LỌC TIN NHẮN RÁC VỚI SPAM-ASSASSIN

Vũ Minh Tuấn[*], Đặng Đình Quân[*], Nguyễn Thanh Hà[+], Trần Quang Anh[#]
[*]Trường Đại học Hà Nội
[+]Sở Thông tin và Truyền thông Hà Nội
[#]Học viện Công nghệ Bưu Chính Viễn Thông

***Tóm tắt*:** Khi dịch vụ tin nhắn trên thiết bị di động trở nên phổ biến và ngày càng phát triển thì cũng là lúc tin nhắn rác tràn ngập các mạng viễn thông. Giống như thư rác, tin nhắn rác không chỉ gây phiền toái, khó chịu cho người sử dụng mà còn gia tăng áp lực lên hạ tầng viễn thông và được sử dụng như một công cụ để thực hiện việc lừa đảo, phát tán vi-rút, mã độc… Trên nền tảng công nghệ phát hiện thư rác, cụ thể là sử dụng SpamAssassin, nhóm nghiên cứu đã phát triển hệ thống phát hiện tin nhắn rác gồm có: Hệ thống sinh bộ luật và phần mềm trên di động. Kết quả thí nghiệm cho thấy đây là một cách tiếp cận khả thi với tỷ lệ phát hiện tin nhắn rác lên đến 94% trong khi tỷ lệ cảnh báo nhầm tin hợp lệ chỉ ở mức thấp hơn 0.15%.

Từ khóa: AntiSpam, Android App, SpamAssassin, Tin nhắn rác, Thư rác, SMS Spam.

I. MỞ ĐẦU

Trong thế giới phẳng ngày nay, ứng dụng công nghệ thông tin để có thể giao tiếp nhanh chóng, thuận tiện và hiệu quả là điều tất yếu. Cùng với điện thoại và thư điện tử, tin nhắn (SMS) đã tạo nên một cuộc cách mạng về trao đổi thông tin. Theo thống kê từ Open University (2014) dựa trên báo cáo của các hãng phân tích thị trường, trong năm 2014, 90% số người trên thế giới nhắn tin ít nhất một lần mỗi ngày. Mỗi tháng, trên 350 tỷ tin nhắn được gửi đi trên toàn thế giới. Người dùng nước Anh đã đóng góp khoảng 140 tỷ tin nhắn trong năm 2014 [1]. Con số này thậm chí còn cao hơn nữa ở những nước đang phát triển. Những thống kê trên cho thấy, mặc dù ra đời sau thư điện tử nhưng tin nhắn SMS lại đang phát triển rất nhanh và đang có xu hướng chiếm ưu thế do một số đặc điểm đặc thù.

Kể từ khi xuất hiện, thư rác đã trở thành vấn đề nan giải và phiền phức, làm đau đầu các nhà cung cấp dịch vụ thư điện tử cũng như người dùng trên toàn thế giới nói chung và Việt Nam nói riêng. Tuy nhiên, khi cuộc chiến chống thư rác vẫn đang hết sức quyết liệt và chưa có hồi kết thì một mối đe dọa mới đối với người dùng mạng viễn thông lại phát sinh và ngày càng nhức nhối. Đó là vấn nạn tin nhắn rác. Điều này cũng không quá khó giải thích. Với số lượng người dùng khổng lồ, liên tục gia tăng trên toàn thế giới, người dùng điện thoại đã và đang trở thành những "con mồi béo bở" để những kẻ phát tán tin nhắn rác, những nhà cung cấp dịch vụ quảng cáo, marketting tha hồ tấn công, lợi dụng để truyền tải thông tin, giới thiệu dịch vụ, sản phẩm. Theo thống kê của Trung tâm ứng cứu khẩn cấp máy tính Việt Nam (VNCERT), mỗi ngày có hàng triệu tin nhắn rác được gửi đi, tương đương với hàng triệu thuê bao bị tấn công bởi những thông tin không có giá trị; đồng thời gây ra áp lực rất lớn lên hạ tầng viễn thông di động trong nước.

Để giải quyết mối đe dọa trên, cần phải kết hợp rất nhiều biện pháp mới có thể xử lý triệt để như chính sách quản lý đối với các công ty viễn thông, các công ty cung cấp dịch vụ quảng cáo, nâng cao ý thức người dùng về bảo mật thông tin cá nhân (địa chỉ thư điện tử, số điện thoại) và các giải pháp kỹ thuật. Đối với giải pháp kỹ thuật, bài toán đặt ra cho các nhà khoa học là phát hiện và ngăn chặn việc phát tán thư rác hiệu quả và kịp thời.

