



(12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ

(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN)
CỤC SỞ HỮU TRÍ TUỆ

(11)



1-0042371

(51)⁷ G06F 21/56

(13) B

(21) 1-2019-04491

(22) 29/06/2018

(86) PCT/CN2018/093791 29/06/2018

(87) WO 2019/007288 10/01/2019

(30) 201710543805.4 05/07/2017 CN

(45) 27/01/2025 442

(43) 25/10/2019 379A

(73) Advanced New Technologies Co., Ltd. (KY)

Cayman Corporate Centre, 27 Hospital Road, George Town, Grand Cayman KY1-9008, Cayman Islands

(72) SUN, Qingqing (CN).

(74) Công ty Luật TNHH T&G (TGVN)

(54) PHƯƠNG PHÁP NHẬN DẠNG ĐỊA CHỈ RỦI RO, THIẾT BỊ ĐIỆN TỬ, VÀ
PHƯƠNG TIỆN LƯU TRỮ ĐỌC ĐƯỢC BẰNG MÁY TÍNH

(21) 1-2019-04491

(57) Sáng chế đề cập đến các phương án mà chúng bộc lộ phương pháp và thiết bị nhận dạng địa chỉ rủi ro, và thiết bị điện tử. Phương pháp nhận dạng địa chỉ rủi ro này bao gồm các bước: sau khi thu chuỗi từ địa chỉ tương ứng theo địa chỉ đầu vào, thực hiện việc nhận dạng rủi ro trên địa chỉ đầu vào này, sử dụng mô hình Markov ẩn và mô hình máy vector hỗ trợ theo từ địa chỉ trùng khớp trong chuỗi từ địa chỉ và ngữ nghĩa trước và sau nó.

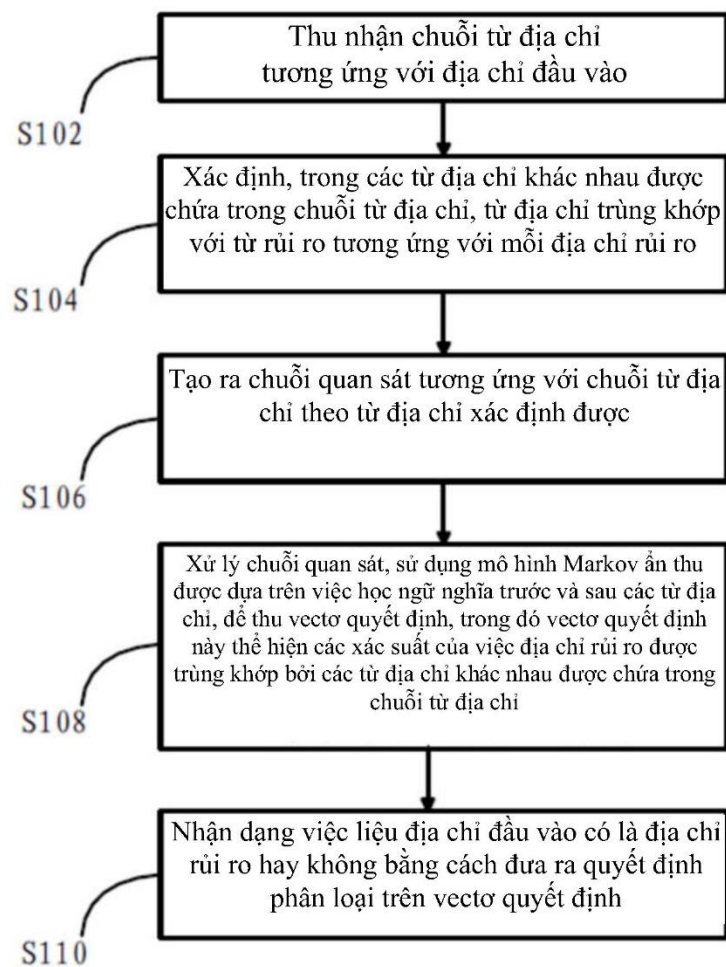


Fig. 2

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến lĩnh vực công nghệ máy tính, và cụ thể là đề cập đến phương pháp và thiết bị nhận dạng địa chỉ rủi ro, và thiết bị điện tử.

Tình trạng kỹ thuật của sáng chế

Các giao dịch tài chính trực tuyến đang ngày càng phát triển, và một số người dùng sử dụng phi pháp tài chính trực tuyến để tiến hành các hoạt động tài chính bất hợp pháp, chẳng hạn như rửa tiền. Để ngăn chặn sự xuất hiện của các giao dịch tài chính trực tuyến bất hợp pháp, đã có một số giải pháp để nhận dạng các địa chỉ rủi ro, mà chúng là các giải pháp nói chung được nhắm đến việc nhận dạng các địa chỉ rủi ro tương đối thông thường.

Ví dụ, việc phân đoạn từ và gán nhãn có thể được thực hiện trên địa chỉ đầu vào, sử dụng thuật toán phân đoạn từ, và cuối cùng việc so khớp từ địa chỉ được thực hiện từng từ một theo thông tin được gán nhãn của các từ địa chỉ khác nhau, để nhận dạng xem liệu địa chỉ đầu vào này có phải là địa chỉ rủi ro hay không thông qua kết quả so khớp.

Dựa trên tình trạng kỹ thuật này, cần thiết có một giải pháp nhận dạng địa chỉ rủi ro chính xác hơn.

Bản chất kỹ thuật của sáng chế

Các phương án của sáng chế cung cấp phương pháp và thiết bị nhận dạng địa chỉ rủi ro, và thiết bị điện tử, để giải quyết vấn đề kỹ thuật sau đây: cần có một giải pháp nhận dạng địa chỉ rủi ro chính xác hơn.

Để giải quyết vấn đề kỹ thuật nêu trên, các phương án của sáng chế được thi hành như sau:

các phương án của sáng chế đề xuất phương pháp nhận dạng địa chỉ rủi ro, phương pháp bao gồm các bước:

thu nhận chuỗi từ địa chỉ (address word sequence) tương ứng với địa chỉ đầu vào;
xác định, trong các từ địa chỉ (address word) khác nhau được chứa trong chuỗi từ địa chỉ này, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro;

tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ xác định được này;

xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ; và

nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

Các phương án của sáng chế đề xuất thiết bị nhận dạng địa chỉ rủi ro, thiết bị này bao gồm:

môđun nhận để thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào;

môđun so khớp để xác định, trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ này, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro;

môđun tạo để tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ xác định được;

môđun gán nhãn để xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ; và

môđun nhận dạng để nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

Các phương án của sáng chế đề xuất thiết bị điện tử, bao gồm:

ít nhất một bộ xử lý; và

bộ nhớ kết nối truyền thông với ít nhất một bộ xử lý này, trong đó

bộ nhớ này lưu trữ các lệnh có thể thực thi được bởi ít nhất một bộ xử lý này, các lệnh này được thực thi bởi ít nhất một bộ xử lý này để cho phép ít nhất một bộ xử lý này:

thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào;

xác định, trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ này, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro;

tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ được xác

định này;

xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ; và

nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

Ít nhất một giải pháp kỹ thuật trong số các giải pháp kỹ thuật được mô tả trên đây được sử dụng trong các phương án của sáng chế có thể đạt được các hiệu quả có lợi sau đây: sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ và mô hình máy vector hỗ trợ để thu kết quả xác định phân loại của địa chỉ đầu vào theo từ địa chỉ thu được sau khi xử lý địa chỉ đầu vào và ngữ nghĩa trước và sau từ địa chỉ, để có thể nhận dạng địa chỉ rủi ro một cách chính xác hơn.

Mô tả vắn tắt các hình vẽ

Để minh họa các giải pháp kỹ thuật trong các phương án của sáng chế hoặc giải pháp kỹ thuật có trước một cách rõ ràng hơn, các hình vẽ kèm theo sẽ được sử dụng trong phần mô tả của các phương án hoặc giải pháp kỹ thuật có trước sẽ được giới thiệu đơn giản dưới đây. Rõ ràng, các hình vẽ kèm theo trong phần mô tả sau đây chỉ đơn thuần thể hiện một số phương án được nêu trong sáng chế, và người có hiểu biết trung bình trong lĩnh vực kỹ thuật vẫn có thể suy ra các hình vẽ khác từ các hình vẽ kèm theo này mà không cần bất kỳ nỗ lực sáng tạo nào.

Fig.1 là sơ đồ giản lược của kiến trúc tổng thể liên quan đến giải pháp của sáng chế theo kịch bản ứng dụng thực tế;

Fig.2 là lưu đồ sơ lược của phương pháp nhận dạng địa chỉ rủi ro được đề xuất trong một phương án của sáng chế;

Fig.3 là lưu đồ sơ lược của phương pháp lập mô hình cho mô hình Markov ẩn được đề xuất trong một phương án của sáng chế;

Fig.4 là lưu đồ sơ lược của phương pháp lập mô hình cho mô hình máy vector hỗ trợ được đề xuất trong một phương án của sáng chế;

Fig.5 là sơ đồ cấu trúc giản lược của thiết bị nhận dạng địa chỉ rủi ro được đề xuất trong một phương án của sáng chế;

Fig.6 là lưu đồ sơ lược trong đó việc nhận dạng địa chỉ rủi ro chỉ ra rằng địa chỉ không có rủi ro, được đề xuất trong một phương án của sáng chế;

Fig.7 là lưu đồ sơ lược trong đó việc nhận dạng địa chỉ rủi ro chỉ ra rằng địa chỉ là rủi ro, được đề xuất trong một phương án của sáng chế; và

Fig.8 là lưu đồ sơ lược của việc lập mô hình và nhận dạng của việc nhận dạng địa chỉ rủi ro được đề xuất trong một phương án của sáng chế.

Mô tả chi tiết sáng chế

Các phương án của sáng chế đề xuất phương pháp và thiết bị nhận dạng địa chỉ rủi ro, và thiết bị điện tử.

Để những người có hiểu biết trung bình trong lĩnh vực tương ứng hiểu rõ hơn về các giải pháp kỹ thuật trong sáng chế, các giải pháp kỹ thuật trong các phương án của sáng chế sẽ được mô tả rõ ràng và đầy đủ dưới đây kết hợp với các hình vẽ kèm theo trong các phương án của sáng chế. Hiển nhiên là các phương án được mô tả chỉ là một số phương án của sáng chế, mà không phải là tất cả. Tất cả các phương án khác thu được bởi những người có trình độ trung bình trong lĩnh vực kỹ thuật dựa trên các phương án của sáng chế mà không cần nỗ lực sáng tạo sẽ nằm trong phạm vi bảo hộ của sáng chế.

Fig.1 là sơ đồ giản lược của kiến trúc tổng thể liên quan đến giải pháp của sáng chế theo kịch bản ứng dụng thực tế. Trong kiến trúc tổng thể này, trước tiên, địa chỉ được nhập vào và được xử lý bởi thiết bị chứa mô hình Markov ẩn (Hidden Markov Model, HMM) để thu vector quyết định tương ứng với địa chỉ đầu vào này, sau đó vector quyết định này được xử lý thêm bởi thiết bị chứa máy vector hỗ trợ (Support Vector Machine, SVM), và cuối cùng, việc liệu địa chỉ đầu vào này có là địa chỉ rủi ro hay không được xác định.

Dựa trên kiến trúc tổng thể nêu trên, giải pháp của sáng chế sẽ được mô tả chi tiết dưới đây.

Các phương án của sáng chế đề xuất phương pháp nhận dạng địa chỉ rủi ro. Fig.2 là lưu đồ sơ lược của phương pháp nhận dạng địa chỉ rủi ro, và lưu trình trên Fig.2 có thể bao gồm các bước sau đây:

S102: chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào được thu nhận.

Chuỗi từ địa chỉ này có thể được thu sau khi xử lý địa chỉ đầu vào này.

Ví dụ, nếu địa chỉ đầu vào chứa nhiều ký tự nhiễu, hoạt động loại bỏ ký tự nhiễu

trước hết có thể được thực hiện trên địa chỉ đầu vào, rồi sau đó chuỗi từ địa chỉ mới được thu. Chuỗi từ địa chỉ cũng có thể là chính địa chỉ đầu vào nếu địa chỉ đầu vào này bao gồm các dấu phân cách từ chẳng hạn như ký tự trắng, hoặc nếu khả năng xử lý, để loại trừ các ký tự nhiễu, của việc nhận dạng địa chỉ rủi ro trong các bước tiếp theo là tốt.

