KỊCH BẢN THUYẾT TRÌNH KHÓA LUẬN TỐT NGHIỆP

*Phần giới thiệu: v:*

# Slide 1

Nhóm em xin chào quý thầy cô trong hội đồng và các bạn!

Lời đầu tiên xin kính chúc quý thầy cô trong hội đồng và các bạn lời chúc tốt đẹp nhất.

Nhóm em gồm có hai thành viên:

* Em là BNV và bạn LHT đến từ lớp KTPM2011.

Dưới sự bùng nổ của cntt nói chung và internet nói riêng, thì máy tính và các dịch vụ internet không còn xa lạ với mọi người, Theo Cục Viễn thông (Bộ Thông tin và Truyền thông), tính đến hết năm 2015, tỷ lệ người dùng Internet tại Việt Nam đã đạt 52% dân số. Điều này có thể dễ dàng nhận ra trong cuộc sống hàng ngày của chúng ta như làm việc văn phòng, giải trí, mua sắm online, mạng xã hội,...

Tuy nhiên đi kèm theo đó là vấn đề về bảo mật và an toàn thông tin. Từ việc phát rối như ẩn file, khóa file,… dến vấn đề như đánh cắp thông tin, tấn công làm hại hệ thống,..

Theo thống kê năm 2015 do BKAV thực hiện thì Trong năm 2015 đã có 62.863 dòng virus máy tính mới xuất hiện tại Việt Nam. 61,7 triệu lượt máy tính đã bị lây nhiễm virus .

Vì vậy, nhóm em chọn nghiên cứu và thực thực hiện là :*”Nghiên cứu một số thuật toán máy học kết hợp với hệ miễn dịch nhân tạo để phát hiện và phân lớp virus máy tính”* dưới sự hướng dẫn của PGS. TS Vũ Thanh Nguyên.

Dựa vào các nghiên cứu trước đó đưa ra cách tiếp cận và giải pháp mới cho vấn đề này

Tiếp theo là những nội dung mà nhóm em sẽ trình bày

# slide 2

Phần trình bày của nhóm em gồm có 4 phần:

* Giới thiệu về đề tài:
* Học máy : SVM ,mạng kohonen
* Miễn dịch nhân tạo.
* Xây dựng demo
* Và Cuối cùng là phần Kết luận

# Slide9

Xin chào hội đồng, sau đây em xin trình bày về phần nghiên cứu hệ miễn dịch nhân tạo,

Hệ miễn dịch nhân tạo liên quan gì đến phát hiện virus máy tính? Đây có lẽ là câu hỏi nhiều người đặt ra nhất. Và trong quá trình trình bày của nhóm em sẽ đi tìm câu trả lời giải thích nó liên quan như thế nào.

Trước tiên em xin trình bày khái quát về hệ miễn dịch con người

Nói về hệ miễn dịch con người:

Mà câu hỏi được đặt ra là:

Miễn dịch là gì? Miễn dịch là trạng thái phòng vệ sinh học đầy đủ để chống nhiễm trùng, bệnh tật, hoặc sự xâm hại sinh học không mong muốn khác và dung hòa các phản ứng để tránh viêm nhiễm, dị ứng, và các bệnh tự miễn

Nếu không có nó, chúng ta chết một cách nhanh chóng, hằng ngày chúng ta tiếp xúc với hàng trăm tác nhân gây hại mới.

Hệ miễn dịch làm việc theo hai cách: Miễn dịch tự nhiên và miễn dịch thu được.

* Miễn dịch tự nhiên: tức là khi sinh ra chúng ta có thể chống chọi lại các nhân gây hại.di truyền
* Miễn dịch thu được: kích thích bởi sự tiếp xúc với vi sinh vật và tạo ra cường độ tăng dần nếu sự tiếp xúc này được lặp đi lặp lại. Ví dụ chúng ta tiêm vaccin phòng bệnh.
* Đề tài nghiên cứu lần này thiên về hướng nghiên cứu miễn dịch thu được.

