

ĐÁNH GIÁ HIỆU SUẤT CÁC THUẬT TOÁN TRÍ TUỆ NHÂN TẠO TRONG PHÂN LOẠI RÁC THẢI

Lê Minh Hóa

Khoa Công nghệ thông tin 2,
Học viện Công nghệ Bưu chính Viễn thông

Tóm tắt: Phân loại rác thải là một bài toán lớn trong thị giác máy tính và hiện nay có nhiều hướng tiếp cận đưa ra giải pháp, trong đó hướng sử dụng trí tuệ nhân tạo đạt mức độ hiệu quả chính xác đáng kể. Trong bài báo này, các thuật toán phân loại trong học máy như cây quyết định, thuật toán rừng ngẫu nhiên, SVM, PCA và mô hình học sâu tiêu biểu VGG16 được nghiên cứu đánh giá so sánh hiệu quả trong việc phân loại. Nghiên cứu này đề xuất thêm một lớp phân loại Softmax sau VGG16 sẽ cho ra kết quả có độ chính xác cao hơn. Các mô hình học sâu được nghiên cứu ở đây sử dụng cốt lõi Mạng nơ-ron tích chập (CNN) là một trong những mạng tiên tiến nhất trong thị giác máy tính, chứng tỏ được khả năng phân loại hình ảnh một cách hiệu quả. Kết quả từ mô hình đề xuất đã được cải thiện với độ chính xác 71.1% so với sử dụng mô hình CNN truyền thống trong điều kiện bộ dữ liệu có số lượng mẫu nhỏ. Trong tương lai gần, các mô hình học sâu sẽ hỗ trợ máy móc việc phân loại rác tự động và không cần nhất thiết can thiệp của con người.

Từ khóa: CNN, Decision tree, Random forest, PCA, phân loại rác, SVM, VGG16.

I. GIỚI THIỆU

Trong thế giới hiện đại ngày nay, việc xử lý khối lượng khổng lồ rác thải sinh hoạt hằng ngày đã trở thành một vấn đề cấp bách cho các quốc gia. Làm sao để giảm nhẹ tác động của rác thải gây ô nhiễm nghiêm trọng tới môi trường là câu hỏi thường trực. Phân loại rác thải là một trong những giải pháp cần kíp trước mắt. Việc tách những nguyên vật liệu có thể tái chế không những đem lại hiệu quả kinh tế khi chỉ cần tinh lọc lại nguyên liệu đầu vào thay vì phải khai thác từ nguồn tài nguyên tự nhiên, mà còn giảm thiểu phần nào sự lãng phí năng lượng trong việc khai thác. Việc phân loại này vẫn đòi hỏi con người tham gia. Do đó, gần đây sử dụng trí tuệ nhân tạo nhằm đẩy mạnh hiệu quả nhận dạng và phân loại tự động rác thải đô thị trở thành một trong nhiều chủ đề nghiên cứu sôi động về bảo vệ môi trường sống, ứng dụng thiết thực cho ngành công nghiệp xử lý rác.

II. MỘT SỐ NGHIÊN CỨU LIÊN QUAN

A. Các thuật toán học máy

Học máy (machine learning) là một phần của trí tuệ nhân tạo trong đó các thuật toán máy tính được sử dụng để

tự học từ dữ liệu. Đến đầu những năm 90, Breiman, Quinlan và các cộng sự đã đề xuất các thuật toán cây quyết định như CART [1], ID3, C4.5 [2]... Cây quyết định phân loại dữ liệu thông qua một chuỗi các luật, quyết định dự đoán đưa ra giá trị gì dựa trên những tình trạng nào. Ở đó, mỗi node của cây sẽ là các thuộc tính, và các nhánh là giá trị lựa chọn của thuộc tính đó. Bằng cách đi theo các giá trị thuộc tính trên cây, cây quyết định sẽ cho biết giá trị dự đoán. Nhóm thuật toán cây quyết định có một điểm mạnh đó là có thể sử dụng cho cả bài toán Phân loại (Classification) và Hồi quy (Regression). Thuật toán cây quyết định vẫn được sử dụng rộng rãi trong một số bài toán hiện nay.

Vào năm 2001, Breiman đưa ra thuật toán rừng ngẫu nhiên (random forest) [3]. Rừng ngẫu nhiên là một thuật toán học có giám sát. Như tên gọi của nó, rừng ngẫu nhiên sử dụng các cây quyết định để làm nền tảng. Rừng ngẫu nhiên là một tập hợp của các cây quyết định, mà mỗi cây được chọn theo một thuật toán dựa vào ngẫu nhiên.