S104: từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro được xác định trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ.

Các địa chỉ rủi ro khác nhau này có thể được hiểu là thư viện địa chỉ rủi ro gồm nhiều địa chỉ rủi ro.

Từ rủi ro có thể được hiểu là từ địa chỉ rủi ro được chứa trong địa chỉ rủi ro. Địa chỉ rủi ro có thể chứa một hoặc nhiều từ rủi ro, và những từ rủi ro này có thể cấu thành nên địa chỉ rủi ro một cách tuần tự theo một quy tắc nhất định hoặc có thể tương đối độc lập.

Ví dụ, phương pháp xác định việc trùng khớp có thể bao gồm các bước: quét tương ứng mỗi từ địa chỉ trong chuỗi từ địa chỉ và so khớp nó với mỗi từ địa chỉ trong địa chỉ rủi ro, và xác định từ địa chỉ trùng khớp (để dễ mô tả, từ địa chỉ như vậy ở dưới đây có thể được gọi là từ địa chỉ trùng khớp).

Ví dụ, nếu chuỗi từ địa chỉ chứa nhiều từ địa chỉ, và từ địa chỉ A trong chuỗi địa chỉ so khớp và trùng khớp với từ rủi ro được chứa trong địa chỉ rủi ro, chuỗi từ địa chỉ có thể được biểu diễn dưới dạng một vector nhiều chiều, với mỗi chiều của vector lần lượt tương ứng với một từ địa chỉ trong chuỗi từ địa chỉ; hơn nữa, chiều, trong vector này, tương ứng với từ địa chỉ A có thể được xác định bằng cách quét, và chiều này được gán nhãn là 1 để xác định rằng từ địa chỉ A trùng khớp; và đối với các từ địa chỉ khác không trùng khớp, các chiều tương ứng được gán nhãn là 0.

S106: chuỗi quan sát tương ứng với chuỗi từ địa chỉ được tạo ra theo từ địa chỉ xác định được.

Có nhiều cách tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ: mỗi từ địa chỉ cần thiết có thể được thu từ chuỗi từ địa chỉ theo từ địa chỉ xác định được và ngữ nghĩa trước và sau nó, và chuỗi quan sát được tạo ra theo từ địa chỉ xác định được và các từ địa chỉ thu được, trong đó các từ địa chỉ được trích xuất này có thể là các từ địa chỉ liên tục, và cũng có thể là các từ địa chỉ không liên tục mà chúng tuân thủ theo một quy tắc nhất định;

Cách tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ cũng có thể là: trước hết, chia tách chuỗi từ địa chỉ để thu nhiều chuỗi từ địa chỉ con, và sau đó tạo ra chuỗi quan sát theo chuỗi từ địa chỉ con cần thiết.

S108: chuỗi quan sát được xử lý, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ.

Ngữ nghĩa trước và sau các từ địa chỉ có thể là ngữ nghĩa được liên kết của các từ địa chỉ trước và sau, và được liên kết với, từ địa chỉ hiện tại, và từ địa chỉ hiện tại, và số lượng các từ địa chỉ được liên kết này có thể là hai hoặc ba, trong đó các từ địa chỉ được liên kết này có thể hoặc là các từ địa chỉ liền kề với từ địa chỉ hiện tại hoặc là các từ địa chỉ cách từ địa chỉ hiện tại một từ.

Cụ thể, từ địa chỉ hiện tại nêu trên có thể là từ địa chỉ xác định được trong bước S106. Ngữ nghĩa trước và sau từ địa chỉ có thể là các từ địa chỉ trước và sau đó, và liền kề với, từ địa chỉ xác định được đóng vai trò là ngữ nghĩa trước và sau từ địa chỉ, ví dụ, quận Chaoyang, Bắc Kinh, Trung Quốc, trong đó ngữ nghĩa trước và sau Bắc Kinh bao gồm Trung Quốc và quận Chaoyang; và cũng có thể sử dụng các từ địa chỉ, trong chuỗi từ địa chỉ, không liền kề nhau để làm ngữ nghĩa trước và sau từ xác định được.

Hơn nữa, để đơn giản hóa các bước xử lý, chuỗi quan sát cũng có thể là chính chuỗi từ địa chỉ hoặc địa chỉ đầu vào. Trong trường hợp này, xác suất trùng khớp địa chỉ rủi ro có thể được gán nhãn trực tiếp dựa trên từ địa chỉ trùng khớp và ngữ nghĩa trước và sau nó theo kết quả gán nhãn của từ địa chỉ trùng khớp này trong chuỗi từ địa chỉ hoặc địa chỉ đầu vào.

S110: việc liệu địa chỉ đầu vào có là địa chỉ rủi ro hay không được nhận dạng bằng cách đưa ra quyết định phân loại trên vector quyết định.

Quyết định phân loại có thể là quyết định phân loại nhị phân được tạo ra trên vector quyết định. Ví dụ, một phân loại có thể là vector quyết định tương ứng với địa chỉ rủi ro, và phân loại kia có thể là vector quyết định tương ứng với địa chỉ không rủi ro. Hơn nữa, việc liệu địa chỉ đầu vào tương ứng với vector quyết định có phải là địa chỉ rủi ro hay không có thể được xác định.

Trong các phương án của sáng chế, đối với bước S102, việc thu nhận chuỗi từ địa

chỉ tương ứng với địa chỉ đầu vào có thể cụ thể bao gồm việc:

nhận địa chỉ đầu vào; và thực hiện việc xử lý làm sạch dữ liệu và việc xử lý phân đoạn từ trên địa chỉ đầu vào này để thu chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào này.

Phương pháp xử lý làm sạch dữ liệu có thể bao gồm bước: loại bỏ ký tự nhiễu khỏi địa chỉ đầu vào để thu chuỗi ký tự địa chỉ được tiêu chuẩn hóa; hoặc điều chỉnh thứ tự sắp xếp của các từ địa chỉ khác nhau trong địa chỉ đầu vào.

Phương pháp xử lý phân đoạn từ có thể bao gồm bước: phân đoạn chuỗi ký tự địa chỉ được tiêu chuẩn hóa bằng ký tự cụ thể chẳng hạn như ký tự trắng hoặc dấu chấm phẩy, để thu chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào.

Như đã nêu trước đây, nếu địa chỉ đầu vào chứa nhiều ký tự nhiễu thì hoạt động loại bỏ ký tự nhiễu trước hết có thể được thực hiện trên địa chỉ đầu vào, rồi sau đó chuỗi từ địa chỉ mới được thu.

Ví dụ, trước hết, việc làm sạch dữ liệu được thực hiện trên một số ký tự nhiễu hiện có trong địa chỉ đầu vào để loại bỏ các ký tự nhiễu này, trong đó, ví dụ, các ký tự nhiễu này có thể bao gồm ít nhất một ký tự trong các ký tự sau: các ký tự trắng thừa, các ký tự nửa độ rộng (half-width), “|”, “:”, “~”, v.v., và địa chỉ đầu vào được làm sạch có thể là, ví dụ, chuỗi ký tự địa chỉ được tiêu chuẩn hóa cách nhau bởi các ký tự trắng; sau đó, việc gán nhãn phân đoạn từ được thực hiện: chuỗi ký tự địa chỉ được làm sạch này được phân đoạn theo khoảng trắng, các từ phụ trợ, chẳng hạn như “of” và “the”, xuất hiện trong chuỗi ký tự địa chỉ này được gán nhãn theo kiểu từ điển, và các từ phụ trợ này thường không thể được sử dụng trong việc so khớp và nhận dạng chi tiếp theo; và cuối cùng, chuỗi từ địa chỉ gồm các từ địa chỉ được thu sau việc gán nhãn phân đoạn từ này. Sau khi thực hiện việc làm sạch và phân đoạn từ trên địa chỉ đầu vào, chuỗi từ địa chỉ được thu, sao cho tiếp đó các hoạt động chẳng hạn như nhận dạng các từ địa chỉ rủi ro có thể được thực hiện trên chuỗi từ địa chỉ nhanh chóng, hiệu quả và chính xác.

Như đã nêu trước đây, chuỗi từ địa chỉ cũng có thể là chính địa chỉ đầu vào nếu địa chỉ đầu vào bao gồm các dấu phân cách từ chẳng hạn như ký tự trắng, hoặc nếu khả năng xử lý, để loại trừ các ký tự nhiễu, của việc nhận dạng địa chỉ rủi ro trong các bước tiếp theo là tốt; như vậy, bước xử lý địa chỉ đầu vào có thể được lược bỏ.

Trong các phương án của sáng chế, đối với bước S104, việc xác định, trong các từ

địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro có thể cụ thể bao gồm việc:

so khớp tương ứng các từ địa chỉ khác nhau trong chuỗi từ địa chỉ, sử dụng các từ rủi ro tương ứng với mỗi địa chỉ rủi ro; và

nếu một từ địa chỉ trong số các từ địa chỉ này được so khớp thành công thì gán nhãn nó và xác định nó là từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro.

Việc gán nhãn có thể là việc: gán nhãn tương ứng, với số hoặc ký tự, kết quả so khớp của các từ địa chỉ khác nhau trong chuỗi từ địa chỉ tương ứng, bao gồm việc gán nhãn kết quả chỉ ra sự trùng khớp và việc gán nhãn kết quả chỉ ra sự không trùng khớp sau khi so khớp trong chuỗi từ địa chỉ, và tạo ra vector gán nhãn bằng các số hoặc các ký tự chỉ ra các kết quả gán nhãn này với nhau.

Ví dụ, chuỗi từ địa chỉ hoặc các từ địa chỉ khác nhau trong chuỗi quan sát tương ứng với chuỗi từ địa chỉ này được quét toàn cục, so khớp, và gán nhãn, và nếu từ địa chỉ thứ hai A trong chuỗi từ địa chỉ nào đó khớp với từ địa chỉ a trong tập hợp từ địa chỉ tương ứng với địa chỉ rủi ro thì từ địa chỉ A được gán nhãn là 1, nếu không thì từ địa chỉ A được gán nhãn là 0. Hơn nữa, vector gán nhãn được thu, mà nó cũng có thể được gọi là vector gán nhãn ban đầu, chẳng hạn như $[0, 0, 1, 0, 0, 0]$.

Các từ địa chỉ khác nhau trong chuỗi từ địa chỉ được so khớp tương ứng, và phương pháp so khớp cụ thể có thể bao gồm các bước: thực hiện việc quét và so khớp địa chỉ rủi ro toàn cục trên chuỗi từ địa chỉ (thông tin địa chỉ rủi ro bao gồm các quốc gia, các khu vực, và các thành phố lớn), trong đó các thuật toán so khớp được sử dụng, ví dụ, có thể bao gồm thuật toán so khớp sự tương tự của chuỗi ký tự, thuật toán so khớp sự tương tự của cách phát âm, và thuật toán so khớp khoảng cách hiệu chỉnh, v.v.

Hơn nữa, việc xác định, trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro còn bao gồm việc:

nếu không có từ địa chỉ nào được so khớp thành công thì xác định rằng địa chỉ đầu vào không phải là địa chỉ rủi ro.

Nếu các từ địa chỉ khác nhau trong chuỗi từ địa chỉ nhất định không khớp với các từ rủi ro trong các địa chỉ rủi ro thì có thể coi rằng chuỗi từ địa chỉ không chứa từ rủi ro,

và tương ứng, xác suất mà địa chỉ đầu vào không phải là địa chỉ rủi ro là lớn hơn, do đó các hoạt động tiếp theo trên địa chỉ đầu vào này có thể được kết thúc.

Trong các phương án của sáng chế, đối với bước S106, việc tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ theo từ địa chỉ xác định được có thể cụ thể bao gồm việc:

đối với từ địa chỉ xác định được, thực hiện tương ứng như sau: tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ theo từ địa chỉ này và các từ được liên kết của từ địa chỉ này trong chuỗi từ địa chỉ, trong đó các từ được liên kết này phản ánh ngữ nghĩa trước và sau từ địa chỉ này trong chuỗi từ địa chỉ.