Vậy thì thành phần chính của hệ miễn dịch là gì: kháng thể và kháng nguyên. Trong hệ miễn dịch nhân tạo thì nó được gọi là sạch(benign) hay virus.

Hệ miễn dịch thu được bên cạnh việc phải chọn ra những tế bào có khả năng nhận biết được những tế bào lạ (antigen hay NonSelf), còn phải loại bỏ những tế bào nhận biết được những tế bào do cơ thể tạo ra (Self - antigen hay đơn giản là Self). Tương ứng với hai quá trình này ta có quá trình chọn lọc tích cực và quá trình chọn lọc tiêu cực.

Có 3 quá trình quan trọng trong hệ miễn dịch:

* Đa dạng thế hệ.
* Chọn lọc tích cực
* Chọn lọc tiêu cực.

Cũng tương ứng với 3 thuật toán mà nhóm em nghiên cứu.

Tiếp theo là đến mạng nơ-ron nhân tạo.

Đây là một mô hình mạng nơ-ron nhân tạo. Trọng mạng nơ-ron nhân tạo thường quan tâm

* Các tập đầu vào.
* Trọng số
* Bộ tổng.
* Ngưỡng.
* Hàm truyền.
* Đầu ra

# Slide 10

Hiện nay, có rất nhiều phần mềm diệt virus, nhiều phần mềm diệt virus tốt, tuy nhiên, khi sử dụng, dưới góc nhìn kỹ thuật, chúng ta đều nhận ra những phần mềm diệt virus này vẫn còn các vấn đề chưa tốt.

Em xin kể ra 2 vấn đề:

Thứ nhất: đối với việc quét các tập tin, các phần mềm diệt virus sử dụng cơ chế xem xét tập tin đang quét có trùng với mẫu trong dữ liệu virus. Nếu tìm thấy thì sẽ nhận dạng tập tin đó là virus. Điều này có nghĩa là nếu như không cập nhật cơ sở dữ liệu hoặc cơ sở dữ liệu không chứa mẫu của 1 virus nào đó, thì sẽ không nhận dạng được con virus này.

Thứ hai:

việc cập nhật dữ liệu theo đặc điểm của từng loại virus nhiều lúc rất lâu, và theo thời gian, khi cơ sở dữ liệu càng lớn thì việc quét virus càng mất nhiều thời gian.

Từ đó, vấn đề đặt ra là làm sao có thể nhận dạng được cả những virus mới, mà trong dữ liệu virus lúc đầu không có mẫu, làm sao để tối giản cơ sở dữ liệu bằng cách không lưu trữ từng mẫu một virus, làm sao để khi quét mỗi tập tin, chúng ta không mất thêm một khoảng thời gian quá lớn bởi vì việc phải so sánh thêm rất nhiều các mẫu mới trong cơ sở dữ liệu.

Từ vấn đề đã đặt ra, ứng dụng mạng nơ ron nhân tạo trong việc nhận dạng virus. Việc sử dụng mạng nơ ron giải quyết một cách tuyệt vời các vấn đề kể trên. Mạng nơ ron có ưu điểm:

- khi sử dụng một cách thích hợp, có thể dùng để nhận dạng các mẫu mới tương đồng với các mẫu cũ nhờ khả năng chống nhiễu,

- thứ 2, dữ liệu của mạng nơ ron là ma trận trọng số, nên khi lưu trữ, không cần phải lưu trữ số lượng lớn mẫu virus để nhận dạng, chỉ cần lưu trữ ma trận trọng số của mạng nơ ron.

- Thứ 3: việc nhận dạng virus bằng mạng nơ ron, là đưa các chuỗi của tập tin vào mạng nơ ron để tính ra kết quả, khi ta cập nhật dữ liệu bằng cách cập nhật trọng số mạng nơron, thì thời gian quét virus của mạng nơ ron không thay đổi, có nghĩa là không mất thêm thời gian để quét virus.

Từ ý tưởng đó, xây dựng nên bài toán nhận dạng virus bằng mạng nơ ron gồm có 4 bước để giải.

