫu nhiên cũng giúp bạn tạo ra kết quả khá đặc biệt. Điều tự nhiên là các thuật toán sử dụng phương pháp học tập kết hợp luôn cho kết quả cao hơn các thuật toán khác. Cũng có lý khi các thuật toán Hồi quy tuyến tính, K-Nearest Neighborhood, Naive Bayes cho kết quả kém hơn so với các cách tiếp cận khác. Lý do cho điều này là vì các thuật toán này tiếp cận như một phân loại tuyến tính

và nó không phù hợp với tập dữ liệu của chúng tôi.

B. Dựa trên đồ thị

Sau quá trình tính toán, điểm danh tiếng của người dùng được cho trong hình 10



Hình 10: Phân phối CallRank

Một nhận xét dễ dàng là điểm danh tiếng của nhóm người dùng bắt chước hợp pháp cao hơn nhiều so với điểm của nhóm spammer. Bảng III cho thấy khoảng tin cậy 95% của mỗi nhóm.

Loại hình	Ràng buộc
Người dùng đáng tin cậy	18.0715 - 25.0715
LoanSpam	1,489 - 3,489
DebtSpam	0,3921 - 2,3921
CheatSpam	0,26 - 2,26
EstateSpam	0,1214 - 1,1214
Khác	0,2967 - 2,2967

BẢNG III: Khoảng tin cậy

Mặc dù kết quả thu được có thể chấp nhận được, nhưng phương pháp này có một hạn chế đáng chú ý. Một tiêu chí nổi bật để phát hiện SPITers là hầu hết SPITers có tỷ lệ call_in rất thấp call_to trong đó "call_in" là số cuộc gọi đã nhận và "call_in" là số lượng cuộc gọi đã bắt đầu. Tuy nhiên, trong cách tiếp cận dựa trên đồ thị được đề xuất, có sự thiếu hụt đơn đặt hàng của các nút đóng vai trò là một yếu tố khác để cân bằng mỗi cung giữa hai người dùng. Chúng tôi dự định giải quyết vấn đề này như một phần của công việc trong tương lai của chúng tôi.

V. KẾT LUẬN

Công việc này tập trung vào việc phát triển một cơ chế đối phó với những kẻ tấn công SPIT bằng hai cách tiếp cận: dựa trên tính năng và dựa trên đồ thị. Với tính năng dựa trên, tập dữ liệu được thu thập từ iCaller ứng dụng được sử dụng để trích xuất 11 tính năng hữu ích của các cuộc tấn công SPIT. Các tiêu chí đó phục vụ cho quá trình đào tạo tám

các thuật toán phân loại có giám sát (Hồi quy logistic, k lân cận gần nhất, Naive Bayes, Cây quyết định, Rừng ngẫu nhiên, Đóng túi, AdaBoost và XGBoost). Một phần y tạm thời của các thuật toán đó được tiến hành và thuật toán thuật toán hiện đại nhất, XGBoost, được tiết lộ dựa trên đánh giá của ROC đường cong. Với dựa trên đồ thị, chúng tôi áp dụng một số sửa đổi từ thuật toán CallRank và cũng cho kết quả khá đặc biệt Tuy · nhiên, vẫn còn thiếu thông tin về số lượng cuộc gọi từ mỗi nút. Do đó, trong công việc sau này, chúng tôi có ý định thêm một chức năng phạt để hiển thị tỷ lệ giữa cuộc gọi đã nhận và cuộc gọi bắt đầu ở mỗi nút để tăng độ chính xác của phương pháp của chúng tôi nói chung.

NGƯỜI GIỚI THIỆU

[1] V. Balasubramaniyan, M. Ahamad và H. Park, “Callrank: Chống lại sự phi báng bằng cách sử dụng thời lượng cuộc gọi, mạng xã hội và danh tiếng toàn cầu.” 01 năm 2007.

[2] P. Kolan và R. Dantu, “Bảo vệ kỹ thuật xã hội chống lại việc nhập thư rác bằng giọng nói,” TAAS, vol. 2, 03 năm 2007.

[3] D. Shin, J. Ahn, và C. Shim, “Đa cấp độ xám liên tục: Một thuật toán bảo vệ thư rác bằng giọng nói,” Network, IEEE, vol. 20, trang 18 - 24, 10 2006.

[4] H. Yan, K. Sripanidkulchai, HB Zhang, Z.-Y. Shae, và D. Saha, “Kết hợp việc lấy dấu vân tay tích cực vào các hệ thống ngăn ngừa khạc nhổ,” 2006.

[5] R. Schlegel, S. Niccolini, S. Tartarelli, và M. Brunner, “Khung phòng chống thư rác qua điện thoại internet (nhổ),” 01 2007, trang 1 - 6.

[6] M. Nassar, R. State, and O. Festor, “Giám sát lưu lượng truy cập bằng cách sử dụng máy vectơ hỗ trợ,” 01 2008.

[7] M. Nassar, O. Dabbebi, R. Badonnel, và O. Festor, “Quản lý rủi ro trong cơ sở hạ tầng voip sử dụng máy vectơ hỗ trợ,” 11 2010, trang 48 - 55.

[8] Y.-S. Wu, S. Bagchi, N. Singh và R. Wita, “Phát hiện spam trong cuộc gọi qua ip thoại thông qua phân nhóm bán giám sát,” 06 2009, trang 307–316.

[9] T. Kusumoto, E. Chen, và M. Itoh, “Sử dụng các mẫu cuộc gọi để phát hiện những người gọi liên lạc không mong muốn,” 07 2009, trang 64–70.

[10] RJ Ben Chikha, T. Abbes và A. Bouhoula, “Một thuật toán phát hiện khạc nhổ dựa trên hành vi cuộc gọi của người dùng,” trong Hội nghị Quốc tế lần thứ 21 về Phần mềm, Viễn thông và Mạng Máy tính năm 2013 - (SoftCOM 2013), 2013, pp. 1–5.

[11] R. jabeur ben chikha, T. Abbes, W. Chikha, và A. Bouhoula, “Cách tiếp cận dựa trên hành vi để phát hiện thư rác qua các cuộc tấn công điện thoại ip,” International Journal of Information Security, vol. Ngày 15, 03 năm 2015.

[12] Y. Rebahi và D. Sisalem, “Các nhà cung cấp dịch vụ hút và vấn đề thư rác,” trong Hội thảo thứ 2 về Bảo mật thoại qua IP, 2005.

[13] P. Patankar, G. Nam, G. Kesidis, và C. Das, “Khám phá các mô hình chống thư rác trong hệ thống voip quy mô lớn,” 07 2008, trang 85–92.

[14] Y. Soupionis và D. Gritzalis, “Aspf: Dựa trên chính sách chống phỉ báng thích ứng khuôn khổ,” 09 2011, trang 153 - 160.

[15] B. Mathieu, S. Niccolini và D. Sisalem, “Sdrs: Hệ thống phát hiện và phản ứng spam bằng giọng nói qua ip,” Bảo mật và Quyền riêng tư, IEEE, vol. 6, trang 52 - 59, 01/2009.

[16] J. Quittek, S. Niccolini, S. Tartarelli, M. Stiemerling, M. Brunner, và T. Ewald, “Phát hiện các cuộc gọi khạc nhổ bằng cách kiểm tra các mẫu giao tiếp của con người,” 06 2007, trang 1979–1984.