Nếu chuỗi từ địa chỉ chứa nhiều từ địa chỉ xác định được, nhiều chuỗi quan sát tương ứng có thể được tạo ra, và cũng có thể chọn để tạo ra một chuỗi quan sát tương ứng với một từ địa chỉ trong số các từ địa chỉ xác định được. Từ được liên kết có thể được hiểu là từ mà nó có mối quan hệ logic nhất định hoặc mối quan hệ liên kết cụ thể với các từ trước và sau nó, hoặc từ được giả định để tạo ra mối quan hệ liên kết với các từ trước và sau nó. Thông thường, các từ được liên kết là ở liền kề với từ hiện tại, ví dụ, chúng có thể là hai từ trước hoặc sau, và liền kề với, từ hiện tại.

Trong các phương án của sáng chế, mô hình Markov ẩn có thể được thu từ trước dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ.

Các phương án của sáng chế đề xuất phương pháp lập mô hình cho mô hình Markov ẩn. Fig.3 là lưu đồ sơ lược của phương pháp lập mô hình này, và lưu trình trên Fig.3 có thể bao gồm các bước sau đây:

S202: trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước, và thiết lập mô hình ban đầu chứa các tham số mô hình Markov ẩn theo các tham số ban đầu này, trong đó các mẫu huấn luyện này là các địa chỉ rủi ro hoặc các địa chỉ không rủi ro;

S204: tạo ra các chuỗi quan sát tương ứng với các mẫu huấn luyện theo các từ địa chỉ được chứa trong các mẫu huấn luyện và ngữ nghĩa trước và sau các từ địa chỉ này; và

S206: huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn.

Các mẫu huấn luyện bao gồm các mẫu khẳng định và các mẫu phủ định; địa chỉ rủi ro có thể là mẫu khẳng định, và địa chỉ không rủi ro có thể là mẫu phủ định. Bằng

cách huấn luyện mô hình Markov ẩn với các mẫu khẳng định và phủ định, có thể đạt được hiệu quả huấn luyện tốt hơn.

Ví dụ, trong một cách thi hành, các mẫu phủ định có thể là các địa chỉ của 235 quốc gia hoặc khu vực không được thừa nhận trên toàn thế giới, và các mẫu khẳng định có thể là dữ liệu địa chỉ cho tất cả các quốc gia và khu vực được thừa nhận. Cần lưu ý rằng, nói chung, địa chỉ rủi ro chứa từ rủi ro, và tất nhiên, địa chỉ không rủi ro cũng có thể chứa từ rủi ro.

Các từ địa chỉ trong bước S204 có thể được hiểu là các từ địa chỉ, mà chúng khớp với các từ rủi ro tương ứng với các địa chỉ rủi ro, trong các mẫu huấn luyện. Trong địa chỉ mẫu được xác định trước, việc liệu địa chỉ mẫu này có chứa từ rủi ro hay không và việc liệu từ địa chỉ có là từ rủi ro hay không đã được gán nhãn từ trước. Ở đây, địa chỉ mẫu là địa chỉ mẫu huấn luyện được chứa trong các mẫu huấn luyện.

Hơn nữa, chuỗi quan sát cần thiết được trích xuất theo từ địa chỉ được gán nhãn trong địa chỉ mẫu và ngữ nghĩa trước và sau nó; và thông thường, 2 hoặc 3 từ trước và sau và được liên kết với từ địa chỉ được gán nhãn này được chọn làm ngữ nghĩa trước và sau từ địa chỉ này, mà tất cả chúng cùng nhau tạo ra chuỗi quan sát.

Hơn nữa, theo mô hình gán nhãn ban đầu thu được và chuỗi quan sát, các tham số mô hình Markov ẩn được huấn luyện cho đến khi thu được các tham số mô hình Markov ẩn thích hợp, và sau đó mô hình Markov ẩn cần thiết được xác định theo các tham số mô hình Markov ẩn thu được thông qua việc huấn luyện này.

Trong phương pháp lập mô hình cho mô hình Markov ẩn trên Fig.3, số lượng và quy mô của các địa chỉ mẫu huấn luyện sẽ ảnh hưởng trực tiếp đến chất lượng của kết quả huấn luyện. Khi thư viện địa chỉ được sử dụng cho các địa chỉ mẫu huấn luyện là toàn diện hơn và có phạm vi bao phủ rộng hơn, việc cải thiện độ chính xác của việc nhận dạng của mô hình là ưu việt, và do đó, phương pháp này có thể đạt được hiệu quả lập mô hình to lớn trong trường hợp có đủ địa chỉ mẫu huấn luyện.

Các tham số ban đầu bao gồm: vector xác suất ban đầu π , ma trận chuyển tiếp trạng thái q_{ij} , và dạng tương tự; và

việc trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước có thể cụ thể bao gồm việc:

dựa trên các mẫu huấn luyện, thu vector xác suất ban đầu bằng cách thực hiện việc gán nhãn xác suất tương ứng trên các từ địa chỉ được chứa trong các mẫu huấn luyện; và thu ma trận chuyển tiếp trạng thái của các địa chỉ mẫu theo các xác suất chuyển tiếp trạng thái, giữa từ rủi ro và từ không rủi ro, của các từ địa chỉ được chứa trong các mẫu huấn luyện. Ở đây, từ rủi ro là từ địa chỉ được chứa trong địa chỉ rủi ro, từ không rủi ro là từ địa chỉ được chứa trong địa chỉ không rủi ro, và một số từ địa chỉ có thể là cả từ rủi ro lẫn từ không rủi ro.

Có thể có nhiều mẫu huấn luyện. Nhìn chung, mỗi mẫu huấn luyện có thể là một địa chỉ mẫu. Khi việc gán nhãn xác suất được thực hiện, địa chỉ mẫu được chọn từ tập hợp làm mẫu huấn luyện; ví dụ, sau khi việc gán nhãn xác suất được thực hiện trên các từ địa chỉ trong địa chỉ mẫu nhất định, từ địa chỉ thứ hai trong địa chỉ mẫu này là từ địa chỉ trùng khớp, và thu được vector xác suất ban đầu $\pi_t = [0, 1, 0, 0, 0, 0]$, trong đó 1 thể hiện từ địa chỉ trùng khớp, và 0 thể hiện từ các địa chỉ không trùng khớp.

Hơn nữa, ma trận chuyển tiếp trạng thái của các địa chỉ mẫu được thu theo các xác suất chuyển tiếp trạng thái, giữa từ rủi ro và từ không rủi ro, của các từ địa chỉ được chứa trong các mẫu huấn luyện. Trong một hoặc nhiều phương án của sáng chế, xác suất chuyển tiếp trạng thái đề cập đến xác suất mà việc chuyển tiếp trạng thái của từ địa chỉ giữa hai trạng thái ẩn, cụ thể là, địa chỉ rủi ro và địa chỉ không rủi ro, có thể xảy ra.

Trong các phương án của sáng chế, việc huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn có thể cụ thể bao gồm việc:

theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện, sử dụng thuật toán Baum-Welch để huấn luyện các tham số mô hình Markov ẩn để thu mô hình Markov ẩn.

Thêm vào đó, các dữ liệu khác để huấn luyện các tham số mô hình Markov ẩn cũng có thể được thu nhận, ví dụ, $O = o_1 o_2 \cdots o_h \cdots o_{n-1} o_n$, trong đó chuỗi O là chuỗi từ địa chỉ chứa từ quốc gia/khu vực rủi ro trùng khớp, và o_h là từ địa chỉ cần được gán nhãn. Nhìn chung, có thể lấy $n = 10$; chuỗi quan sát được thu theo 3 từ trước và sau từ địa chỉ trong ngữ cảnh của o_h cần được gán nhãn, trong đó $S = s_1 s_2 \cdots s_h \cdots s_{n-1} s_n$ là vector gán nhãn

tương ứng với chuỗi từ địa chỉ $\mathcal{Q}$, và thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau trong chuỗi từ địa chỉ; và sau đó các xác suất trùng khớp của các từ địa chỉ khác nhau này tạo thành vector gán nhãn, trong đó vector gán nhãn có thể được hiểu là vector gán nhãn ban đầu;

$P(o_h, s_h | \lambda)$ chỉ ra xác suất mà chuỗi từ địa chỉ o_h và nhãn trùng khớp s_h của nó là chính xác, và được sử dụng để chọn tham số mô hình Markov ẩn cần thiết λ ; và

sau đó, tham số mô hình Markov ẩn λ được huấn luyện theo tham số được mô tả trên đây để thu mô hình Markov ẩn.

Hơn nữa, hàm mục tiêu của mô hình này được định ra là: $\arg \max P(o_h, s_h | \lambda)$, để tạo thuận lợi cho việc thu nhận vector quyết định cần thiết;

Trong các phương án của sáng chế, đối với bước S108, việc xử lý chuỗi quan sát, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định có thể cụ thể bao gồm việc:

xử lý chuỗi quan sát, sử dụng mô hình Markov ẩn thu được bằng cách học ngữ nghĩa trước và sau các từ địa chỉ, và thuật toán Viterbi, để thu vector quyết định,

trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ. Thêm vào đó, vector quyết định cũng có thể thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi quan sát tương ứng với chuỗi từ địa chỉ, và từ địa chỉ không được chứa trong chuỗi quan sát có thể được gán nhãn trực tiếp là 0.

Ví dụ, nếu chuỗi từ địa chỉ [A, B, C, D, E, F] có chuỗi quan sát tương ứng [B, C, D] thì vector quyết định thu được có thể được biểu diễn dưới dạng [0; 0,5; 1; 0,3; 0; 0].

Trong các phương án của sáng chế, đối với bước S110, việc đưa ra quyết định phân loại trên vector quyết định có thể cụ thể bao gồm việc:

đưa ra quyết định phân loại trên vector quyết định, sử dụng mô hình SVM thu được thông qua việc huấn luyện.

Nói chung, vector quyết định được thu bằng việc tính toán thông qua mô hình Markov ẩn, và sau đó quyết định với phân loại nhị phân hoặc nhiều phân loại hơn được thực hiện trên vector quyết định này; và cách giải thích được thực hiện dưới đây chủ yếu

thông qua việc lấy quyết định phân loại nhị phân làm ví dụ.

Trong các phương án của sáng chế, phương pháp lập mô hình cho mô hình máy vectơ hỗ trợ được đề xuất. Fig.4 là lưu đồ sơ lược của phương pháp lập mô hình cho mô hình máy vectơ hỗ trợ, và lưu trình trên Fig.4 có thể bao gồm các bước sau đây:

S302: thu nhận các mẫu huấn luyện cho máy vectơ hỗ trợ;

S304: ánh xạ các mẫu huấn luyện của máy vectơ hỗ trợ này sang không gian đặc điểm đa chiều, để thu không gian đặc điểm mẫu tương ứng với các mẫu huấn luyện của máy vectơ hỗ trợ này;

S306: thu nhận, từ không gian đặc điểm mẫu này, các tham số thể hiện các đặc điểm mẫu, và thiết lập hàm phân biệt để xác định loại của các đặc điểm mẫu này theo các tham số của các đặc điểm mẫu này; và

S308: huấn luyện các tham số mô hình SVM tương ứng trong hàm phân biệt này dựa trên các mẫu huấn luyện của máy vectơ hỗ trợ, để thu mô hình SVM.

Các mẫu huấn luyện cho máy vectơ hỗ trợ có thể là các vectơ quyết định tương ứng với các mẫu huấn luyện thu được bằng cách huấn luyện các địa chỉ mẫu thông qua mô hình Markov ẩn được mô tả trên đây, hoặc có thể là dữ liệu cần được phân biệt khác định rõ đặc trưng của địa chỉ đầu vào.

Cụ thể, SVM có thể ánh xạ địa chỉ vectơ quyết định sang không gian đặc điểm đa chiều bằng cách chọn hàm hạch đa thức đa lũy thừa, và biểu thức là như sau:

$$\kappa(x, x_i) = \left((x \square x_i) + 1 \right)^d$$

SVM được sử dụng để tìm siêu mặt phẳng phân loại tối ưu cho mỗi loại của các đặc điểm mẫu và các đặc điểm khác trong không gian đặc điểm đa chiều của địa chỉ mẫu, tập hợp vectơ hỗ trợ thể hiện các đặc điểm mẫu khác nhau và độ tin cậy VC tương ứng của nó được thu, và hàm phân biệt để xác định loại của mỗi đặc điểm được tạo ra:

$$f(x) = \sum_{i=1}^n \alpha_i y_i \kappa(x, x_i)$$

Tham số mô hình SVM α_i được thu bằng cách huấn luyện dựa trên một lượng lớn dữ liệu địa chỉ chẳng hạn như dữ liệu thư viện địa chỉ toàn cầu, để tối ưu hóa hơn nữa mô hình máy vectơ hỗ trợ này.