Bước 1: Chuẩn bị dữ liệu

Bước 2: Bước 2 xây dựng mạng nơron

Bước 3: Xây dựng SVM để phân lớp các file

Bước 4: Nhận dạng tập tin độc hại dựa vào bước 3, hay bước quét virus.

---

Em xin đi vào trình bày cụ thể từng bước giải bải toán

# SLIDE 11

Bước 1 là xây dựng bộ dữ liệu huấn luyện

Input là thư mục virus và thư mục chứa file sạch

Output là 1 tập hợp các chuỗi độc, và 1 tập hợp các chuỗi sạch

Qua quá trình chọn lọc âm tính, ta loại bỏ dần các chuỗi sạch bên trong các file virus, các chuỗi còn lại sẽ là các chuỗi độc hại, được gán nhãn là virus. Còn tất cả các chuỗi trong các file sạch sẽ được gán nhãn là chuỗi sạch.

# SLIDE 12

Slide tiếp theo là nguyên tắc rút trích chuỗi từ file, và luật để so khớp chuỗi

Về cách rút trích chuỗi, chúng ta sẽ lấy chuỗi với độ dài L= 32 bit (4 byte), và hai chuỗi gần nhau sẽ trùng lên nhau 1 đoạn L/2.

Về việc so khớp chuỗi, nhóm sử dụng 2 luật là Hamming và R-Continuous

# SLIDE 13

Cụ thể quá trình chọn lọc âm tính như sau:

Ta rút trích tất cả các chuỗi trong tập tin virus tạo thành 1 tập hợp đặt tên là Virus String và rút trích tất cả các chuỗi trong tập tin sạch thành 1 tập hợp đặt tên là Benign String, tập hợp Virus String bao gồm các chuỗi độc hại và các chuỗi sạch, nhiệm vụ của chúng ta là loại bỏ các chuỗi sạch này đi, để trong tập này chỉ còn lại các chuỗi độc hại. Chúng ta sẽ lọc các chuỗi sạch bằng cách so sánh từng chuỗi trong tập Virus String với các chuỗi trong Benign String, nếu như phát hiện khớp nhau, thì chuỗi đó là chuỗi sạch và loại bỏ ra khỏi tập Virus String. Kết thúc quá trình, chúng ta sẽ có tập Virus String gồm các chuỗi độc và Benign String gồm các chuỗi sạch.

# SLIDE 14

Em xin đi vào giải bước 2 của bài toán

Là xây dựng bộ phân cụm Kohonen

Đầu tiên em sẽ thực hiện viện trộn các chuỗi sạch và chuỗi virus vào 1 tập với thứ tự ngẫu nhiên gọi là tập huấn luyện.

Tiếp theo em sẽ dùng tập huấn luyện để huấn luyện mạng nơ ron Kohonen, hay dễ hiểu hơn là dùng mạng Kohonen để phân nhóm tập huấn luyện.  
Sau khi phân nhóm kết thúc, em sẽ thực hiện việc mapping các nơ ron của mạng Kohonen, bằng cách cho các chuỗi virus và chuỗi sạch qua mạng Kohonen, với mỗi chuỗi sẽ có một nơ ron thắng, nếu chuỗi đó là virus thì sẽ tăng biến đếm virus của nơ ron thêm 1, nếu chuỗi đó là chuỗi sạch thì sẽ tăng biến đếm chuỗi sạch của nơ ron thêm 1.

Kết thúc quá trình, chúng ta sẽ có 1 mạng Kohonen, với các nơ ron được mapping mức độ nguy hiểm dùng để nhận dạng chuỗi virus.

Cụ thể nếu nơ ron nào chỉ nhận dạng chuỗi virus thì giá trị biến đếm virus của nơ ron đó chính là mức độ nguy hiểm ứng với nơ ron. Sau này, khi có chuỗi nào mà nơ ron này nhận dạng, thì chuỗi đó có mức độ nguy hiểm bằng với mưc độ nguy hiểm mà nơ ron được gán nhãn.

