**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI** (Bold, size 14)

**PHÂN HIỆU TẠI TP. HỒ CHÍ MINH** (Bold, size 14)

**BỘ MÔN CÔNG NGHỆ THÔNG TIN** (Bold, size 14)



**BÁO CÁO ĐỒ ÁN TỐT NGHIỆP**

**(Bold, size 15)**

**ĐỀ TÀI: …..…………………………………………......................**

**………………………………………………………………………**

**...........................................................................................................**

**(Bold, size 16)**

Giảng viên hướng dẫn: LÊ NHẬT TÙNG (size 14, chữ hoa)

Sinh viên thực hiện: TRẦN THỊ HOA (size 14, chữ in hoa)

Lớp : ……… (size 14, chữ in hoa)

Khoá :.…........ (size 14, chữ in hoa)

Tp. Hồ Chí Minh, năm 2017 (size 14)

**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI** (Bold, size 14)

**PHÂN HIỆU TẠI TP. HỒ CHÍ MINH** (Bold, size 14)

**BỘ MÔN CÔNG NGHỆ THÔNG TIN** (Bold, size 14)



**BÁO CÁO ĐỒ ÁN TỐT NGHIỆP**

**(Bold, size 15)**

**ĐỀ TÀI: …..…………………………………………......................**

**………………………………………………………………………**

**...........................................................................................................**

**(Bold, size 16)**

Giảng viên hướng dẫn: LÊ NHẬT TÙNG (size 14, chữ hoa)

Sinh viên thực hiện: TRẦN THỊ HOA (size 14, chữ in hoa)

Lớp : ……… (size 14, chữ in hoa)

Khoá :.…........ (size 14, chữ in hoa)

Tp. Hồ Chí Minh, năm 2017 ( size 14)

TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI **CỘNG HÒA XÃ HỘI CHỦ NGHIÃ VIỆT NAM**

**PHÂN HIỆU TẠI THÀNH PHỐ HỒ CHÍ MINH** Độc lập – Tự do – Hạnh phúc

# NHIỆM VỤ THIẾT KẾ TỐT NGHIỆP

BỘ MÔN: **CÔNG NGHỆ THÔNG TIN**

-------\*\*\*-------

**Mã sinh viên:** **Họ tên SV:**

**Khóa:** **Lớp:**

1. **Tên đề tài**
2. **Mục đích, yêu cầu**
3. **Nội dung và phạm vi đề tài**
4. **Công nghệ, công cụ và ngôn ngữ lập trình**
5. **Các kết quả chính dự kiến sẽ đạt được và ứng dụng**
6. **Giáo viên và cán bộ hướng dẫn**

Họ tên:

Đơn vị công tác:

Điện thoại: Email:

|  |  |
| --- | --- |
| **Ngày tháng 03 năm 2017**  **Trưởng BM Công nghệ Thông tin** | **Đã giao nhiệm vụ TKTN**  **Giáo viên hướng dẫn** |
| **ThS. Nguyễn Thị Hải Bình** |  |

Đã nhận nhiệm vụ TKTN

Sinh viên: Ký tên:

Điện thoại: Email:

**LỜI CẢM ƠN (size 15, bold)**

*(Cách 1 tab, Time newRoman, 20)*

Size 13

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN (size 15, bold)**

*(Cách 1 tab, Time newRoman, 20)*

Size 13

|  |
| --- |
| ***Tp. Hồ Chí Minh, ngày ….… tháng ….… năm ….…***  **Giáo viên hướng dẫn**  **Lê Nhật Tùng** |

**MỤC LỤC (size 15, bold)**

*(Cách 1 tab, Time newRoman, 20)*

BẮT BUỘC DÙNG MỤC LỤC TỰ ĐỘNG

CHƯƠNG 1: TỔNG QUAN ………………………………………..…….………… 1

1.1. Tổng quan về nợ công…………………………………………………..……....... 1

1.1.1. Nợ công Việt Nam …………………………………………………………..… 2

**DANH MỤC CHỮ VIẾT TẮT**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Mô tả** | **Ý nghĩa** | **Ghi chú** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**BẢNG BIỂU, SƠ ĐỒ, HÌNH VẼ (size 15, bold)**

*(Cách 1 tab, Time newRoman, 20, mỗi nội dung trình bày bắt đầu từ 1 trang mới)*

Bảng 1.1: (size 13)…………………………………………………………………........