Bằng cách học so khớp địa chỉ dựa trên việc nhận dạng ngữ nghĩa, kết quả xác

định phân loại của địa chỉ rủi ro được thu theo kết quả so khớp của mỗi từ địa chỉ và ngữ nghĩa trước và sau nó, mà kết quả này có thể nhận dạng một cách hiệu quả các địa chỉ rủi ro hoặc địa chỉ rủi ro giả mạo, và có thể tránh được việc đánh giá sai các địa chỉ không có rủi ro.

Dựa trên cùng ý tưởng, các phương án của sáng chế còn đề xuất thiết bị nhận dạng địa chỉ rủi ro. Fig.5 là sơ đồ cấu trúc giản lược của thiết bị nhận dạng địa chỉ rủi ro tương ứng với Fig.2 được đề xuất trong một phương án của sáng chế. Cấu trúc như được thể hiện trên Fig.5 có thể cụ thể bao gồm:

môđun nhận 101 để thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào;

môđun so khớp 102 để xác định, trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ này, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro;

môđun tạo 103 để tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ xác định được;

môđun gán nhãn 104 để xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ; và

môđun nhận dạng 105 để nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

Sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ và mô hình máy vector hỗ trợ, kết quả xác định phân loại của địa chỉ đầu vào được thu theo từ địa chỉ thu được sau khi xử lý địa chỉ đầu vào và ngữ nghĩa trước và sau nó, mà kết quả này có thể nhận dạng một cách hiệu quả các địa chỉ rủi ro hoặc địa chỉ rủi ro giả mạo, và có thể tránh được việc đánh giá sai các địa chỉ không có rủi ro. Do đó, các vấn đề trong giải pháp kỹ thuật có trước có thể được giải quyết một phần hoặc hoàn toàn.

Hơn nữa, việc môđun nhận 101 thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào có thể cụ thể bao gồm việc:

môđun nhận 101 nhận địa chỉ đầu vào; và thực hiện việc xử lý làm sạch dữ liệu và việc xử lý phân đoạn từ trên địa chỉ đầu vào này để thu chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào này. Thông qua thêm việc xử lý làm sạch và việc xử lý phân đoạn từ

trên địa chỉ đầu vào, chuỗi từ địa chỉ được tiêu chuẩn hóa được thu, để tạo thuận lợi cho hoạt động gán nhãn trên chuỗi từ địa chỉ này trong các bước tiếp theo, điều này có thể cải thiện hiệu quả của việc gán nhãn xác suất của các từ địa chỉ xác định được trong chuỗi từ địa chỉ này.

Hơn nữa, việc môđun so khớp 102 xác định, trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ này, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro có thể cụ thể bao gồm việc:

môđun so khớp 102 so khớp tương ứng các từ địa chỉ khác nhau trong chuỗi từ địa chỉ, sử dụng các từ rủi ro tương ứng với mỗi địa chỉ rủi ro; và

nếu một từ địa chỉ trong số các từ địa chỉ này được so khớp thành công thì gán nhãn nó và xác định nó là từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro.

Hơn nữa, việc môđun so khớp 102 xác định, trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ này, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro có thể còn bao gồm việc:

nếu không có từ địa chỉ nào được so khớp thành công thì xác định rằng địa chỉ đầu vào không phải là địa chỉ rủi ro.

Bằng cách gán nhãn từ địa chỉ thông qua môđun so khớp 102, việc lọc trước nhanh chóng địa chỉ đầu vào rủi ro và địa chỉ đầu vào không có rủi ro có thể được thực hiện, điều này có thể cải thiện hiệu quả làm việc của việc nhận dạng địa chỉ rủi ro.

Hơn nữa, việc môđun tạo 103 tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ xác định được có thể cụ thể bao gồm việc:

đối với từ địa chỉ xác định được, thực hiện tương ứng như sau: tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ theo từ địa chỉ này và các từ được liên kết của từ địa chỉ này trong chuỗi từ địa chỉ, trong đó các từ được liên kết này phản ánh ngữ nghĩa trước và sau từ địa chỉ này trong chuỗi từ địa chỉ. Ở đây, ngữ nghĩa trước và sau từ địa chỉ đề cập đến nhiều từ địa chỉ trước và sau, và được liên kết với, từ địa chỉ trùng khớp, và số lượng từ địa chỉ được liên kết này có thể là hai hoặc ba, trong đó các từ địa chỉ được liên kết này có thể là các từ địa chỉ được liên kết liên tục với từ địa chỉ trùng khớp và cũng có thể là các từ địa chỉ được liên kết và đặt cách khỏi từ địa chỉ trùng khớp.

Hơn nữa, việc thu mô hình Markov ẩn dựa trên việc học ngữ nghĩa trước và sau

các từ địa chỉ bao gồm việc:

trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước, và thiết lập mô hình ban đầu chứa các tham số mô hình Markov ẩn theo các tham số ban đầu này, trong đó các mẫu huấn luyện này là các địa chỉ rủi ro hoặc các địa chỉ không rủi ro;

tạo ra các chuỗi quan sát tương ứng với các mẫu huấn luyện theo các từ địa chỉ được chứa trong các mẫu huấn luyện và ngữ nghĩa trước và sau các từ địa chỉ này; và

huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn.

Đối với mô hình Markov ẩn này, chuỗi quan sát bao gồm từ rủi ro trùng khớp và ngữ nghĩa trước và sau từ rủi ro này được sử dụng để huấn luyện các tham số mô hình Markov ẩn, để thu mô hình Markov ẩn (HMM) cần thiết, điều này có thể cải thiện độ chính xác của việc nhận dạng rủi ro địa chỉ đầu vào bởi mô hình Markov ẩn này.

Hơn nữa, các tham số ban đầu bao gồm: vectơ xác suất ban đầu và ma trận chuyển tiếp trạng thái; và

việc trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước có thể cụ thể bao gồm việc:

dựa trên nhiều mẫu huấn luyện, thu vectơ xác suất ban đầu bằng cách thực hiện việc gán nhãn xác suất tương ứng trên các từ địa chỉ được chứa trong các mẫu huấn luyện; và thu ma trận chuyển tiếp trạng thái của các địa chỉ mẫu theo các xác suất chuyển tiếp trạng thái, giữa từ rủi ro và từ không rủi ro, của các từ địa chỉ được chứa trong các mẫu huấn luyện. Ở đây, từ rủi ro là từ địa chỉ được chứa trong địa chỉ rủi ro, từ không rủi ro là từ địa chỉ được chứa trong địa chỉ không rủi ro, và một số từ địa chỉ có thể là cả từ rủi ro lẫn từ không rủi ro.

Hơn nữa, việc huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn có thể cụ thể bao gồm việc:

theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện, sử dụng thuật toán Baum-Welch để huấn luyện các tham số mô hình Markov ẩn để thu mô hình Markov ẩn.

Cần lưu ý rằng trong quá trình lập mô hình của mô hình Markov ẩn, số lượng và

quy mô của các địa chỉ mẫu huấn luyện sẽ ảnh hưởng trực tiếp đến chất lượng của kết quả huấn luyện. Khi thư viện địa chỉ toàn cầu được sử dụng cho các địa chỉ mẫu huấn luyện là toàn diện hơn và có phạm vi bao phủ rộng hơn, tỷ lệ nhận dạng trong khi sử dụng mô hình này sẽ được cải thiện rất nhiều, và do đó, có thể đạt được hiệu quả lập mô hình rất tốt trong trường hợp có đủ địa chỉ mẫu huấn luyện.

Hơn nữa, việc môđun gán nhãn 104 xử lý chuỗi quan sát, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định có thể cụ thể bao gồm việc:

xử lý chuỗi quan sát, sử dụng mô hình Markov ẩn thu được bằng cách học ngữ nghĩa trước và sau các từ địa chỉ, và thuật toán Viterbi, để thu vector quyết định. Vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ.

Hơn nữa, việc môđun nhận dạng 105 đưa ra quyết định phân loại trên vector quyết định có thể cụ thể bao gồm việc:

đưa ra quyết định phân loại trên vector quyết định, sử dụng mô hình máy vector hỗ trợ (SVM) thu được thông qua việc huấn luyện.

Hơn nữa, việc thu mô hình máy vector hỗ trợ thông qua việc huấn luyện bao gồm việc:

thu nhận các mẫu huấn luyện cho máy vector hỗ trợ;

ánh xạ các mẫu huấn luyện của máy vector hỗ trợ này sang không gian đặc điểm đa chiều, để thu không gian đặc điểm mẫu tương ứng với các mẫu huấn luyện của máy vector hỗ trợ này;

thu nhận, từ không gian đặc điểm mẫu này, các tham số thể hiện các đặc điểm mẫu, và thiết lập hàm phân biệt để xác định loại của các đặc điểm mẫu này theo các tham số của các đặc điểm mẫu này; và

huấn luyện các tham số mô hình SVM tương ứng trong hàm phân biệt này dựa trên các mẫu huấn luyện của SVM, để thu mô hình SVM.

Các mẫu huấn luyện có thể được hiểu là vector quyết định trong các phương án được mô tả trên đây, hoặc dữ liệu cần được phân biệt khác định rõ đặc trưng của địa chỉ đầu vào.

Nhìn chung, đối với vector quyết định được tính toán thông qua mô hình Markov

ẩn, SVM cần được sử dụng để ánh xạ vector quyết định này sang không gian đặc điểm đa chiều và sau đó đưa ra quyết định phân loại nhị phân. Tất nhiên, quyết định phân loại có thể được đưa ra trên một số vector quyết định, mà chúng là dễ xử lý, mà không cần ánh xạ các vector quyết định này sang không gian đặc điểm đa chiều, ví dụ, quyết định phân loại tuyến tính có thể được đưa ra, mà điều này có thể làm giảm độ khó tính toán và thúc đẩy tốc độ xử lý.

Dựa trên cùng ý tưởng, các phương án của sáng chế còn đề xuất thiết bị điện tử, bao gồm:

ít nhất một bộ xử lý; và

bộ nhớ kết nối truyền thông với ít nhất một bộ xử lý này, trong đó

bộ nhớ này lưu trữ các lệnh có thể thực thi được bởi ít nhất một bộ xử lý này, các lệnh này được thực thi bởi ít nhất một bộ xử lý này để cho phép ít nhất một bộ xử lý này:

thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào;

xác định, trong các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ này, từ địa chỉ trùng khớp với từ rủi ro tương ứng với mỗi địa chỉ rủi ro;

tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ được xác định này;

xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được trùng khớp bởi các từ địa chỉ khác nhau được chứa trong chuỗi từ địa chỉ; và

nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

Để giải thích rõ hơn về khái niệm có tính mới của một hoặc nhiều phương án của sáng chế, hai phương án được liệt kê cụ thể dưới đây:

Fig.6 là lưu đồ sơ lược trong đó việc nhận dạng địa chỉ rủi ro chỉ ra rằng địa chỉ không có rủi ro, được đề xuất trong một phương án của sáng chế, mà nó có thể cụ thể bao gồm việc:

đầu tiên, văn bản của địa chỉ đầu vào là: REPUBLICA DE SIRIA 7495 SANTA FE Santa Fe AR;

sau đó, quét toàn cục địa chỉ đầu vào này và gán nhãn địa chỉ đầu vào này là trùng khớp với từ địa chỉ được thừa nhận (quốc gia hoặc thành phố): SIRIA, và thu vector gán nhãn ban đầu $[0, 0, 1, 0, 0, 0, 0, 0, 0]$ của địa chỉ đầu vào, trong đó vector gán nhãn ban đầu này có thể được hiểu là vector xác suất ban đầu;

hơn nữa, thu vector quyết định $[0,01; 0,02; 0,02; 0,02; 0,01; 0; 0; 0; 0]$ thông qua mô hình gán nhãn Markov ẩn; và

cuối cùng, mô hình phân loại SVM đưa ra quyết định rằng: địa chỉ đầu vào không trùng khớp với địa chỉ được thừa nhận, trong đó địa chỉ được thừa nhận được mô tả ở đây có thể được hiểu là địa chỉ rủi ro.