Ngược lại, bao gồm các nơ ron không nhận dạng được virus và k nhận dạng được chuỗi sạch, hoặc chỉ nhận dạng chuỗi sạch, hoặc nhận dạng cả chuỗi sạch và chuỗi virus sẽ được gán mức độ nguy hiểm = 0.

# SLIDE 17

Bước tiếp theo của bài toán, bước 3, là xây dựng bộ phân lớp tập tin

Trước khi nói về cách giải của nhóm em ở bước này, em xin trình bày 1 cách giải đơn giản.

Đó là ta với mỗi tập tin, ta sẽ rút trích tất cả ra một tập các chuỗi, sau đó ta sẽ đưa từng chuỗi qua mạng Kohonen để tính mức độ nguy hiểm, từ đó tổng hợp mức độ nguy hiểm của tập tin. Tiếp theo ta chọn 1 mức ngưỡng T nào đó, nếu như mức độ nguy hiểm của tập tin >= T, thì tập tin đó là virus, ngước lại sẽ là tập tin sạch.

Tuy nhiên, việc xác định giá trị ngưỡng này không phải dễ dàng, ta cần thực nghiệm trên một số lượng tập tin lớn, bao gồm cả virus và chương trình sạch, từ đó lựa chọn giá trị ngưỡng phù hợp.

Nhóm thực hiện sử dụng một giải pháp hiệu quả hơn nhiều bằng việc xây dựng một bộ phân lớp, học từ dữ liệu thực nghiệm từ đó xác định nhãn (virus hoặc sạch) cho các tập tin mới.

Để xây dựng bộ phân lớp, đầu tiên chúng ta cần phải tạo một bộ dữ liệu huấn luyện cho bộ phân lớp này, gọi là bộ dữ liệu huấn luyện phân lớp.

Cụ thể ta sẽ chọn 1 tập tin nào đó trong tập dữ liệu ban đầu mà ta đã biết chính xác là sạch hay virus. Sau đó ta sẽ tính toán mức độ nguy hiểm và gắn nhãn cho nó. Ta được một bộ trong bộ dữ liệu huấn luyện có cấu trúc gồm 2 thành phần, gồm mức độ nguy hiểm và nhãn.

Việc tiếp theo là ta xây dựng 1 tập các bộ dữ dữ liệu như trên để được bộ dữ liệu huấn luyện phân lớp.

Bởi vì cấu trúc mỗi bộ dữ liệu huấn luyện khá đơn giản nên ta có thể sử dụng một mạng nơ-ron nhân tạo với giải thuật học lan truyền ngược để học các dữ liệu huấn luyện này.

Sau khi huấn luyện xong mạng nơ ron nhân tạo này, có nghĩa là ta đã xong bước xây dựng bộ phân lớp tập tin.

Và cuối cùng là ta dùng bộ phân cụm Kohonen và bộ phân lớp tập tin để nhận dạng virus, là bước 4 của bài toán

Ở bước này, qua trình diễn ra đơn giản như sau, với mỗi tập tin sẽ dùng bộ phân cụm Kohonen tính mức độ nguy hiểm, và dùng bộ phân lớp để xác định mức độ nguy hiểm đó sẽ tương ứng là file virus hay file sạch.

Phần kết luận: Xuân

# SLIDE 23

Qua quá trình thực hiện đề tài, Nhóm đã đạt được các nội dung:

Về lý thuyết

Nắm được khái niệm virus, tầm quan trọng của bảo mật đối với doanh nghiệp

Hiểu được hệ miễn dịch nhân tạo, mạng nơ ron nhân tạo và một số thuật toán máy học khác.

Về thực nghiệm:

- Xây dựng ứng dụng để nghiên cứu

- Cài đặt các thuật toán máy học và mạng nơ ron trong quá trình nghiên cứu

# SLIDE 24

Tuy đạt được nhiều kết quả nhưng đề tài còn có mặt hạn chế, như  
Hướng nghiên cứu còn gần với những nghiên cứu trước, chưa tiến xa.