Bảng 1.2: ..........................................................................................................................

Sơ đồ 1.1:………………………………………………………………………………..

Hình 1.1: ………………………………………………………………………………..

**Ghi chú:**

* + Xếp sau trang Mục lục
  + Chữ số thứ nhất chỉ tên chương
  + Chữ số thứ hai chỉ thứ tự bảng biểu, sơ đồ, hình,…trong mỗi chương
  + Ở cuối mỗi bảng biểu, sơ đồ, hình,…trong mỗi chương phải có ghi chú, giải thích, nêu rõ nguồn trích hoặc sao chụp,…

**CHƯƠNG 1. MỞ ĐẦU**

### Đặt vấn đề

Thị giác máy tính là một lĩnh vực đã và đang được ứng dụng rộng khắp trong hiều mặt của cuộc sống, là kĩ thuật liên quan đến việc nhận dạng và phân biệt đối tượng trong thực tế thông qua ảnh chụp của chúng. Kĩ thuật này ngày càng được áp dụng rộng rãi vào nhiều lĩnh vực của nông nghiệp như tự động hóa, nông nghiệp chính xác, phân loại nông sản, nhận dạng (loài cây, cỏ dại, trái cây...). Thực chất, đa số bài toán của thị giác máy tính là bài toán phân lớp. Chẳng hạn, việc xác định giới tính của một người thông qua ảnh chân dung của người đó, thực chất là cách xác phân lớp nam hoặc nữ, hay nhận dạng biểu cảm của con người: tức giận, vui vẻ , bất ngờ, lo lắng,…

Chính vì vậy, việc ứng dụng các thuật toán Machine learning vào thị giác máy tính ngày càng được chú trọng. Đặt biệt là các bài toán phân lớp. Trong Machine learning có nhiều bài toán phân lớp chẳng hạn như: kNN, Logistic Regression, SVM,… Nhìn chung các ,thuật toán trên ban đầu giải quyết bài toán phân lớp đơn giản (tức là gồm hai lớp - gọi là binary classification ), ngoài trừ thuật toán kNN. Nhưng sau đó, chúng được cải tiến để giải quyết các bài toán đa lớp (Multi-class).

Trong số các thuật toán kể ở trên, SVM (Support Vector Machine) là một thuật toán dễ sử dụng, đơn giản nhưng lại đem lại một kết quả với hiệu suất cao. Thậm chí khi đem so sánh với các thuật toán khác ở trong việc phân lớp (mặc dù SVM vẫn có thể áp dụng cho các bài toán hồi quy, nhưng em cảm thấy nó vẫn thích hợp cho bài toán phân lớp hơn), thì kết quả của SVM có thể gọi là nhỉnh hơn đôi phần.

Trong nội dung báo cáo thực hành chuyên môn lần này em sẽ tiến hành nghiêng cứu về thuật toán SVM, đồng thời sẽ áp dụng nó cho bài toán nhận diện chữ viết tay bằng một demo nhỏ.

##### Bài toán nhận

### Mục tiêu nghiên cứu

Mục tiêu đề tài là tiến hành nghiên cứu về thuật toán SVM. Hiểu rõ được ý tưởng triển khai của SVM.

* 1. Phạm vi

- Nghiêng cứu về ý tưởng của SVM

- Các cơ sở về toán học

- Phương pháp biên mềm (Soft-margin) nhằm giúp SVM có thể chấp nhận nhiểu

- Phương pháp Kernel trick giải quyết các bài toán phi tuyến tính

### Cấu trúc báo cáo thực tập chuyên môn

* + 1. Chương 1: Mở đầu
    2. Chương 2: Giới thiệu chung về Machine Learning
    3. Chương 3: Thuật toán SVM

**CHƯƠNG 2. GIỚI THIỆU CHUNG VỀ MACHINE LEARNING**

##### 2.1 Tổng quan về trí tuệ nhân tạo

Ngày nay, lĩnh vực AI - Artificial Intelligence (Trí tuệ nhân tao) đang phát triển một cách vũ bão và ngày càng len lỏi vào mọi lĩnh vực trong đời sống. Mục đích chính của lĩnh vực AI là tập trung nghiêng cứu và mô phỏng lại quá trình học tập, suy nghĩ va tư duy của con người để máy có thể học. Các khái niệm này lần đầu tiên được John McCarthy, một nhà khoa học máy tính Mỹ, vào năm 1956 tại Hội nghị The Dartmouth.