Trong bài báo này, với kinh nghiệm và những kết quả nghiên cứu đã đạt được trong lĩnh vực phát

Tác giả liên lạc: Vũ Minh Tuấn,
email: minhtuan_fit@hanu.edu.vn
Đến tòa soạn: 10/11/2016, chỉnh sửa: 24/1/2016,
chấp nhận đăng: 24/1/2016

hiện và ngăn chặn thư rác, chúng tôi đề xuất một phương pháp học máy lọc tin nhắn rác sử dụng SpamAssassin. Từ phương pháp được đề xuất, một mô hình khép kín để huấn luyện, tạo ra các bộ luật đã được xây dựng, thực nghiệm và đánh giá kết quả. Từ đó, tạo tiền đề để thực hiện những nghiên cứu khác dựa trên công nghệ này.

Bài báo được trình bày với cấu trúc như sau: Phần II giới thiệu về các công nghệ lọc thư rác hiện tại, đồng thời phân tích sự tương đồng và khác biệt giữa thư rác và tin nhắn rác. Phần tiếp theo, chúng tôi trình bày về phương án sử dụng SpamAssassin, một giải pháp được sử dụng rộng rãi trong ngăn chặn thư rác, áp dụng cho việc lọc tin nhắn rác. Mô hình triển khai chi tiết cũng được đề cập trong phần này. Trong phần IV, trọng tâm của bài nghiên cứu xoay quanh thí nghiệm được thực hiện với tệp dữ liệu mẫu; kết quả thu được của thí nghiệm được phân tích để đánh giá chính xác về tiềm năng của phương pháp và mô hình triển khai. Cuối cùng là phần Thảo luận, tóm lược lại vấn đề nghiên cứu và đề cập đến hướng phát triển tiếp theo.

II. KIẾN THỨC NỀN TẢNG

A. So sánh thư điện tử và tin nhắn SMS

Ý tưởng của nghiên cứu này là ứng dụng công nghệ lọc thư rác vào tin nhắn rác. Chính vì vậy, việc hiểu rõ sự tương đồng và khác biệt giữa hai loại "spam" này rất quan trọng.

Về mục đích sử dụng, cả thư điện tử và tin nhắn đều là phương tiện trao đổi thông tin của người dùng. Tuy nhiên, do giao thức trao đổi dữ liệu không giống nhau nên mỗi loại đều mang những đặc điểm riêng. Thư điện tử đã và đang được sử dụng rộng rãi nhất do lịch sử hình thành sớm hơn và có những lợi thế nhất định khi được coi như một kênh giao tiếp chính thống bằng văn bản. Người dùng có thể gửi thư đến một nhóm rất nhiều người dùng khác một cách dễ dàng mà không hề phải trả thêm phí dịch vụ. Thư điện tử gửi đi có thể được công nhận như một loại bằng chứng để lấy căn cứ làm việc.

Tuy nhiên, điểm bất lợi của tin nhắn so với thư điện tử lại nằm ở tính tiện dụng và tức thì. Người dùng có thể gửi tin nhắn SMS bất kỳ lúc nào với thiết bị di động cầm tay và tin nhắn đó sẽ được chuyển đến người nhận ngay lập tức. Việc này rất thuận tiện khi người nhận không có máy tính hoặc không có kết nối internet. Tuy nhiên, đây cũng chính là lý do khiến cho tin nhắn SMS nhanh chóng bị những kẻ phát tán tin nhắn rác lợi dụng để "tấn công" người sử dụng điện thoại di động. Do hệ thống phát hiện và ngăn chặn tin nhắn rác chưa phát triển như công nghệ chặn thư rác nên tin nhắn rác có khả năng đến được với người dùng cao hơn. Theo thống kê vào năm 2014 của tạp chí Business-2-Communication, tỷ lệ mở tin nhắn của người dùng là 98% trong khi tỷ lệ đó ở thư điện tử chỉ là 22% khi nhận được thông điệp từ các chiến dịch marketing trên di động [2].

Cấu trúc và dữ liệu của tin nhắn và thư điện tử cũng có những điểm tương đồng, khác biệt nhất định [3]. Điều này được thể hiện trong bảng 1.

Độ dài giới hạn của tin nhắn SMS là 160 ký tự, do đó, người dùng có xu hướng viết tắt để giảm phát sinh cước phí. Mỗi ngôn ngữ đều có những chữ viết tắt nhất định như: "*Anh*" => "*a*", "*Em*" =>"*e*", trong tiếng Việt hay "*Thank you*" => "*tks*", "*you*" => "*u*" trong tiếng Anh. Trong khi đó, thư điện tử không giới hạn về độ dài nên người gửi sẽ tự do và linh động hơn trong việc soạn thảo nội dung gửi đi. Phần cấu trúc thư điện tử và tin nhắn SMS tuy có đôi chút khác biệt (bảng 1) nhưng về cơ bản, các trường dữ liệu của cả hai đều tương đồng. Đây là cơ sở rất quan trọng để ứng dụng công nghệ phát hiện thư rác vào tin nhắn rác SMS.