Chưa có nhiều kinh nghiệm về nghiên cứu.

Chương trình thử nghiệm chưa tối ưu nhất.

# SLIDE 25

Về hướng phát triển của đề tài:

…

# SLIDE 26 +27

Đây là danh sách các tài liệu tham khảo trong quá trình nhóm thực hiện đề tài

Tiếp theo, em xin trình bày phần demo của đề tài.

New

Phần Ý Tưởng Giải Quyết Các Bài Toán: Xuân

Đang ở Slide 8

Hiện nay, có rất nhiều phần mềm diệt virus, nhiều phần mềm diệt virus khá tốt, tuy nhiên chúng vẫn còn một số nhược điểm. Đó là

- Không nhận dạng được mẫu virus mới

- Cập nhật dữ liệu đôi khi rất lâu

Từ đó, vấn đề đặt ra:

- làm sao có thể nhận dạng được virus mới.

- giảm thời gian cập nhật dữ liệu

Từ vấn đề đã đặt ra, nhóm em đã nghĩ ra ý tưởng, ứng dụng mạng nơ ron nhân tạo trong việc nhận dạng virus. Việc sử dụng mạng nơ ron nhân tạo giải quyết một cách tuyệt vời các vấn đề kể trên. Mạng nơ ron nhân tạo có ưu điểm:

- Thứ 1: khi sử dụng một cách thích hợp, có thể dùng để nhận dạng các mẫu mới tương đồng với các mẫu cũ nhờ khả năng chống nhiễu.

- Thứ 2, dữ liệu của mạng nơ ron nhân tạo là ma trận trọng số. nên khi cập nhật dữ liệu, chỉ cập nhật ma trận này. Không cần phải cập nhật quá nhiều mẫu virus.

Từ ý tưởng đó, nhóm em xây dựng nên bài toán nhận dạng virus bằng mạng nơ ron nhận tạo mà bên trong đó lại bao gồm 4 bài toán nhỏ hơn.

Bài toán 1: Xây dựng bộ dữ liệu huấn luyện.

Bài toán 2: Tính mức độ nguy hiễm của chuỗi

Bài toán 3: Nhận dạng tập tin virus

Bài toán 4: Quét virus.

Bốn bài toán tuy riêng rẽ nhưng liên quan chặt chẽ với nhau, bài toán sau chỉ giải được khi ta giải được bài toán trước.

Bải toán 1 là xây dựng dữ liệu huần luyện để ta có thể tính mức độ nguy hiểm của chuỗi ở bài toán 2.

Từ việc tính mức độ nguy hiểm của chuỗi ở bài toán 2, chính là tiền đề để tính mức độ nguy hiểm của tập tin rồi từ đó nhận dạng tập tin virus ở bài toán 3.

Và cuối cùng, dựa vào kết quả bài toán 2 và 3, ta tổng hợp lại để ra được kết quả là thuật toán quét virus ở bài toán 4.

---

Em xin đi vào trình bày cụ thể từng bài toán

Qua slide 9

Đầu tiên là bài toán xây dựng dữ liệu huấn luyện.  
Nội dung của bài toán 1 là từ các tập tin virus và tập tin sạch đã biết trước, ta rút trích ra các chuỗi nhị phân và phân chia ra thành 2 tập hợp riêng biệt. tập hợp đầu tiền là các chuỗi nhị phân độc hại thường xuất hiện trong tập tin virus gọi là tập chuỗi virus. Tập hợp thứ hai là các chuỗi nhị phân xuất hiện trong các tập tin sạch gọi là tập chuỗi sạch.

Cách giải:

Qua quá trình chọn lọc âm tính, ta loại bỏ dần các chuỗi sạch bên trong các tập tin virus, các chuỗi còn lại sẽ là các chuỗi độc hại, được gán nhãn là chuỗi virus. Còn tất cả các chuỗi trong các file sạch sẽ được gán nhãn là chuỗi sạch.