Fig.7 là lưu đồ sơ lược trong đó việc nhận dạng địa chỉ rủi ro chỉ ra rằng địa chỉ là rủi ro, được đề xuất trong một phương án của sáng chế, mà nó có thể cụ thể bao gồm việc:

đầu tiên, văn bản của địa chỉ đầu vào là: Rostovskaya, 31a, Simferopol 5 Other RU;

sau đó, quét toàn cục địa chỉ đầu vào này và gán nhãn địa chỉ đầu vào này là trùng khớp với từ địa chỉ được thừa nhận (quốc gia hoặc thành phố): Simferopol, và thu vector gán nhãn ban đầu $[0, 0, 1, 0, 0, 0]$ của địa chỉ đầu vào;

hơn nữa, thu được vector quyết định $[0,9; 0,9; 0,9; 0,9; 0,5; 0,1]$ thông qua mô hình gán nhãn Markov ẩn; và

cuối cùng, mô hình phân loại SVM đưa ra quyết định rằng: địa chỉ đầu vào trùng khớp địa chỉ được thừa nhận,

trong đó địa chỉ được thừa nhận được mô tả ở đây có thể được hiểu là địa chỉ rủi ro được mô tả trên đây.

Fig.8 là lưu đồ sơ lược của việc lập mô hình và nhận dạng của việc nhận dạng địa chỉ rủi ro được đề xuất trong một phương án của sáng chế, mà nó có thể cụ thể bao gồm việc:

khi lập mô hình cho mô hình nhận dạng địa chỉ rủi ro, thực hiện việc huấn luyện tham số, và trước tiên, thu nhận tuyển tập địa chỉ huấn luyện, mà nó có thể được hiểu là các địa chỉ mẫu;

sau đó, làm sạch tuyển tập địa chỉ huấn luyện này và tách chúng thành các chuỗi ký tự địa chỉ được tiêu chuẩn hóa, mà chúng có thể được hiểu là chuỗi từ địa chỉ;

sau đó, quét toàn cục, so khớp và gán nhãn tuyến tập địa chỉ huấn luyện đầu vào này, để thu mô hình gán nhãn ban đầu và các tham số ban đầu.

Hơn nữa, theo mô hình gán nhãn ban đầu này, các tham số mô hình Markov ẩn được huấn luyện để thu mô hình Markov ẩn cần thiết. Lại thêm nữa, các tham số của SVM được huấn luyện theo vector quyết định được xuất ra qua mô hình Markov ẩn, để cuối cùng thu mô hình SVM cần thiết.

Khi thực hiện việc quét và nhận dạng địa chỉ rủi ro, trước tiên người dùng nhập địa chỉ; sau đó, việc xử lý tiêu chuẩn hóa và quét toàn cục địa chỉ rủi ro và gán nhãn được thực hiện trên địa chỉ đầu vào này để thu chuỗi từ địa chỉ được gán nhãn, và hơn nữa, chuỗi quan sát được thu dựa trên từ rủi ro và ngữ nghĩa trước và sau nó.

Lại thêm nữa, việc tính toán xác suất rủi ro và gán nhãn được thực hiện trên chuỗi từ địa chỉ (chuỗi quan sát) này thông qua mô hình Markov ẩn để tạo ra vector quyết định.

Cuối cùng, SVM đưa ra quyết định phân loại nhị phân theo vector quyết định này để xác định xem liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không.

Mô hình Markov ẩn và máy vector hỗ trợ được sử dụng để đánh giá và nhận dạng địa chỉ đầu vào dựa trên từ địa chỉ rủi ro và ngữ nghĩa trước và sau nó, và độ chính xác của việc nhận dạng có thể được cải thiện một cách hiệu quả.

Các phương án cụ thể của sáng chế đã được mô tả trên đây. Các phương án khác nằm trong phạm vi của các yêu cầu bảo hộ kèm theo. Trong một số trường hợp, các hành động hoặc các bước được chỉ rõ trong các yêu cầu bảo hộ có thể được thực hiện theo thứ tự khác với các hành động hoặc các bước trong các phương án và vẫn có thể đạt được các kết quả mong muốn. Thêm vào đó, các quy trình được mô tả trong các hình vẽ không nhất thiết phải theo một thứ tự cụ thể hoặc thứ tự liên tiếp như được thể hiện để đạt được các kết quả mong muốn. Trong một số cách thi hành, việc xử lý đa nhiệm và xử lý song song cũng là khả thi hoặc có thể là ưu điểm.

Các phương án khác nhau trong phần mô tả đều được mô tả theo cách lũy tiến. Đối với các phần giống nhau hoặc tương tự nhau giữa các phương án, tham chiếu có thể được thực hiện cho nhau. Đối với mỗi phương án, phần của nó khác với các phương án khác được mô tả chủ yếu. Cụ thể, các phương án về thiết bị, thiết bị điện tử, phương tiện lưu trữ máy tính phi khả biến được mô tả theo cách tương đối đơn giản vì chúng về cơ bản là tương tự như các phương án về phương pháp, và đối với các phần liên quan, có

thể tham chiếu đến các phần của phần mô tả trong các phương án về phương pháp.

Thiết bị, thiết bị điện tử và phương tiện lưu trữ máy tính phi khả biến được đề xuất trong các phương án của sáng chế tương ứng với phương pháp. Do đó, thiết bị, thiết bị điện tử và phương tiện lưu trữ máy tính phi khả biến này cũng có các hiệu ứng kỹ thuật có lợi tương tự như của phương pháp tương ứng. Do các hiệu ứng kỹ thuật có lợi của phương pháp đã được mô tả chi tiết ở trên, các hiệu ứng kỹ thuật có lợi của thiết bị, thiết bị điện tử và phương tiện lưu trữ máy tính phi khả biến tương ứng sẽ không được mô tả ở đây.

Trong những năm 1990, sự cải thiện về công nghệ có thể được phân biệt một cách rõ ràng thành sự cải thiện trên phần cứng (ví dụ, sự cải thiện trên cấu trúc mạch chẳng hạn như điốt, tranzito, và bộ chuyển mạch) hoặc sự cải thiện trên phần mềm (sự cải thiện trên thủ tục phương pháp). Tuy nhiên, với sự phát triển của các công nghệ khác nhau, những sự cải thiện về nhiều thủ tục phương pháp hiện nay có thể được xem là các cải thiện trực tiếp về các cấu trúc mạch phần cứng. Hầu hết tất cả các nhà thiết kế đều lập trình các thủ tục phương pháp được cải thiện vào trong các mạch phần cứng để thu được các cấu trúc mạch phần cứng tương ứng. Do đó, không thể nói rằng sự cải thiện về thủ tục phương pháp không thể được thi hành qua việc sử dụng mô đun thực thể phần cứng. Ví dụ, thiết bị logic lập trình được (Programmable Logic Device, PLD) (ví dụ, mảng cổng lập trình được dạng trường (Field Programmable Gate Array, FPGA)) là mạch tích hợp như vậy, và các chức năng logic của nó được xác định bởi thiết bị lập trình bởi người dùng. Các nhà thiết kế tự lập trình để “tích hợp” hệ thống kỹ thuật số vào mảnh PLD, mà không cần mời các nhà sản xuất chip thiết kế và sản xuất chip mạch tích hợp chuyên dụng. Hơn nữa, hiện tại, việc lập trình chủ yếu được thi hành bằng cách sử dụng phần mềm trình “biên dịch logic”, thay vì sản xuất thủ công một chip mạch tích hợp. Phần mềm “trình biên dịch logic” này là tương tự như trình biên dịch phần mềm được sử dụng để phát triển và viết chương trình, và mã gốc trước khi biên dịch cũng cần được viết bằng ngôn ngữ lập trình cụ thể, được gọi là ngôn ngữ mô tả phần cứng (Hardware Description Language, HDL). Có nhiều loại HDL, chẳng hạn như ABEL (ngôn ngữ biểu thức luận lý cải tiến, Advanced Boolean Expression Language), AHDL (ngôn ngữ mô tả phần cứng của Altera, Altera Hardware Description Language), Confluence (Hợp lưu), CUPL (ngôn ngữ lập trình của đại học Cornell, Cornell University Programming

Language), HDCL, JHDL (ngôn ngữ mô tả phần cứng Java, Java Hardware Description Language), Lava, Lola, MyHDL, PALASM, và RHDL (ngôn ngữ mô tả phần cứng Ruby, Ruby Hardware Description Language), trong số đó, VHDL (ngôn ngữ mô tả phần cứng mạch tích hợp tốc độ rất cao, Very-High-Speed Integrated Circuit Hardware Description Language) và Verilog hiện đang được sử dụng phổ biến nhất. Những người có hiểu biết trung bình trong lĩnh vực kỹ thuật cũng nên biết rằng mạch phần cứng để thi hành thủ tục phương pháp logic có thể dễ dàng được thu chỉ bằng cách lập trình logic một cách hợp lý thủ tục phương pháp sử dụng một số ngôn ngữ mô tả phần cứng được mô tả trên đây và lập trình thủ tục phương pháp này vào trong mạch tích hợp.

Bộ điều khiển có thể được thi hành theo cách thích hợp bất kỳ. Ví dụ, bộ điều khiển này có thể ở dạng bộ vi xử lý hoặc bộ xử lý, và phương tiện có thể đọc bằng máy tính lưu trữ các mã chương trình có thể đọc bằng máy tính (ví dụ, phần mềm hoặc phần sụn) có thể thực thi được bởi bộ (vi) xử lý, cổng logic, bộ chuyển mạch, mạch tích hợp chuyên dụng (Application Specific Integrated Circuit, ASIC), bộ điều khiển logic lập trình được, và bộ vi điều khiển được nhúng. Các ví dụ về bộ điều khiển bao gồm, nhưng không bị giới hạn ở, các bộ vi điều khiển dưới đây: ARC 625D, Atmel AT91SAM, Microchip PIC18F26K20, và Silicone Labs C8051F320. Bộ điều khiển bộ nhớ cũng có thể được thi hành dưới dạng một phần của bộ logic điều khiển của bộ nhớ. Những người có hiểu biết trung bình trong lĩnh vực kỹ thuật cũng biết rằng, ngoài việc thi hành bộ điều khiển qua các mã chương trình đọc được bằng máy tính thuần túy, các bước của phương pháp có thể được lập trình logic để cho phép bộ điều khiển thi hành chức năng tương tự dưới dạng cổng logic, bộ chuyển mạch, mạch tích hợp chuyên dụng, bộ điều khiển logic lập trình được và bộ vi điều khiển được nhúng. Do đó, bộ điều khiển như vậy có thể được xem là thành phần phần cứng, và các thiết bị được chứa trong đó để thi hành các chức năng khác nhau cũng có thể được xem là các cấu trúc bên trong thành phần phần cứng này. Theo cách khác, các thiết bị để thi hành các chức năng khác nhau này thậm chí có thể được coi là cả các môđun phần mềm để thi hành phương pháp này và các cấu trúc bên trong các thành phần phần cứng này.

Hệ thống, các thiết bị, các môđun hoặc các đơn vị được minh họa trong các phương án nêu trên có thể được thi hành bởi chip máy tính hoặc thực thể, hoặc được thi hành bởi sản phẩm có chức năng cụ thể. Thiết bị thi hành thông thường là máy tính. Ví dụ,

máy tính này có thể là máy tính cá nhân, máy tính xách tay, điện thoại tế bào, điện thoại camera, điện thoại thông minh, thiết bị trợ giúp số cá nhân, thiết bị trình diễn phương tiện, thiết bị định vị, thiết bị thư điện tử, bảng điều khiển trò chơi, máy tính bảng, thiết bị đeo được, hoặc sự kết hợp của các thiết bị bất kỳ trong số các thiết bị này.

Để dễ mô tả, thiết bị nêu trên được chia thành các môđun khác nhau dựa trên các chức năng, và các môđun này được mô tả riêng biệt. Tất nhiên, khi một hoặc nhiều phương án của sáng chế được thi hành, các chức năng của các đơn vị khác nhau có thể được thi hành trong một hoặc nhiều đoạn phần mềm và/hoặc phần cứng.

Những người có hiểu biết trung bình trong lĩnh vực cần hiểu rằng các phương án của sáng chế có thể được đề xuất dưới dạng phương pháp, hệ thống, hoặc sản phẩm chương trình máy tính. Do đó, các phương án của sáng chế có thể được thi hành dưới dạng các phương án hoàn toàn là phần cứng, các phương án hoàn toàn là phần mềm, hoặc các phương án kết hợp phần mềm và phần cứng. Hơn nữa, các phương án của sáng chế có thể ở dạng sản phẩm chương trình máy tính được thi hành trên một hoặc nhiều phương tiện lưu trữ sử dụng được bằng máy tính (bao gồm, nhưng không bị giới hạn ở, bộ nhớ đĩa từ, CD-ROM, bộ nhớ quang và dạng tương tự) bao gồm các mã chương trình sử dụng được bằng máy tính trong đó.