Trong lịch sử phát triển AI, các nhà nghiên cứu phân thành 4 hướng tiếp cận chính:

- Hành động như người (acting humanly)

- Suy nghĩ như người (thinking humanly)

- Suy nghĩ hợp lý (thinking rationally)

- Hành động hợp lý (acting rationally)

Trong đó, mức độ mô phỏng máy tính như người là khó nhất và đây cũng là mục tiêu mà các nhà khoa học đang hướng tới. Đến năm 1950, mục tiêu trên được “khai phá” nhờ bài kiểm tra “Turing Test”. Bài test này được Alan Turing đề xuất nhằm mục đích để kiểm tra hệ thống máy tính đã đạt đến khả năng thông minh hay chưa. Bài test tiến hành bao gồm một người đặt câu hỏi, một người trả lời câu hỏi và một máy tính phân tích trả lời câu hỏi. Nếu quá trình trao đổi đó, người đặt câu hỏi không còn phân biệt được đâu là câu trả lời của con người đâu là câu trả lời của máy tính, thì máy tính được xem là thông minh. Từ cách tiếp cận này đã cho ra các nhánh sau đây:

- Natural language processing: xử lý ngôn ngữ tự nhiên

- Computer vision : thị giác máy tính

- Knowledge representation: biểu diễn tri thức

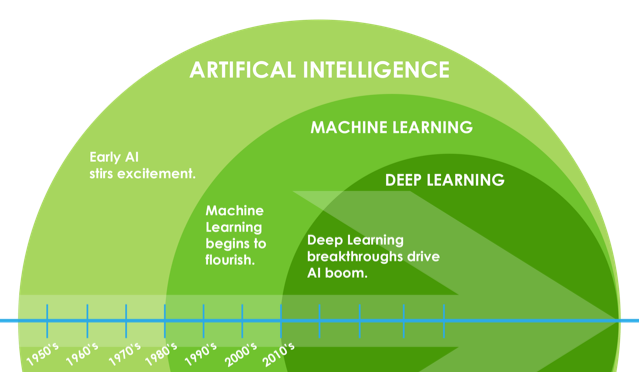
- Machine learnig: máy học

- Automated reasoning: Tự suy luận

- Robotics: Rô bốt học

- ….

##### 2.2 Machine learning (Máy học)



Machine learning là một tập con của trí tuệ nhân tạo. Theo Wikipedia Machine learning được định nghĩa như sau.

“Machine learnig là một lĩnh vực của trí tuệ nhân tạo liên quan đến việc nghiên cứu và xây dựng các kĩ thuật cho phép các hệ thống "học" tự động từ dữ liệu để giải quyết những vấn đề cụ thể. Ví dụ như các máy có thể "học" cách phân loại thư điện tử xem có phải thư rác (spam) hay không và tự động xếp thư vào thư mục tương ứng. Học máy rất gần với suy diễn thống kê (statistical inference) tuy có khác nhau về thuật ngữ.

Học máy có liên quan lớn đến thống kê, vì cả hai lĩnh vực đều nghiên cứu việc phân tích dữ liệu, nhưng khác với thống kê, học máy tập trung vào sự phức tạp của các giải thuật trong việc thực thi tính toán. Nhiều bài toán suy luận được xếp vào loại bài toán NP-khó, vì thế một phần của học máy là nghiên cứu sự phát triển các giải thuật suy luận xấp xỉ mà có thể xử lý được.

Học máy có hiện nay được áp dụng rộng rãi bao gồm máy truy tìm dữ liệu, chẩn đoán y khoa, phát hiện thẻ tín dụng giả, phân tích thị trường chứng khoán, phân loại các chuỗi DNA, nhận dạng tiếng nói và chữ viết, dịch tự động, chơi trò chơi và cử động rô-bốt (robot locomotion).”

Như vậy thông qua định nghĩa trên ta có thể thấy được ML là một lĩnh vực tập trung nghiêng cứu về dữ liệu (data), và từ tập dữ liệu đó (do người lập trình đưa vào học do máy tính tự động thu thập) để từ đó máy tính có thể xử lý các bài toán liên quan.