Qua slide 10

Đây là sơ đồ quá trình chọn lọc âm tính

Quá trình chọn lọc âm tính đơn giản là ta rút trích các chuỗi trong file virus tạo thành tập chuỗi virus, rút trích các chuỗi trong file sạch tạo thành tập chuỗi sạch, rồi sau đó, lấy từng chuỗi trong tập chuỗi virus so khớp với các chuỗi trong tập chuỗi sạch, nếu khớp nhau thì loại bỏ chuỗi ra khỏi tập chuỗi virus.

tiếp theo là nguyên tắc rút trích chuỗi từ file, và luật để so khớp chuỗi

Qua slide 11

Về cách rút trích chuỗi, chúng ta sẽ lấy chuỗi với độ dài L= 32 bit (4 byte), và hai chuỗi gần nhau sẽ trùng lên nhau 1 đoạn L/2.

Về việc so khớp chuỗi, nhóm sử dụng 2 luật là Hamming và R-Continuous

Tiếp theo là bài toán thứ 2, nhận dạng chuỗi độc

Qua slide 12:

Nội dung bài toán là từ 1 chuỗi nhị phân được rút trích từ 1 tập tin bất kì, ta tính toán mức độ nguy hiểm của chuỗi đó.

Cách giải:

Để giải bài toán này, em dùng mạng nơ ron Kohonen, em xin gọi tắt là mạng Kohonen

Việc giải bài toán này, đơn giản là việc cài đặt mạng Kohonen, để mạng Kohonen có khả năng nhận dạng và tính toán mức độ nguy hiểm của chuỗi

Đây là sơ đồ quá trình cài đặt mạng Kohonen.

Đầu tiên ta sẽ thực hiện trộn các chuỗi sạch và chuỗi virus có được ở bài toán 1 theo thứ tự ngẫu nhiên để tạo thành tập huấn luyện, việc trộn này nhằm mục đích giúp cho trong quá trình huấn luyện mạng, các nơ ron phân cực một cách đồng đều. và kết quả huấn luyện của mạng Kohonen chính xác hơn.

Tiếp theo, ta luyện mạng nơ ron Kohonen,

Qua slide 13  
đây là quá trình huấn luyện mạng Kohonen, các bước huấn luyện được trình bày trên sơ đồ

Sau khi luyện xong, ta thực hiện mapping các nơ ron với 1 mức độ nguy hiểm, mức độ nguy hiểm này = số chuỗi virus mà nơ ron nhận dạng được. tuy nhiên, nơ ron nào nhận dạng cả chuỗi sạch, ta sẽ gán mức độ nguy hiểm của của nơ ron này = 0, việc này giúp giảm tỉ lệ nhận diện nhầm tập tin sạch thành tập tin virus.

Sau này, khi một nơ ron nhận dạng được một chuỗi nào đó, thì độ nguy hiểm của chuỗi đó được xác định = mức độ nguy hiểm mà nơ ron được gán cho.

Vậy là ta đã hoàn thành được bài toán 2, cũng là bài toán quan trọng nhất, đó là tính mức độ nguy hiểm của chuỗi.

Tiếp theo, ta đi vào chi tiết bài toán 3, là bài toán nhận dạng tập tin virus.

Qua slide 14

Nội dung của bài toán là với một tập tin bất kỳ, ta xác định tập tin này là virus hay tập tin sạch.

Ta đi vào cách giải bài toán 3:

Trước khi nói về cách giải của nhóm em ở bước này, em xin trình bày 1 cách giải đơn giản.

Đó là với mỗi tập tin, ta dùng mạng Kohonen để tính mức độ nguy hiểm của các chuỗi trong tập tin đó rồi tổng hợp thành mức độ nguy hiểm của tập tin. Tiếp theo ta chọn 1 mức ngưỡng T nào đó, nếu như mức độ nguy hiểm của tập tin >= T, thì tập tin đó là virus, ngước lại sẽ là tập tin sạch.

Tuy nhiên, việc xác định giá trị ngưỡng này không phải dễ dàng, ta cần thực nghiệm trên một số lượng tập tin lớn, bao gồm cả virus và chương trình sạch, từ đó lựa chọn giá trị ngưỡng phù hợp.