Sáng chế được mô tả với tham chiếu đến các lưu đồ và/hoặc các sơ đồ khối của phương pháp, thiết bị (hệ thống) và sản phẩm chương trình máy tính theo các phương án của sáng chế. Cần hiểu rằng các lệnh chương trình máy tính có thể thi hành mỗi quy trình và/hoặc mỗi khối trong các lưu đồ và/hoặc các sơ đồ khối này và những sự kết hợp của các quy trình và/hoặc các khối trong các lưu đồ và/hoặc các sơ đồ khối này. Các lệnh chương trình máy tính này có thể được cung cấp cho máy tính đa năng, máy tính chuyên dụng, bộ xử lý được nhúng, hoặc bộ xử lý của các thiết bị xử lý dữ liệu lập trình được khác để tạo ra bộ máy, sao cho máy để thi hành chức năng được chỉ rõ trong một hoặc nhiều quy trình trong các lưu đồ và/hoặc trong một hoặc nhiều khối trong các sơ đồ khối được tạo ra bởi các lệnh được thực thi bởi bộ xử lý của máy tính này hoặc các thiết bị xử lý dữ liệu lập trình được khác này.

Các lệnh chương trình máy tính này cũng có thể được lưu trữ trong bộ nhớ đọc được bằng máy tính mà nó có thể hướng dẫn máy tính hoặc thiết bị xử lý dữ liệu lập trình được khác vận hành theo cách cụ thể, sao cho các lệnh được lưu trữ trong bộ nhớ

đọc được bằng máy tính này tạo ra sản phẩm, bao gồm thiết bị ra lệnh. Thiết bị ra lệnh này thi hành chức năng được chỉ rõ trong một hoặc nhiều quy trình trong các lưu đồ và/hoặc trong một hoặc nhiều khối trong các sơ đồ khối này.

Các lệnh chương trình máy tính này cũng có thể được tải lên trên máy tính hoặc các thiết bị xử lý dữ liệu lập trình được khác, sao cho chuỗi các bước hoạt động được thực hiện trên máy tính hoặc các thiết bị xử lý dữ liệu lập trình được khác này, do đó tạo ra việc xử lý được thi hành bởi máy tính. Do đó, các lệnh được thực thi trên máy tính hoặc các thiết bị lập trình được khác này cung cấp các bước để thi hành chức năng được chỉ rõ trong một hoặc nhiều quy trình trong các lưu đồ và/hoặc trong một hoặc nhiều khối trong các sơ đồ khối này.

Theo cấu hình điển hình, thiết bị tính toán bao gồm một hoặc nhiều đơn vị xử lý trung tâm (central processing unit, CPU), giao diện đầu vào/đầu ra, giao diện mạng, và bộ nhớ.

Bộ nhớ này có thể bao gồm phương tiện đọc được bằng máy tính chẳng hạn như bộ nhớ phi vĩnh cửu, bộ nhớ truy cập ngẫu nhiên (RAM), và/hoặc bộ nhớ phi khả biến, ví dụ, bộ nhớ chỉ đọc (ROM) hoặc RAM tác động nhanh (flash RAM). Bộ nhớ này là một ví dụ về phương tiện đọc được bằng máy tính.

Phương tiện đọc được bằng máy tính bao gồm các phương tiện vĩnh cửu và phi vĩnh cửu cũng như các phương tiện có thể và không thể tháo ra được, và có thể lưu trữ thông tin theo phương pháp hoặc công nghệ bất kỳ. Thông tin này có thể là lệnh có thể đọc được bằng máy tính, cấu trúc dữ liệu, môđun của chương trình, hoặc dữ liệu khác. Một ví dụ về phương tiện lưu trữ của máy tính bao gồm, nhưng không bị giới hạn ở, bộ nhớ thay đổi pha (phase change memory, PRAM), bộ nhớ truy cập ngẫu nhiên tĩnh (static random access memory, SRAM), bộ nhớ truy cập ngẫu nhiên động (dynamic random access memory, DRAM), các loại bộ nhớ truy cập ngẫu nhiên (random access memory, RAM) khác, bộ nhớ chỉ đọc (read-only memory, ROM), bộ nhớ chỉ đọc lập trình được xóa được bằng điện (electrically erasable programmable read-only memory, EEPROM), bộ nhớ tác động nhanh (flash memory) hoặc các công nghệ bộ nhớ khác, bộ nhớ chỉ đọc dạng đĩa compac (compact disk read-only memory, CD-ROM), đĩa đa năng số (digital versatile disc, DVD) hoặc các bộ lưu trữ quang học khác, băng cát xet, bộ lưu trữ băng từ/đĩa từ hoặc các thiết bị lưu trữ từ tính khác, hoặc phương tiện không truyền

dẫn khác bất kỳ, và có thể được sử dụng để lưu trữ thông tin truy cập được vào thiết bị tính toán. Theo định nghĩa ở đây, phương tiện đọc được bằng máy tính không bao gồm phương tiện chuyển tiếp đọc được bằng máy tính, chẳng hạn như các tín hiệu dữ liệu được điều biến và các sóng mang.

Cần lưu ý thêm rằng các thuật ngữ "bao gồm", "gồm" hoặc các biến thể khác bất kỳ của chúng được dự định để bao trùm sự bao gồm không loại trừ, sao cho quy trình, phương pháp, vật phẩm hoặc thiết bị bao gồm một chuỗi phần tử không chỉ bao gồm các phần tử đó, mà còn bao gồm các phần tử khác không được liệt kê rõ ràng, hoặc còn bao gồm các phần tử có hữu đối với quy trình, phương pháp, vật phẩm hoặc thiết bị này. Khi không có thêm các giới hạn, một phần tử được định ra bởi "bao gồm một ..." không loại trừ việc quy trình, phương pháp, vật phẩm hoặc thiết bị bao gồm phần tử đó còn có các phần tử giống hệt khác.

Sáng chế có thể được mô tả trong ngữ cảnh thông thường của các lệnh có thể thực thi được bằng máy tính được thực thi bằng máy tính, ví dụ, môđun chương trình. Nhìn chung, môđun chương trình này bao gồm đoạn chương trình, chương trình, đối tượng, thành phần, cấu trúc dữ liệu, và loại tương tự để thực thi một tác vụ cụ thể hoặc thi hành một loại dữ liệu trừu tượng cụ thể. Sáng chế cũng có thể được thực hành trong các môi trường tính toán phân tán. Trong các môi trường tính toán phân tán này, tác vụ được thực thi bởi thiết bị xử lý từ xa được kết nối qua mạng truyền thông. Trong các môi trường tính toán phân tán này, môđun chương trình có thể được đặt trong phương tiện lưu trữ máy tính cục bộ và từ xa, bao gồm thiết bị lưu trữ.

Các phương án khác nhau trong phần mô tả đều được mô tả theo cách lũy tiến. Đối với các phần giống nhau hoặc tương tự nhau giữa các phương án, tham chiếu có thể được thực hiện cho nhau. Đối với mỗi phương án, phần của nó khác với các phương án khác được mô tả chủ yếu. Cụ thể, các phương án về hệ thống được mô tả theo cách tương đối đơn giản vì chúng về cơ bản là tương tự như các phương án về phương pháp, và đối với các phần liên quan, có thể tham chiếu đến các phần của phần mô tả trong các phương án về phương pháp.

Các phần mô trên đây chỉ đơn thuần là các phương án của sáng chế, và không được sử dụng để giới hạn sáng chế. Đối với người có hiểu biết trung bình trong lĩnh vực tương ứng, sáng chế có thể có các thay đổi và biến thể khác nhau. Sự cải biến, thay thế tương

đương, cải tiến và dạng tương tự bất kỳ được thực hiện trong phạm vi và nguyên lý của sáng chế sẽ được bao hàm trong phạm vi của các yêu cầu bảo hộ của sáng chế.

YÊU CẦU BẢO HỘ

1. Phương pháp nhận dạng địa chỉ rủi ro, bao gồm các bước:

thu nhận chuỗi từ địa chỉ (address word sequence) tương ứng với địa chỉ đầu vào;
xác định từ địa chỉ (address word) trong chuỗi từ địa chỉ này, từ địa chỉ xác định được này khớp với từ rủi ro tương ứng với địa chỉ rủi ro;

tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ xác định được này;

xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được khớp bởi các từ địa chỉ được chứa trong chuỗi từ địa chỉ; và

nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

2. Phương pháp theo điểm 1, trong đó bước thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào bao gồm các bước:

nhận địa chỉ đầu vào;

thực hiện việc xử lý làm sạch dữ liệu và việc xử lý phân đoạn từ trên địa chỉ đầu vào này để thu chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào này.

3. Phương pháp theo điểm 1, trong đó bước xác định từ địa chỉ trong chuỗi từ địa chỉ khớp với từ rủi ro tương ứng với địa chỉ rủi ro bao gồm các bước:

so khớp tương ứng các từ địa chỉ trong chuỗi từ địa chỉ, sử dụng các từ rủi ro tương ứng với địa chỉ rủi ro này; và

nếu một từ địa chỉ trong số các từ địa chỉ này được so khớp thành công thì gán nhãn từ địa chỉ được so khớp này và xác định từ địa chỉ được so khớp này là từ địa chỉ khớp với từ rủi ro tương ứng với địa chỉ rủi ro này.

4. Phương pháp theo điểm 3, trong đó bước xác định từ địa chỉ trong chuỗi từ địa chỉ khớp với từ rủi ro tương ứng với địa chỉ rủi ro còn bao gồm bước:

nếu không có từ địa chỉ nào được so khớp thành công thì xác định rằng địa chỉ đầu vào không phải là địa chỉ rủi ro.

5. Phương pháp theo điểm 3, trong đó bước tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ theo từ địa chỉ xác định được bao gồm bước:

tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ theo từ địa chỉ xác định được và các từ được liên kết của từ địa chỉ xác định được trong chuỗi từ địa chỉ, trong đó các từ được liên kết này phản ánh ngữ nghĩa trước và sau từ địa chỉ xác định được trong chuỗi từ địa chỉ.

6. Phương pháp theo điểm 1, trong đó bước thu mô hình Markov ẩn dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ bao gồm các bước:

trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước, và thiết lập mô hình ban đầu chứa các tham số mô hình Markov ẩn theo các tham số ban đầu này, trong đó các mẫu huấn luyện này là các địa chỉ rủi ro hoặc các địa chỉ không rủi ro;

tạo ra các chuỗi quan sát tương ứng với các mẫu huấn luyện theo các từ địa chỉ được chứa trong các mẫu huấn luyện và ngữ nghĩa trước và sau các từ địa chỉ này; và

huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn.

7. Phương pháp theo điểm 6, trong đó các tham số ban đầu bao gồm: vector xác suất ban đầu và ma trận chuyển tiếp trạng thái; và bước trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước bao gồm các bước:

dựa trên nhiều mẫu huấn luyện được xác định trước, thu vector xác suất ban đầu bằng cách thực hiện việc gán nhãn xác suất tương ứng trên các từ địa chỉ được chứa trong các mẫu huấn luyện này; và

thu ma trận chuyển tiếp trạng thái của các địa chỉ mẫu theo các xác suất chuyển tiếp trạng thái, giữa từ rủi ro và từ không rủi ro, của các từ địa chỉ được chứa trong các mẫu huấn luyện này.

8. Phương pháp theo điểm 6, trong đó bước huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn bao gồm bước:

theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện, sử dụng thuật toán Baum-Welch để huấn luyện các tham số mô hình Markov ẩn để thu mô hình Markov ẩn.

9. Phương pháp theo điểm 1, trong đó bước xử lý chuỗi quan sát, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định bao gồm bước:

xử lý chuỗi quan sát, sử dụng mô hình Markov ẩn thu được bằng việc học ngữ nghĩa trước và sau các từ địa chỉ, và thuật toán Viterbi, để thu vector quyết định.

10. Phương pháp theo điểm 1, trong đó bước đưa ra quyết định phân loại trên vector quyết định bao gồm bước:

đưa ra quyết định phân loại trên vector quyết định, sử dụng mô hình máy vector hỗ trợ thu được thông qua việc huấn luyện.