Một điểm đáng quan tâm nữa khi so sánh thư điện tử và tin nhắn là phương thức truyền và lưu trữ thông tin. Thư điện tử được lưu trữ trên máy chủ; người dùng sử dụng các phần mềm như MS Outlook, Thunderbird… để tải về máy tính cá nhân hoặc thiết bị di động để đọc. Với cơ chế này, việc lọc thư rác có thể thực hiện rất thuận lợi trên chính máy chủ lưu trữ thư trong khi tin nhắn được chuyển trực tiếp về thiết bị di động của người dùng và lưu trữ trên đó nên công nghệ phát hiện và ngăn chặn tin nhắn rác sẽ gặp nhiều khó khăn hơn.

B. Công nghệ lọc tin nhắn rác SMS

Hiện tại, trên thế giới cũng đã có một số hướng nghiên cứu về công nghệ phát hiện tin nhắn rác. Tạp chí khoa học quốc tế "Expert System with Applications" số 39 (2012) có đăng một bài báo tương đối tổng quan về các phương pháp và cơ sở dữ liệu mẫu liên quan đến lọc tin nhắn rác của Sarah .D, Mark .B và Derek .G chủ yếu dựa trên nội dung tin nhắn [4]. Trong phần đầu bài, chúng tôi đánh giá sự tương đồng và khác biệt giữa thư điện tử và tin nhắn. Đây chính là cơ sở để phát triển những thuật toán lọc tin nhắn rác dựa trên lọc thư rác. Những nghiên cứu mới nhất về phương pháp phát hiện tin nhắn rác dựa trên nội dung đều được đề cập trong bài báo, cụ thể như: cơ chế phân loại với SVMs, k-NN hay Bayes… Có những nhà nghiên cứu đã khá thành công như Junaid và Farooq trong việc sử dụng các thuật toán tiến hóa để lọc tin nhắn rác [5]. Bên cạnh cách tiếp cận dựa trên nội dung tin nhắn, còn có một số phương pháp dựa trên các đặc điểm của tin nhắn như tần suất gửi tin, độ dài tin nhắn, khoảng cách giữa người gửi - người nhận… để phát hiện tin nhắn rác.

Trong báo cáo tại hội thảo SPSM'13, tác giả Akshay Narayan (2013) đã công bố kết quả thí nghiệm để đánh giá hiệu quả của một số phần mềm chặn tin nhắn rác trên thiết bị di động trên hệ điều hành Android [6]. Tất cả các ứng dụng được thử nghiệm đều hoạt động theo cơ chế xây dựng "danh sách đen" (blacklist) để nhận diện tin nhắn rác. Có một số ứng dụng chặn chính xác tin nhắn từ những số lạ nhưng không ổn định mà phụ thuộc vào tập dữ liệu kiểm thử.

Tại Việt Nam, hiện tại cũng đã có một số nhóm nghiên cứu và doanh nghiệp tập trung vào hướng nghiên cứu này. Năm 2013, Trung tâm Ứng cứu khẩn cấp máy tính Việt Nam (VNCERT) đã xây dựng đề án "Xây dựng hệ thống ngăn chặn, phòng chống thư rác". Tuy nhiên, mặc dù đề án bao gồm rất nhiều phân hệ khác nhau nhưng phần liên quan đến tin nhắn rác chỉ dừng lại ở việc điều phối và xử lý tin nhắn rác chứ không đề cập đến việc phát hiện tin nhắn rác; và đặc biệt là tin nhắn rác bằng Tiếng Việt.

Bảng 1. Bảng so sánh thư điện tử và tin nhắn SMS

Yếu tố so sánh	Thư điện tử	Tin nhắn SMS
Độ tài	Không giới hạn	160 ký tự (Với bộ mã 7-bit) 140 ký tự (Với bộ mã 8-bit) 70 ký tự (Với bộ mã 16-bit)
Kiểu trao đổi thông tin	Tự do	Có xu hướng viết tắt, lược bỏ các cấu trúc ngữ pháp phức tạp và không có dấu (nếu dùng tiếng Việt)
Dữ liệu	Người gửi (Địa chỉ thư), người nhận (Địa chỉ thư), thời gian, tiêu đề, CC, nội dung văn bản, tệp tin đính kèm, ảnh (nếu có)	Người gửi (Số điện thoại), người nhận (số điện thoại), thời gian, nội dung văn bản, trung tâm nhắn tin, bộ mã hóa
Lưu trữ	Máy chủ thư điện tử	Thiết bị người dùng cuối

Trước đó, vào năm 2011, nhà nghiên cứu Nguyễn Đức Tuân (VNCERT) cũng chủ trì một đề tài mang tên "Nghiên cứu và xây dựng hệ thống tiếp nhận và điều phối xử lý tin nhắn SMS rác tại Việt Nam". Tương tự như đề án trên, đề tài này cũng chỉ tập trung vào việc tiếp nhận và điều phối xử lý chứ chưa chú trọng đến phần phát hiện tin nhắn rác. Hiện tại, Công ty Cổ phần Viễn thông Quân đội Viettel công bố có thể chặn tới 90% tin nhắn rác [7]. Tuy nhiên, công nghệ cụ thể đơn vị này sử dụng lại không được công bố rộng rãi nên rất hạn chế trong việc nghiên cứu và tiếp cận thông tin.