Nhóm thực hiện sử dụng một giải pháp hiệu quả hơn nhiều bằng việc xây dựng một bộ phân lớp tập tin, học từ dữ liệu thực nghiệm từ đó xác định nhãn (virus hoặc sạch) cho các tập tin mới.

Để xây dựng bộ phân lớp, đầu tiên chúng ta cần phải tạo một bộ dữ liệu huấn luyện cho bộ phân lớp này, gọi là bộ dữ liệu huấn luyện phân lớp.

Cụ thể, ta sẽ chọn 1 tập tin nào đó trong tập dữ liệu ban đầu mà ta đã biết chính xác là virus hay sạch. Sau đó ta sẽ tính toán mức độ nguy hiểm và gắn nhãn (là virus hay sạch) cho nó. Ta được một bộ có cấu trúc gồm 2 thành phần: mức độ nguy hiểm và nhãn.

Việc tiếp theo là ta xây dựng 1 tập các bộ dữ dữ liệu như trên để được bộ dữ liệu huấn luyện phân lớp.

Bởi vì cấu trúc mỗi bộ dữ liệu huấn luyện khá đơn giản nên ta có thể sử dụng một mạng nơ-ron nhân tạo với giải thuật học lan truyền ngược (gọi tắt là mạng lan truyền ngược) để học các dữ liệu huấn luyện này.

Qua slide 15

Đây là quá trình huấn luyện mạng lan truyền ngược, quá trình huấn luyện được trình bày trên sơ đồ

Sau khi huấn luyện xong mạng nơ ron nhân tạo này, ta đã có thể nhận dạng tập tin là virus hay sạch dựa vào mức độ nguy hiểm của nó. Vậy là ta đã giải xong bài toán nhận dạng tập tin virus.

Và cuối cùng là ta dùng bộ phân cụm Kohonen và bộ phân lớp tập tin để quét virus, là bước 4 của bài toán

Qua slide 16

Nội dung bài toán:  
Cho 1 thư mục chứa 1 hoặc nhiều tập tin, nhận dạng các file virus trong thư mục đó.

Cách giải:

Ở bước này, qua trình diễn ra đơn giản như sau, với mỗi tập tin sẽ dùng bộ phân cụm Kohonen tính mức độ nguy hiểm , và dùng bộ phân lớp để xác định mức độ nguy hiểm đó sẽ tương ứng là file virus hay file sạch.

Phần kết luận: Xuân

Slide: 21

Đây là sơ đồ đánh giá sự kết hợp các thuật toán trong đề tài nghiên cứu của nhóm em.  
Đề tài tập trung nghiên cứu và mạng nơ ron Kohonen, bên cạnh đó kết hợp thuật toán chọn lọc âm tính và mạng nơ ron lan truyền ngược.

Slide ???: Kết quả nghiên cứu:

Slide 22:

Qua quá trình thực hiện đề tài, Nhóm đã đạt được các nội dung:

Về lý thuyết

Nắm được khái niệm virus, tầm quan trọng của bảo mật đối với doanh nghiệp

Hiểu được hệ miễn dịch nhân tạo, mạng nơ ron nhân tạo và một số thuật toán máy học khác.

Về thực nghiệm:

- Xây dựng ứng dụng để nghiên cứu

- Cài đặt các thuật toán máy học và mạng nơ ron trong quá trình nghiên cứu

Slide 23:

Tuy đạt được nhiều kết quả nhưng đề tài còn có mặt hạn chế, như  
Hướng nghiên cứu còn gần với những nghiên cứu trước, chưa tiến xa.

Chưa có nhiều kinh nghiệm về nghiên cứu.

Chương trình thử nghiệm chưa tối ưu nhất.

Slide 24:

Về hướng phát triển của đề tài:

…

Slide 25:

Đây là danh sách các tài liệu tham khảo trong quá trình nhóm thực hiện đề tài

Tiếp theo, em xin trình bày phần demo của đề tài.