11. Phương pháp theo điểm 10, trong đó bước thu mô hình máy vector hỗ trợ thông qua việc huấn luyện bao gồm các bước:

thu nhận các mẫu huấn luyện cho máy vector hỗ trợ;

ánh xạ các mẫu huấn luyện của máy vector hỗ trợ này sang không gian đặc điểm đa chiều, để thu không gian đặc điểm mẫu tương ứng với các mẫu huấn luyện của máy vector hỗ trợ này;

thu nhận, từ không gian đặc điểm mẫu này, các tham số thể hiện các đặc điểm mẫu, và thiết lập hàm phân biệt để xác định loại của các đặc điểm mẫu này theo các tham số của các đặc điểm mẫu này; và

huấn luyện các tham số máy vector hỗ trợ tương ứng trong hàm phân biệt này dựa trên các mẫu huấn luyện của máy vector hỗ trợ, để thu mô hình máy vector hỗ trợ.

12. Thiết bị điện tử, thiết bị này bao gồm:

bộ xử lý; và

bộ nhớ để lưu trữ các lệnh có thể thực thi được bởi bộ xử lý này;

trong đó bộ xử lý được tạo cấu hình để:

thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào;

xác định từ địa chỉ trong chuỗi từ địa chỉ này, từ địa chỉ xác định được này khớp với từ rủi ro tương ứng với địa chỉ rủi ro;

tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ xác định được này;

xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vector quyết định, trong đó vector quyết định

này thể hiện các xác suất của việc địa chỉ rủi ro được khớp bởi các từ địa chỉ được chứa trong chuỗi từ địa chỉ; và

nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

13. Thiết bị theo điểm 12, trong đó trong việc thu nhận chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào, bộ xử lý còn được tạo cấu hình để:

nhận địa chỉ đầu vào; và

thực hiện việc xử lý làm sạch dữ liệu và việc xử lý phân đoạn từ trên địa chỉ đầu vào này để thu chuỗi từ địa chỉ tương ứng với địa chỉ đầu vào này.

14. Thiết bị theo điểm 12, trong đó trong việc xác định từ địa chỉ trong chuỗi từ địa chỉ khớp với từ rủi ro tương ứng với địa chỉ rủi ro, bộ xử lý còn được tạo cấu hình để:

so khớp tương ứng các từ địa chỉ trong chuỗi từ địa chỉ, sử dụng các từ rủi ro tương ứng với địa chỉ rủi ro này; và

nếu một từ địa chỉ trong số các từ địa chỉ này được so khớp thành công thì gán nhãn từ địa chỉ được so khớp này và xác định từ địa chỉ được so khớp này là từ địa chỉ khớp với từ rủi ro tương ứng với địa chỉ rủi ro này.

15. Thiết bị theo điểm 14, trong đó trong việc xác định từ địa chỉ trong chuỗi từ địa chỉ khớp với từ rủi ro tương ứng với địa chỉ rủi ro, bộ xử lý còn được tạo cấu hình để:

nếu không có từ địa chỉ nào được so khớp thành công thì xác định rằng địa chỉ đầu vào không phải là địa chỉ rủi ro.

16. Thiết bị theo điểm 14, trong đó trong việc tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ theo từ địa chỉ xác định được, bộ xử lý còn được tạo cấu hình để:

tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ theo từ địa chỉ xác định được và các từ được liên kết của từ địa chỉ xác định được trong chuỗi từ địa chỉ, trong đó các từ được liên kết này phản ánh ngữ nghĩa trước và sau từ địa chỉ xác định được trong chuỗi từ địa chỉ.

17. Thiết bị theo điểm 12, trong đó bộ xử lý còn được tạo cấu hình để:

trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước, và thiết lập mô hình ban đầu chứa các tham số mô hình Markov ẩn theo các tham số ban đầu này, trong đó các mẫu huấn luyện này là các địa chỉ rủi ro hoặc các địa chỉ không rủi ro;

tạo ra các chuỗi quan sát tương ứng với các mẫu huấn luyện theo các từ địa chỉ được chứa trong các mẫu huấn luyện và ngữ nghĩa trước và sau các từ địa chỉ này; và

huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn.

18. Thiết bị theo điểm 17, trong đó các tham số ban đầu bao gồm: vectơ xác suất ban đầu và ma trận chuyển tiếp trạng thái; và trong việc trích xuất các tham số ban đầu theo các mẫu huấn luyện được xác định trước, bộ xử lý còn được tạo cấu hình để:

dựa trên nhiều mẫu huấn luyện được xác định trước, thu vectơ xác suất ban đầu bằng cách thực hiện việc gán nhãn xác suất tương ứng trên các từ địa chỉ được chứa trong các mẫu huấn luyện này; và

thu ma trận chuyển tiếp trạng thái của các địa chỉ mẫu theo các xác suất chuyển tiếp trạng thái, giữa từ rủi ro và từ không rủi ro, của các từ địa chỉ được chứa trong các mẫu huấn luyện này.

19. Thiết bị theo điểm 17, trong đó trong việc huấn luyện các tham số mô hình Markov ẩn theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện để thu mô hình Markov ẩn, bộ xử lý còn được tạo cấu hình để:

theo mô hình ban đầu và các chuỗi quan sát tương ứng với các mẫu huấn luyện, sử dụng thuật toán Baum-Welch để huấn luyện các tham số mô hình Markov ẩn để thu mô hình Markov ẩn.

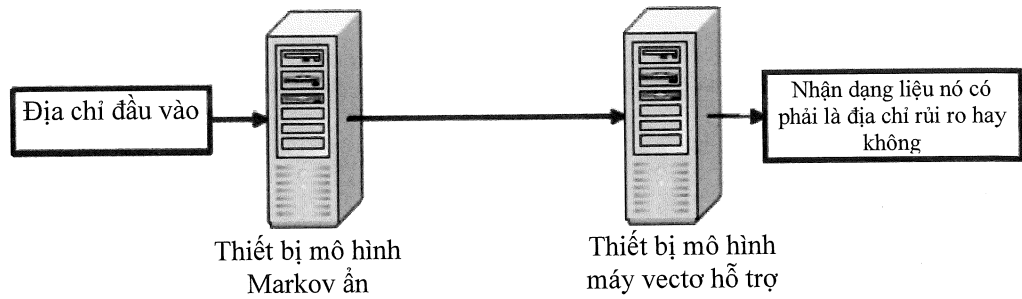
20. Phương tiện lưu trữ đọc được bằng máy tính, lưu trữ trên đó chương trình máy tính mà, khi được thực thi bởi bộ xử lý của thiết bị, khiến cho thiết bị này thực hiện phương pháp nhận dạng địa chỉ rủi ro, phương pháp này bao gồm các bước:

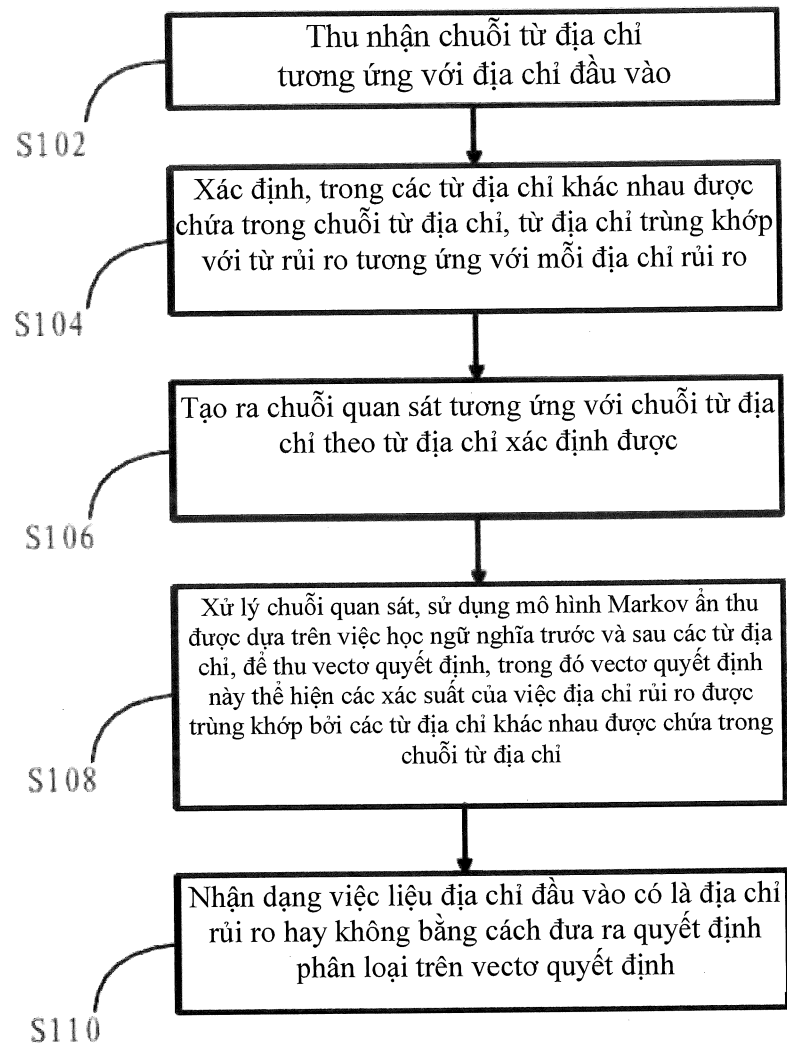
thu nhận chuỗi từ địa chỉ (address word sequence) tương ứng với địa chỉ đầu vào; xác định từ địa chỉ (address word) trong chuỗi từ địa chỉ này, từ địa chỉ xác định được này khớp với từ rủi ro tương ứng với địa chỉ rủi ro;

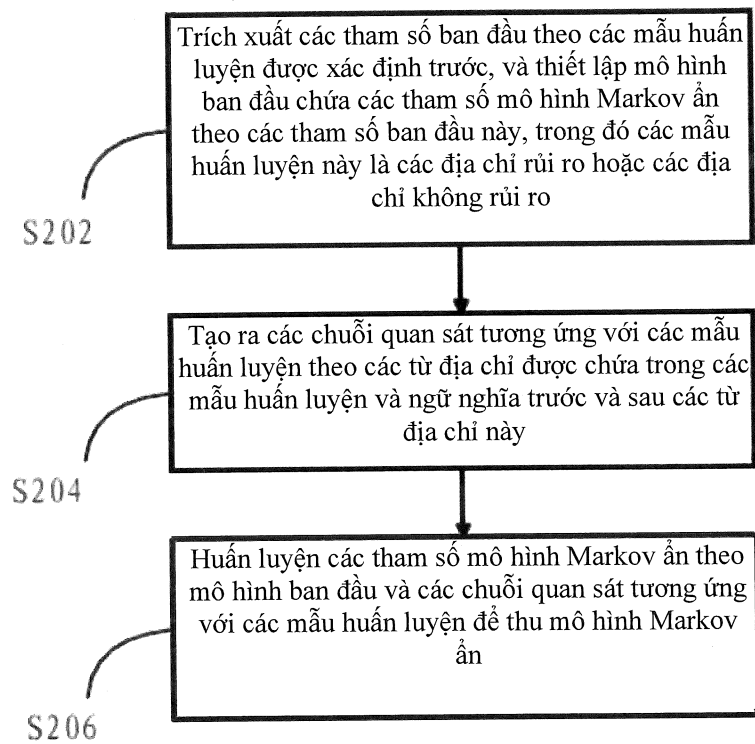
tạo ra chuỗi quan sát tương ứng với chuỗi từ địa chỉ này theo từ địa chỉ xác định được này;