Phân loại Machine learning. Dựa vào phương thức học ta có thể chia machine learining ra thành 4 loại sau:

- Supervised Learning (Học có giám sát): đây là thuật toán dự đoán đầu ra (outcome) của một dữ liệu mới (new input) dựa trên các cặp (input, outcome) đã biết từ trước. Cặp dữ liệu này còn được gọi là (data, label), tức (dữ liệu, nhãn). Supervised learning là nhóm phổ biến nhất trong các thuật toán Machine Learning. Học có giám sát tiếp tục được chia thành các nhánh khác bao gồm:

+ Các thuât toán Classification (Phân loại): giải quyết các bài toán dạng: “Cái này có dạng là A hay B? ”. Cụ thể là cái bài toán như: xác định giới tính một người, nhận dạng chữ viết tay, xác định email là spam hay không?…. Các thuật toán nổi tiếng trrong nhóm này bao gồm: Logistic Regression, SVM, KNN,….

+ Các thuật toán Regression (Hồi quy): giải quyết các bài toán dạng: “Với dữ liệu như thế này sẽ có kết quả xấp xỉ là bao nhiêu”. Cụ thể là cái bài toán như: “Dự đoán giá nhà với một tập dữ liệu diện tích, khoảng cách đến trung tâm thành phố,….”. Các thuật toán điển hình là: Linear Regression,Stepwise Regression,….

- Học không giám sát: tương tự với học có giám sát, phương pháp học không giám sát cũng tiến hành sát định đầu ra của một dữ liệu mới thông qua tập dữ liệu có sẵn. Tuy nhiên điểm khác biệt là học không giám sát, tập dữ liệu có sẵn (data training) chưa sát định được đầu ra (kết quả). Các bài toán Unsupervised learning được tiếp tục chia nhỏ thành hai loại:

+ Clustering (phân nhóm): Một bài toán phân nhóm toàn bộ dữ liệu thành các nhóm nhỏ dựa trên sự liên quan giữa các dữ liệu trong mỗi nhóm. Ví dụ bài toán xây dựng hệ thống sản phẩm liên quan.

+ Association: Là bài toán khi chúng ta muốn khám phá ra một quy luật dựa trên nhiều dữ liệu cho trước. Ví dụ: những khách hàng nam mua quần áo thường có xu hướng mua thêm đồng hồ hoặc thắt lưng; những khán giả xem phim Spider Man thường có xu hướng xem thêm phim Bat Man, dựa vào đó tạo ra một hệ thống gợi ý khách hàng (Recommendation System), thúc đẩy nhu cầu mua sắm.

- Semi-Supervised Learning (Học bán giám sát): Các bài toán khi chúng ta có một lượng lớn dữ liệu nhưng chỉ một phần trong chúng được gán nhãn (có kết quả sẵn) được gọi là Semi-Supervised Learning. Những bài toán thuộc nhóm này nằm giữa hai nhóm được nêu bên trên.

- Reinforcement Learning (Học tăng cường): là các bài toán giúp cho một hệ thống tự động xác định hành vi dựa trên hoàn cảnh để đạt được lợi ích cao nhất (maximizing the performance). Hiện tại, Reinforcement learning chủ yếu được áp dụng vào Lý Thuyết Trò Chơi (Game Theory), các thuật toán cần xác định nước đi tiếp theo để đạt được điểm số cao nhất.

**CHƯƠNG 3. KẾT QUẢ VÀ KIẾN NGHỊ**

3.1 Kết quả đạt được

3.2 Kiến nghị

**PHỤ LỤC**

Phụ lục 1: hướng dẫn cài đặt

Phụ lục 2: hướng dẫn sử dụng

**TÀI LIỆU THAM KHẢO**

1. T.S Lê Phan Thị Diệu Thảo, Nguyễn Thảo Phương, *Nợ công và những vấn đề cần bàn thêm*, Nhà xuất bản Kinh Tế.
2. Boulding, K.E. (1995), *Economics analysis*, Hamish Hamilton, London.
3. <Http://dantri.com.vn/c76/s76-402058/nhung-moc-chinh-cua-khung-hoang-no-chau-au.htm>, “Những mốc chính của khủng hoảng nợ châu Âu”