III. MÔ HÌNH LỌC TIN NHẮN RÁC VỚI SPAMASSASSIN

A. Giới thiệu về SpamAssassin

SpamAssassin là một hệ thống phần mềm có chức năng phân tích, đánh giá các thư điện tử nhận được và đưa ra kết luận rằng thư đó là thư rác hay thư hợp lệ. Hệ thống hoạt động dựa trên nguyên tắc so sánh các phần khác nhau của một thư điện tử với các bộ luật được định nghĩa sẵn. Với mỗi luật tương ứng, thư điện tử sẽ được tăng hoặc giảm điểm đánh giá. Một thư điện tử đạt điểm đến ngưỡng (threshold) đủ cao thì sẽ bị coi là thư rác [8].

Luật của SpamAssassin bao gồm ba phần: loại luật (*header* hoặc *body*), mô tả luật (*describe*), điểm số (*score*). Dưới đây là một luật điển hình thường thấy trong các bộ luật của SpamAssassin:

```
header FROM_STARTS_WITH_NUM
        From =~ /^\d\d/

describe FROM_STARTS_WITH_NUM
        From: starts with nums

score FROM_STARTS_WITH_NUM     0.390
```

Hình 1. Mẫu luật SpamAssassin

Luật mẫu trên được đặt tên là `FROM_STARTS_WITH_NUM`. Với luật này, SpamAssassin sẽ kiểm tra xem phần tiêu đề của thư điện tử có bắt đầu bằng hai chữ số hay không dựa trên Regular Expression (Regex). Nếu thư nào khớp với luật này sẽ được tăng số điểm tương ứng.

Mặc dù hiện tại đang có rất nhiều sản phẩm phát hiện và chặn thư rác nhưng SpamAssassin vẫn là một trong những hệ thống phổ biến nhất.

B. Lọc tin nhắn rác với SpamAssassin

Theo những kết quả nghiên cứu và thực nghiệm của chúng tôi, sử dụng SpamAssassin để phát hiện và chặn thư rác (với ngôn ngữ tiếng Việt) rất khả quan. Tại ngưỡng 2.5 của SpamAssassin, tỷ lệ phát hiện thư rác tiếng Việt là 81.4% và hoàn toàn không có lỗi đánh dấu nhầm thư hợp lệ thành thư rác (0%) [9]. Căn cứ trên kết quả đó, cùng với những đánh giá về tính tương đồng của thư điện tử và tin nhắn SMS ở phần II.A, chúng tôi cho rằng việc ứng dụng SpamAssassin là khả thi và có cơ sở.

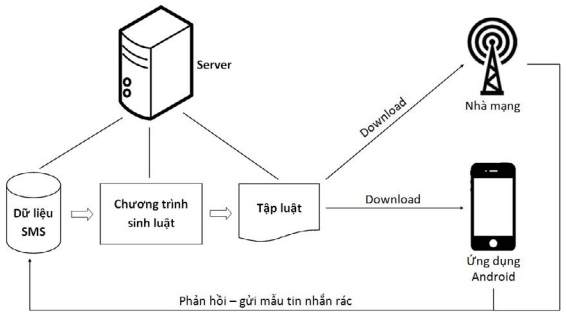
Để mô hình ứng dụng này hoạt động hiệu quả, cần thực hiện một số cải biến như sau:

- Thay đổi bộ luật để phù hợp với đặc thù của tin nhắn rác SMS.
- Thu thập tập dữ liệu tin nhắn SMS (cả tin hợp lệ và tin rác) phục vụ việc huấn luyện hệ thống học máy.
- Cập nhật và tối ưu để hệ thống có thể chạy độc lập, tích hợp với các SMS Gateway của nhà mạng hoặc chạy trên chính thiết bị di động của người dùng cuối.

Sau khi thực hiện những điều chỉnh phù hợp, thực nghiệm sẽ được thực hiện trên tập dữ liệu mẫu. Chi tiết về hệ thống, cách thức vận hành được mô tả chi tiết trong phần tiếp theo.

C. Mô tả hệ thống

Hệ thống bao gồm hai phần chính: Hệ thống học máy dựa trên dữ liệu mẫu để sinh ra bộ luật và Phần mềm chạy trên thiết bị di động sử dụng bộ luật được sinh ra để phát hiện và ngăn chặn tin nhắn rác. Hình 2 thể hiện cấu trúc của toàn bộ hệ thống.