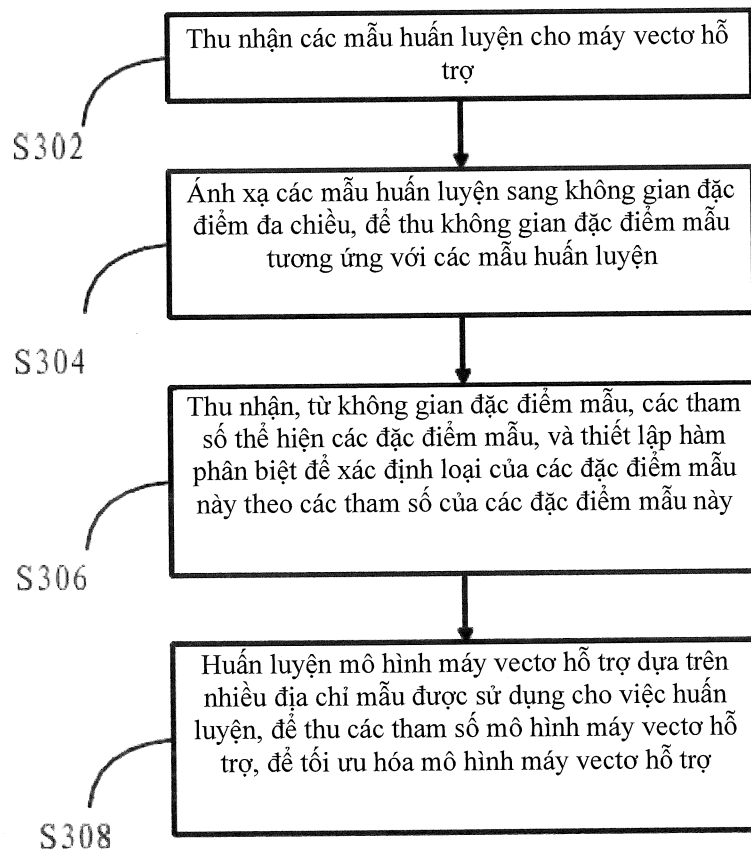
xử lý chuỗi quan sát này, sử dụng mô hình Markov ẩn thu được dựa trên việc học ngữ nghĩa trước và sau các từ địa chỉ, để thu vectơ quyết định, trong đó vectơ quyết định này thể hiện các xác suất của việc địa chỉ rủi ro được khớp bởi các từ địa chỉ được chứa trong chuỗi từ địa chỉ; và

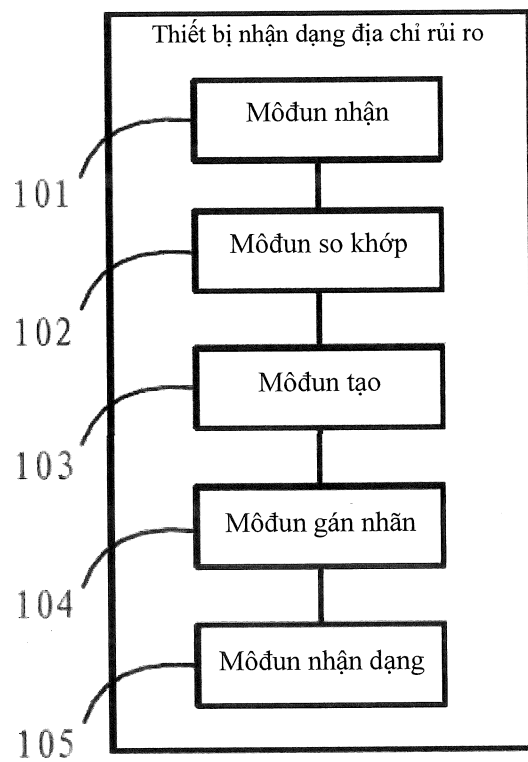
nhận dạng việc liệu địa chỉ đầu vào có phải là địa chỉ rủi ro hay không bằng cách đưa ra quyết định phân loại trên vector quyết định này.

**Fig. 1**

**Fig. 2**

**Fig. 3**

**Fig. 4**

**Fig. 5**

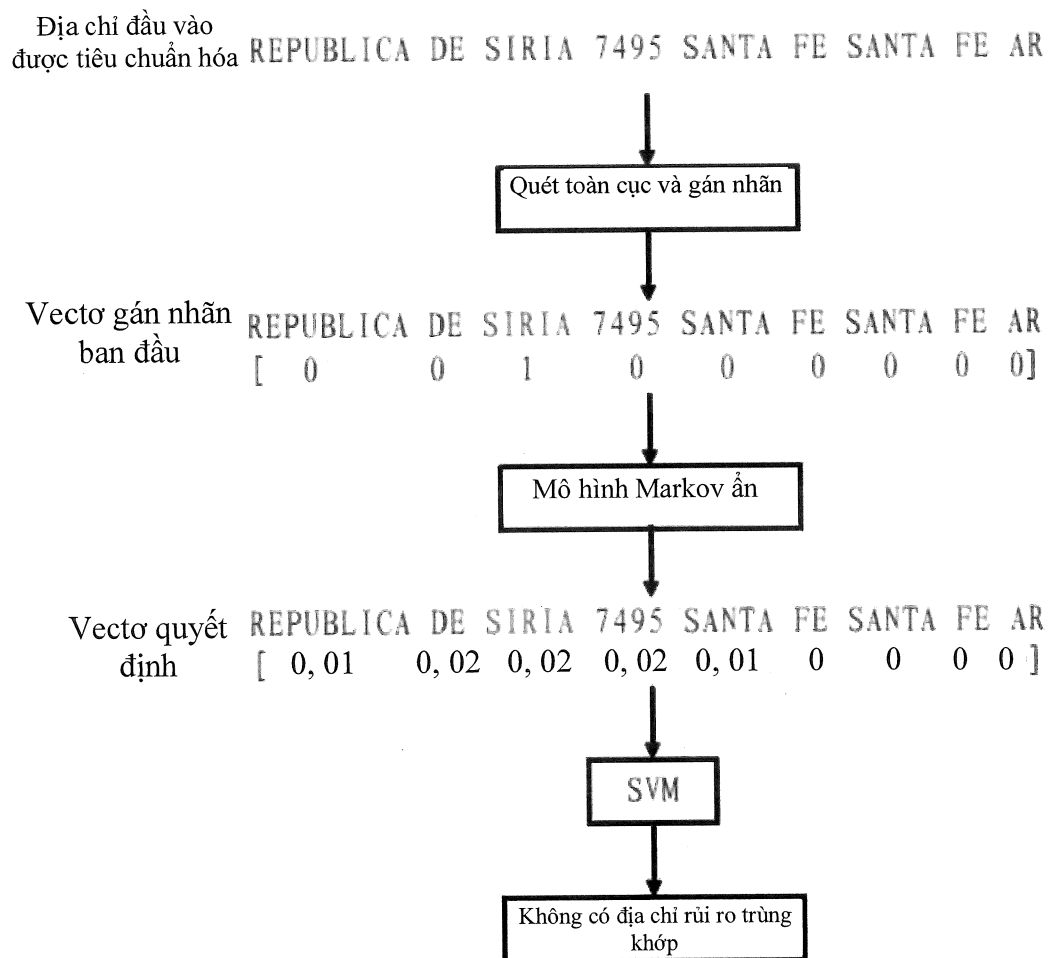
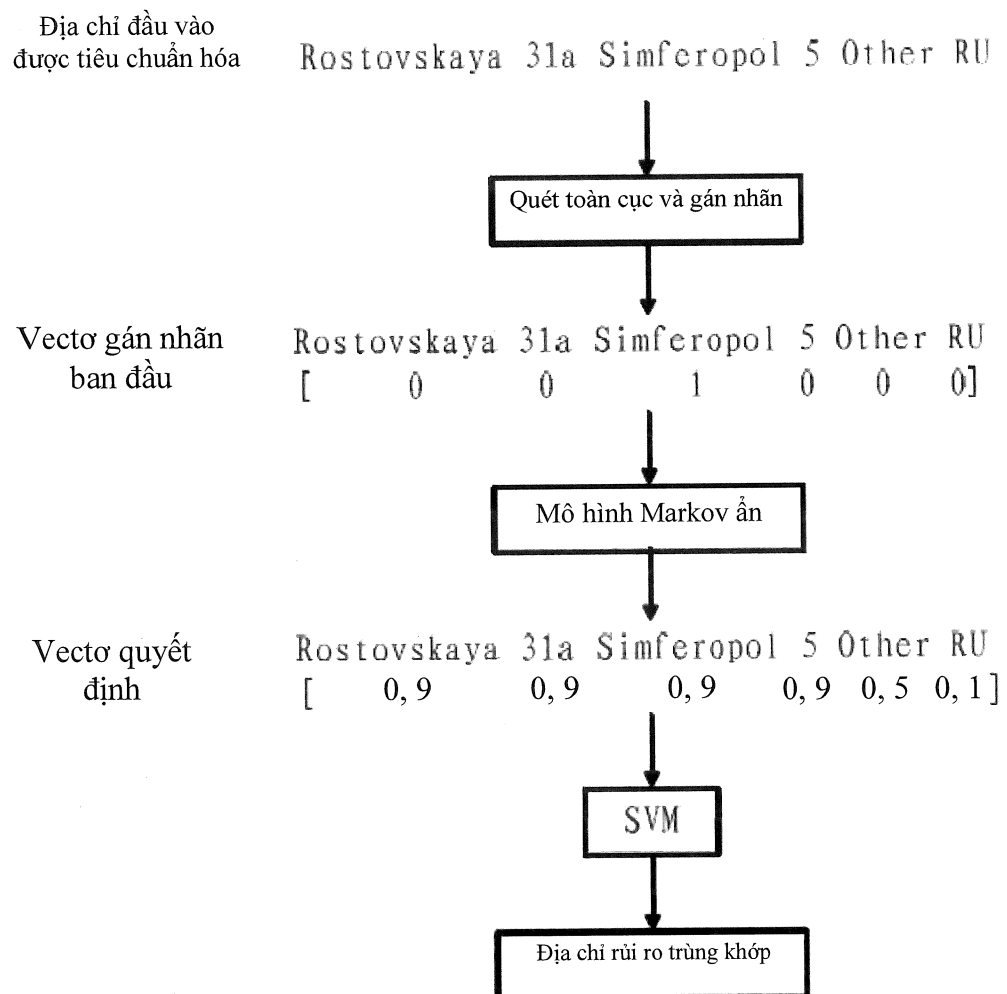
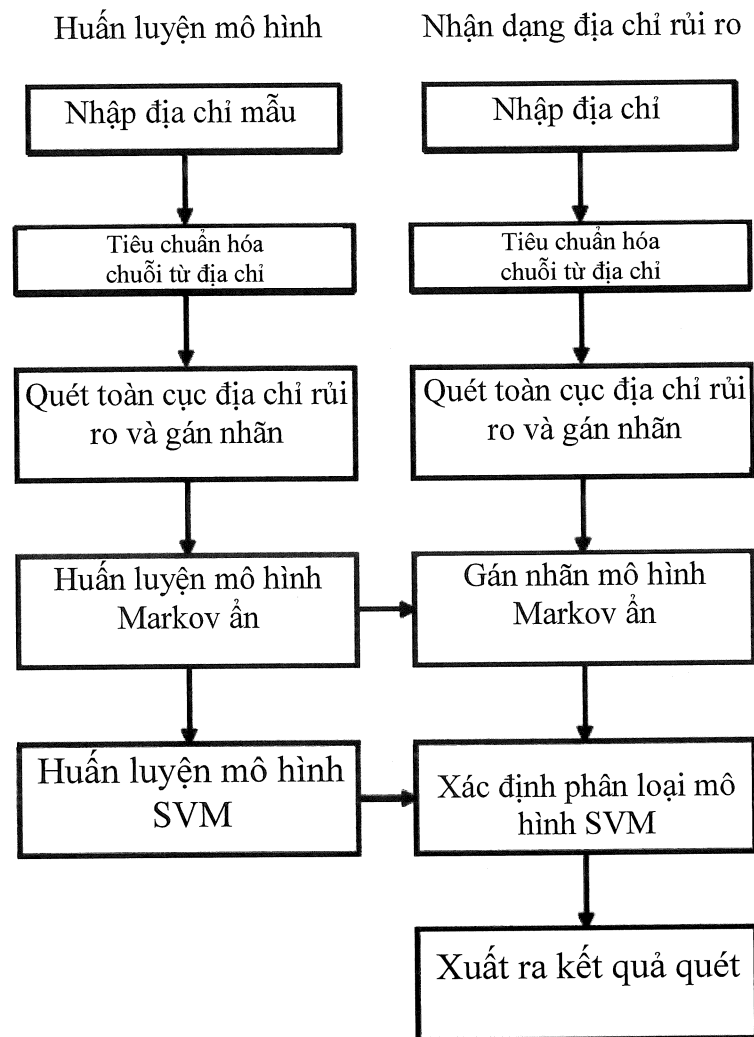


Fig. 6

**Fig. 7**

**Fig. 8**