Hình 2. Mô hình triển khai hệ thống

1) Hệ thống học máy sinh bộ luật

Như đã đề cập ở phần III.A, SpamAssassin phát hiện thư rác dựa trên nguyên lý lọc dùng luật có trọng số. Hình thức này được áp dụng tương tự việc lọc tin nhắn rác (hình 3); tức là đối với mỗi tin nhắn m (1...n), nếu khớp với một luật r (1...n) thì sẽ nhận được một điểm trọng số tương ứng. Khi tổng điểm này đạt hoặc lớn hơn ngưỡng T thì tin nhắn m bị coi là Spam.

Trước khi thực hiện huấn luyện và sinh bộ luật, nhóm nghiên cứu phải chuẩn bị tập dữ liệu mẫu. Tập dữ liệu này cần được phân loại và gắn nhãn (tin rác hay tin hợp lệ). Chi tiết về tập dữ liệu được mô tả kỹ hơn ở phần IV.A. Việc xây dựng các luật được thực hiện với các bước như sau:

Bước 1: Tách từ có nghĩa từ tin nhắn: Chúng tôi sử dụng bộ công cụ vnTokenizier để tách các từ từ nội dung tin nhắn [11]. Các từ này sẽ được xuất ra dưới dạng danh sách theo định dạng XML.

Bước 2: Lựa chọn từ khóa để xây bộ luật:

Việc lựa chọn từ khóa để xây dựng bộ luật được dựa trên công thức tính như sau:

$$V_{ts} = P(E|H) = \frac{P(E \wedge H)}{P(H)} \quad (1)$$

$$V_{th} = P(\bar{E}|H) = \frac{P(\bar{E} \wedge H)}{P(\bar{H})} \quad (2)$$

Trong đó: V_{ts} và V_{th} lần lượt thể hiện mối liên hệ giữa từ khóa t với thư rác và từ khóa t với thư hợp lệ.

Sau khi có V_{ts} và V_{th}, N từ khóa có giá trị cao nhất tính theo tỷ lệ $R_t = V_{ts} / V_{th}$ sẽ được chọn. N sẽ là số lượng tập luật (cũng chính là yếu tố quyết định hiệu suất của tập luật). Nếu E là giả thuyết một tin nhắn là spam và H là giả thuyết một tin nhắn là ham và đều chứa từ khóa t thì công thức tính V_{ts} và V_{th} dựa trên lý thuyết xác suất có điều kiện (Conditional Propability) như sau:

$$(E) = \frac{A + C}{A + B + C + D} \quad (3)$$

$$P(\bar{E}) = \frac{B + D}{A + B + C + D} \quad (4)$$

$$P(H) = \frac{A + B}{A + B + C + D} \quad (5)$$

$$P(E \wedge H) = \frac{A}{A + B + C + D} \quad (6)$$

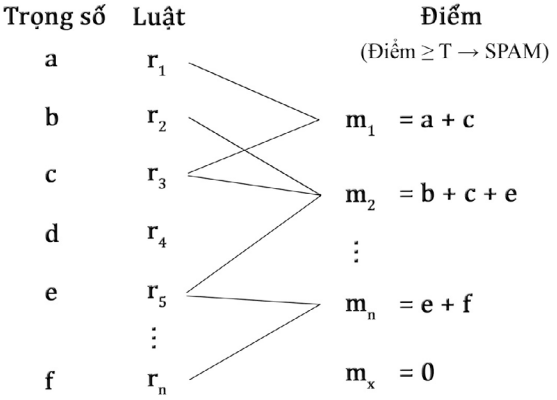
$$P(\bar{E} \wedge H) = \frac{B}{A + B + C + D} \quad (7)$$

Trong đó: Với từ khóa t, A và B là số lần tin nhắn spam và hàm chứa t; C và D là số lần tin nhắn spam và hàm không chứa t. Giá trị xác suất trong (1) và (2) sẽ được tính theo công thức trên.

Nhóm nghiên cứu nhận thấy phần lớn tin nhắn rác đều có dấu hiệu nhận dạng đặc trưng cho từng loại tin nhắn rác. Các dấu hiệu này có thể nhận ra được thông qua việc quan sát dữ liệu. Ví dụ: tin nhắn bán sim số đẹp (chứa nhiều số điện thoại), tin nhắn lừa đảo (chứa cụm từ yêu cầu thuê bao soạn tin nhắn đến tổng đài),… Có thể dùng Regular Expression để nhận diện những mẫu tin nhắn đó, ví dụ:

```
/([0-9]{4,}[^0-9].{1,}){3,}/s
/s(0|o)(@|a)n.+gu(j|i).+?\d+/is
/cl(i|1)p ?sexy?/i
```

Chính vì vậy, các luật heuristics nói trên được bổ sung vào bộ luật để tăng thêm hiệu quả đối với tin nhắn rác có cấu trúc.



Hình 3. Mô tả nguyên lý lọc dùng bộ luật có trọng số

Bước 3: Cập nhật điểm cho bộ luật:

Các luật được ánh xạ vào một mạng nơron một lớp theo nguyên tắc: Mỗi luật tương ứng với một nút mạng, điểm số của luật tương ứng với trọng số nút mạng để thực hiện tối ưu hóa theo phương pháp Stochastic Gradient Descent [12]. Cụ thể như sau: Hệ thống sử dụng hàm chuyển giao tuyến tính (8) và hàm kích hoạt logsig (9) để ghép trọng số với không gian điểm số các luật.

$$f(x) = \sum_{i=1}^{N} w_i x_i \quad (8)$$

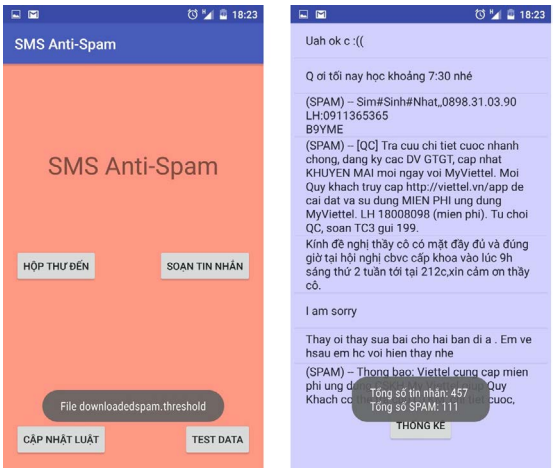
$$y(x) = \frac{1}{1 + e^{-f(x)}} \quad (9)$$

Trong đó: w_i biểu diễn điểm số i và x_i thể hiện việc một tin nhắn có kích hoạt luật i hay không; hàm chuyển giao (8) sẽ trả về điểm số của tin nhắn.

2) Phần mềm trên di động

Mặc dù bộ luật được sinh ra hoàn toàn có thể áp dụng trên các SMS Gateway đặt tại các nhà mạng để ngăn chặn tin nhắn rác trước khi chúng được chuyển đến thiết bị của người dùng nhưng việc triển khai gặp rất nhiều khó khăn cả về thủ tục triển khai, phối hợp với nhà mạng cũng như vấn đề kĩ thuật. Chính vì vậy, để có thể thực hiện thí nghiệm sớm, dễ dàng hơn, nhóm nghiên cứu đã xây dựng một phần mềm trên nền tảng Android dành cho các thiết bị di động của người dùng cuối. Phần mềm sử dụng trực tiếp bộ luật để phân tích và lọc tin nhắn rác ngay trên thiết bị di động của người dùng.

Về mặt giao diện người dùng và chức năng, phần mềm gồm có ba chức năng chính (hình 4):



Hình 4. Giao diện phần mềm trên di động

- Tải tập luật từ máy chủ: Bộ luật được sinh ra trên máy chủ sẽ được phần mềm tải về thông qua kết nối Internet. Bộ luật này sẽ được cập nhật thường xuyên.
- Phát hiện và đánh dấu tin nhắn rác: Dựa vào bộ luật tải về, phần mềm chạy thuật toán với dữ liệu là các tin nhắn có trong hộp thư đến của điện thoại. Tin nhắn rác sẽ được đánh dấu để người dùng dễ dàng nhận biết.
- Gửi mẫu tin nhắn rác về máy chủ: Các tin nhắn rác sẽ được gửi ngược lại về máy chủ phục vụ việc huấn luyện và cập nhật bộ luật (với sự cho phép của người dùng).

Do đây là phiên bản thử nghiệm nên chưa thực sự hoàn thiện về giao diện cũng như cung cấp đầy đủ các chức năng. Tuy nhiên, hai tính năng quan trọng nhất là đánh dấu tin nhắn rác và cập nhật bộ luật qua Internet đã được triển khai và đáp ứng những yêu cầu tối thiểu để thực hiện thí nghiệm, đánh giá kết quả của hệ thống.

IV. THÍ NGHIỆM VÀ KẾT QUẢ

A. Tập dữ liệu mẫu

Tập dữ liệu nhóm nghiên cứu sử dụng gồm có 2781 tin nhắn, được chia thành hai nhóm tập:

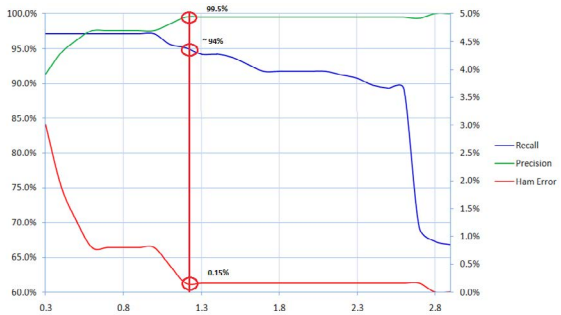
- Nhóm dùng để huấn luyện: 476 tin nhắn rác và 1470 tin nhắn hợp lệ.
- Nhóm dùng cho thí nghiệm: 205 tin nhắn rác và 630 tin nhắn hợp lệ.

Nguồn dữ liệu có được chủ yếu là do thành viên nhóm nghiên cứu đóng góp. Đối với nhóm dữ liệu huấn luyện, nhóm nghiên cứu phải đọc và gắn mác (tin rác hay tin hợp lệ) cho từng tin nhắn. Kết quả của thí nghiệm được ghi lại, phân tích và đánh giá trong phần tiếp theo của báo cáo.

B. Phân tích và đánh giá kết quả

Sau khi huấn luyện với gần 2000 tin nhắn, bộ luật được sinh ra và áp dụng thí nghiệm trên 835 tin nhắn thuộc nhóm thí nghiệm. Phần mềm được cài đặt trên điện thoại HTC One M7 với thông số kỹ thuật như sau: CPU Qualcomm® Snapdragon™ 600, quad-core, 1.7GHz; RAM 2GB DDR2; OS Android™ with HTC Sense™ [13]. Tính trung bình, phần mềm mất 0,0134 giây để tính toán và đánh giá xem một tin nhắn trong tập thí nghiệm là tin nhắn hợp lệ hay tin nhắn rác dựa trên bộ luật đã được sinh ra. Kết quả thí nghiệm được mô tả trong hình 5 với các đại lượng có đơn vị là tỷ lệ phần trăm (%):

- Recall: Tỷ lệ tin nhắn rác được phát hiện.
- Precision: Tỷ lệ cảnh báo chính xác (Độ tin cậy của kết quả).
- Ham Error: Tỷ lệ tin nhắn hợp lệ bị lọc nhầm.



Hình 5. Kết quả thí nghiệm với tập dữ liệu 825 tin nhắn

Theo như kết quả ghi nhận, phần mềm đạt hiệu quả tốt nhất khi đạt ngưỡng 1.175. Với ngưỡng này, tỷ lệ recall đạt tới 94% trong khi tỷ lệ chặn nhầm tin hợp lệ chỉ ở mức 0.15%. Như vậy, tỷ lệ cảnh báo chính xác ở ngưỡng này lên đến 99.5%. Ở ngưỡng thấp hơn (1.0), mặc dù tỷ lệ tin nhắn rác được phát hiện tăng lên đáng kể (gần 96%) nhưng tỷ lệ Ham Error cũng tăng lên rất cao so với ngưỡng 1.175 (0.6%) và tỷ lệ cảnh báo chính xác chỉ còn khoảng 97%. Đặc biệt, khi tăng ngưỡng lên đến 2.8 thì tỷ lệ phát hiện tin nhắn rác giảm mạnh, chỉ còn 67%.

V. THẢO LUẬN

SpamAssassin là một trong những phương pháp rất phổ biến và hiệu quả để phát hiện và lọc thư rác. Trên cơ sở những nghiên cứu đạt được với công nghệ lọc thư nhắn rác tiếng Việt, nhóm nghiên cứu đã triển khai hệ thống lọc tin nhắn rác SMS. Các bộ luật được sinh ra và tối ưu, sau đó chạy trên phần mềm lọc tin nhắn rác cài trực tiếp trên thiết bị di động của người dùng cuối.

Mặc dù tập dữ liệu mẫu còn rất giới hạn nhưng kết quả thu được tương đối khả quan. Kết quả đó sẽ là cơ sở để nhóm nghiên cứu tiếp tục hoàn thiện giải pháp và sản phẩm với những dự định trong tương lai gần:

- Trước hết, nhóm sẽ mở rộng cách tiếp cận, nghiên cứu và thực hiện thí nghiệm với một số phương pháp khác để so sánh kết quả với phương pháp hiện tại hướng đến một giải pháp hiệu quả phát hiện và ngăn chặn tin nhắn rác.
- Triển khai trên hệ thống chặn tin nhắn rác trên SMS Gateway của các nhà mạng.
- Phát triển phần mềm trên các nền tảng khác như Windows, iOS, Blackberry 10…
- Mở rộng phạm vi thu thập dữ liệu mẫu để kết quả thử nghiệm khách quan và chính xác hơn.
- Nghiên cứu thêm các đặc trưng của tin nhắn rác (tập trung vào tin nhắn rác tiếng Việt) để nâng cao hiệu quả phát hiện và ngăn chặn tin nhắn rác.
- Bổ sung thêm các tính năng như: tự động gửi mẫu tin nhắn rác về hệ thống, có thư mục riêng chứa tin nhắn rác để người dùng không phải tiếp xúc với những tin nhắn rác ngay cả khi tin đã được gửi đến thiết bị.

TÀI LIỆU THAM KHẢO

[1] T. O. University, "2014 Text Messaging Usage Statistics," 3 Dec 2014. [Online]. Available: http://www.openuniversity.edu/news/news/2014-text-messaging-usage-statistics. [Accessed 10 Nov 2016].

[2] A. Doherty, "SMS Versus Email Marketing," 28 July 2014. [Online]. Available: http://www.business2community.com/digital-marketing/sms-versus-email-marketing-0957139. [Accessed 5 November 2016].

[3] M. Tauf q Nuruzzaman, Changmoo Lee, Mohd. Fikri Azli bin Abdullah, Deokjai Choi, "Simple SMS spam f ltering on independent mobile phone," in 11th IEEE Conference on Computer and Information Technology, Cyprus, 2011.

[4] Sarah Jane Delany, Mark Buckley, Derek Greene, "SMS spam f ltering: Methods and data," Expert Systems with Applications, vol. 39, no. 10, p. 9899-9908, 2012.

[5] Muhammad Bilal Junaid, Muddassar Farooq, "Using evolutionary learning classif ers to do MobileSpam (SMS) f ltering," in Proceedings of the 13th annual conference on Genetic and evolutionary computation (GECCO'11), New York, NY, USA, 2011.

[6] Akshay Narayan, Prateek Saxena, "The Curse of 140 Characters: Evaluating the Eff cacy of SMS Spam Detection on Android," in Security and Privacy in Smartphones and Mobile Devices, Berlin, Germany, 2013.

[7] V. T. Corp., "Ứng dụng công nghệ mới, Viettel chặn được hơn 90% tin nhắn rác," 3 2 2016. [Online]. Available: http://viettelteleco m.vn/index.php/chi-tiet-tin-tuc/ung-dung-cong-nghe-moi-viettel-chan-duoc-hon-90-tin-nhan-rac. [Accessed 12 10 2016].

[8] A. Schwartz, SpamAssassin, Sebastopol, CA: O'Reilly, 2004.

[9] Minh Tuan Vu, Quang Anh Tran, Frank Jiang, Van Quan Tran, "Multilingual Rules for Spam Detection," Journal of Machine to Machine Communications, vol. 1, p. 107-122, 2014.

[10] Gordon V. Cormack, José María Gómez Hidalgo, Enrique Puertas Sánz, "Feature Engineering for Mobile (SMS) Spam Filtering," in SIGIR'07, Amsterdam, The Netherlands, 2007.

[11] L. H. Phuong, "vnTokenizer -- Vietnamese word segmentation | Lê Hồng Phương," College of Science, Vietnam National University, [Online]. Available: http://mim.hus.vnu.edu.vn/phuonglh/softwares/vnTokenizer. [Accessed 8 Nov 2016].

[12] John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, Donald B. Rubin, Bayesian Data Analysis, CRC Press, 2004.

[13] H. Corporation, "HTC One (M7) Specs and Reviews | HTC United States," HTC Corporation, [Online]. Available: http://www.htc.com/us/smartphones/htc-one-m7/. [Accessed 8 Nov 2016].

SMS SPAM FILTERING WITH SPAMASSASSIN

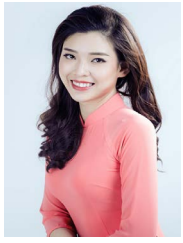
***Abstract*:** As the mobile messaging service becomes more popular and evolving, it is time for spam messages to f ood the telecommunication networks. Like spam, junk messages not only annoy users, but also put pressure on the telecommunications infrastructure and are used as a tool to carry out phishing scams, malicious code and virus spreading... On the basis of spam detection technology, specif cally with SpamAssassin, we has developed a spam detection system, including: rule-generating system and applications for mobile devices. The results show that this is a feasible approach with a 94% spam detection rate while a false false positive rate of 0.15%.

Vũ Minh Tuấn là giảng viên tại khoa Công nghệ thông tin - Trường Đại học Hà Nội. Hiện tại, anh đang làm nghiên cứu sinh ngành Hệ thống thông tin tại Học viện Công nghệ Bưu chính Viễn thông.

Lĩnh vực nghiên cứu và chuyên môn bao gồm: AntiSpam, công nghệ phần mềm và phân tích, thiết kế hệ thống thông tin.

Đặng Đình Quân là giảng viên tại khoa Công nghệ thông tin - Trường Đại học Hà Nội.

Lĩnh vực nghiên cứu và chuyên môn bao gồm: AntiSpam, học máy và giải thuật tiến hóa.



Nguyễn Thanh Hà hiện đang công tác tại Sở Thông tin và Truyền thông Hà Nội. Hiện chị đang làm nghiên cứu sinh ngành Hệ thống thông tin tại Học viện Công nghệ Bưu chính Viễn thông.

Lĩnh vực nghiên cứu và chuyên môn bao gồm: AntiSpam, công nghệ phần mềm và hệ thống thông tin.

PGS.TS Trần Quang Anh hiện là Phó Giám đốc Học viện Công nghệ Bưu chính Viễn thông. Ông hoàn thành chương trình thạc sĩ và tiến sĩ tại Đại học Trường Thanh Hoa, Trung Quốc.

Lĩnh vực nghiên cứu của ông bao gồm: an ninh mạng, thuật toán tiến hóa, chống thư rác…