Trước đó, Support Vector Machines (SVM) [4] được đề xuất bởi Vladimir N. Vapnik và các đồng nghiệp của ông tại Nga và sau đó trở nên phổ biến trong những năm 90 nhờ ứng dụng giải quyết các bài toán phi tuyến tính. SVM sử dụng không gian giả thuyết các hàm tuyến tính trên không gian đặc trưng nhiều chiều, dựa trên lý thuyết tối ưu và lý thuyết thống kê. Không gian dữ liệu nhập ban đầu sẽ được ánh xạ vào không gian đặc trưng và trong không gian đặc trưng này mặt siêu phẳng phân chia tối ưu sẽ được xác định.

Khi xây dựng dữ liệu đặc trưng để phân loại, tiêu chí quan trọng là cần đảm bảo không để mất nhiều thông tin cũng như không quá tốn kém về mặt chi phí. Năm 1901, Karl Pearson tạo ra thuật toán Phương pháp phân tích thành phần chính - Principle Component Analysis (PCA) [5], với mục đích giải quyết vấn đề dữ liệu có quá nhiều chiều dữ liệu, cần giảm bớt chiều dữ liệu nhằm tăng tốc độ xử lý, nhưng vẫn giữ lại thông tin nhiều nhất có thể (high variance). Hiện nay phương pháp hàm nhân đã được dùng để tăng khả năng áp dụng PCA khi giải quyết các bài toán phi tuyến. Phương pháp này đã được Schölkopf và đồng nghiệp của ông [6] đưa ra với tên gọi là KPCA vào năm 1998.

B. Mô hình học sâu

Học sâu là tập con của học máy trong trí tuệ nhân tạo, có các mạng lưới có khả năng "học" mà không bị giám sát

Tác giả liên hệ: Lê Minh Hóa

Email: hoasac@ptithcm.edu.vn

Đến tòa soạn: 10/2020, chỉnh sửa: 11/2020, chấp nhận đăng: 12/2020

từ dữ liệu không có cấu trúc hoặc không được gắn nhãn. Thời kỳ đầu mặc dù có nhiều mô hình thuật toán học sâu được đưa ra trong khoảng 1986 cho đến 2012, học sâu vẫn không được triển khai rộng rãi do nhiều vấn đề nảy sinh như là vấn đề mất mát đạo hàm, không đủ tập mẫu để huấn luyện hay hiệu suất tính toán của máy tính còn thấp [7]. Cuộc thi phân loại ảnh (ILSVRC) năm 2012 đem lại đột phá cho học sâu khi đại diện là mạng AlexNet [8] gây bất ngờ khi có độ lỗi phân lớp top 5 giảm hơn 10% khi cạnh tranh với các đối thủ sử dụng thuật toán truyền thống xử lý ảnh kết hợp các công cụ trích lọc đặc trưng (SIFT, SURF, FAST, BRISK, AKAZE, ...) và SVM để phân loại các bức ảnh.

Mạng nơ-ron tích chập - Convolutional Neural Networks (CNN) được AlexNet sử dụng, là một trong những mô hình học sâu phổ biến nhất và có ảnh hưởng nhiều nhất trong cộng đồng thị giác máy tính. CNN được sử dụng nhiều trong các bài toán nhận dạng các đối tượng trong ảnh. Năm 1998, Yan LeCun lần đầu huấn luyện mô hình CNN với thuật toán lan truyền ngược cho bài toán nhận dạng chữ viết tay [9].

III. GIẢI PHÁP PHÂN LOẠI

A. Support Vector Machine (SVM)

Support Vector Machine (SVM) được phát triển bởi Vapnik dựa trên lý thuyết học thống kê. Bản chất của phương pháp SVM là chuyển không gian dữ liệu ban đầu thành một không gian mới hữu hạn chiều mà ở đó cho khả năng phân lớp dễ dàng hơn. Điểm làm SVM hiệu quả hơn các phương pháp khác chính là việc SVM không còn bị giới hạn bởi việc phân lớp một cách tuyến tính, hay nói cách khác các siêu phẳng có thể được hình thành từ các hàm phi tuyến.

Để phân loại tốt nhất thì phải xác định siêu phẳng nằm ở càng xa các điểm dữ liệu của tất cả các lớp càng tốt, vì nói chung lề (margin) càng lớn thì sai số tổng quát hóa của thuật toán phân loại càng bé:

$$\langle w, x_i \rangle + b \geq 1 \quad \text{nếu } (y_i = 1) \tag{1}$$

$$\langle w, x_i \rangle + b \leq -1 \quad \text{nếu } (y_i = -1) \tag{2}$$

Mô hình được tổng quát lên không gian nhiều chiều. Khoảng cách từ một điểm (vector) bất kỳ có tọa độ x_i tới siêu mặt phẳng có phương trình $w^T x + b = 0$ được xác định bởi:

$$d(w, b; x_i) = \frac{|\langle w, x_i \rangle + b|}{\|w\|} \tag{3}$$

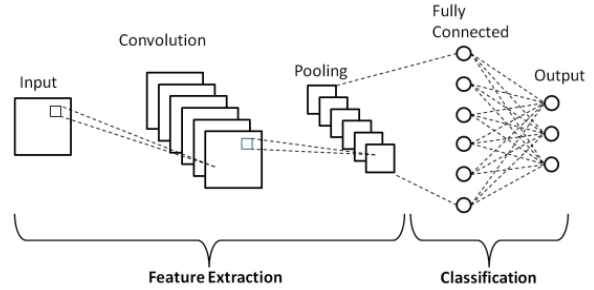
Bài toán tối ưu trong SVM chính là bài toán tìm w và b sao cho margin này đạt giá trị lớn nhất:

$$\min_{x_i; y_i=1} d(w, b; x_i) + \min_{x_i; y_i=-1} d(w, b; x_i) \tag{4}$$

Việc giải trực tiếp bài toán này sẽ rất phức tạp, nhưng trên lý thuyết có cách để đưa nó về bài toán đơn giản hơn là tính toán $\|w\|$ để đạt được giá trị cực tiểu. Tuy nhiên, việc giải bài toán này trở nên phức tạp khi số chiều của không gian dữ liệu và số điểm dữ liệu tăng lên cao. Khi đó để tìm nghiệm thường giải bài toán đối ngẫu của bài toán này.

B. Mạng nơ-ron tích chập (CNN)

Mô hình Mạng CNN [12] là một tập hợp các lớp tích chập chồng lên nhau và sử dụng các hàm kích hoạt phi tuyến như ReLU và tanh để kích hoạt các trọng số trong các node. Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo. Mỗi một lớp được sử dụng các bộ lọc khác nhau, thông thường có hàng trăm hàng nghìn bộ lọc như vậy và kết hợp kết quả của chúng lại.



Hình 1. Mô hình Mạng nơ-ron tích chập [12]

Lớp tích chập $y(t)$ sử dụng quá trình tích chập các bộ lọc trên toàn bộ ma trận ảnh $u(t) * x(t)$, có thể biểu thị chung bằng công thức toán học như sau:

$$y(t) = u(t) * x(t) = \int u(\tau)x(t - \tau)d\tau \tag{5}$$

Khi (5) áp dụng trên ma trận ảnh, thì công thức có thể viết lại thành:

$$y[k] = \sum_j u[j]x[k - j] \tag{6}$$

Ở (6), $x[k]$ có thể được viết là $x[k-j]$, mặt khác nó có thể được chuyển thành $x[k+j]$ cũng không ảnh hưởng đến kết quả tính toán:

$$y[k] = \sum_j u[j]x[k + j] \tag{7}$$

Trong quá trình tích chập của mạng nơ-ron, w đóng vai trò là một bộ lọc trong lớp tích chập, x là đầu vào của lớp này và $f(.)$ là hàm kích hoạt. Sải bước (ξ) là khoảng cách giữa 2 kernel khi quét. Với sải bước bằng 1, kernel sẽ quét 2 ô ngay cạnh nhau, nhưng với sải bước bằng 2, kernel sẽ quét ô số 1 và ô số 3, bỏ qua ô ở giữa. Khi đó (7) được thay thế bằng công thức:

$$y_n[k] = f(w * x_n) = f(\sum_j w[j]x_n[\xi k + j]) \tag{8}$$

Ngoài ra có một số lớp khác để giảm kích thước tính toán bằng cách sử dụng lấy mẫu gộp dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu). Trong quá trình huấn luyện mạng CNN tự động học các giá trị qua các lớp tích chập. Ví dụ trong tác vụ phân lớp ảnh, CNN sẽ cố gắng tìm ra thông số tối ưu cho các bộ lọc tương ứng theo thứ tự: điểm ảnh thô > cạnh > hình dạng > khuôn mặt > đặc trưng mức độ cao. Lớp cuối cùng được dùng để phân lớp hình ảnh.

C. Hồi quy Softmax

Softmax [13] được đặt ở cuối mạng học sâu dùng phổ biến cho việc phân loại. Vector đặc trưng của những lớp trước là đầu vào cho lớp Softmax này. Phương pháp này được sử dụng rộng rãi khi cần phân loại nhiều nhóm đối tượng. Để biểu diễn mô hình gọn hơn, mô hình Softmax sử dụng ký hiệu đại số tuyến tính, biểu diễn dưới dạng vector $q=Wx+b$, một dạng phù hợp hơn cho cả toán học và lập trình:

$$q_n = W^T \hat{x}_n + w_0 = [q_{n,1} q_{n,2} \dots q_{n,k}]^T \quad (9)$$

Một hàm softmax có thể được sử dụng để chuyển đổi giá trị thành xác suất. Các hàm thường dùng là

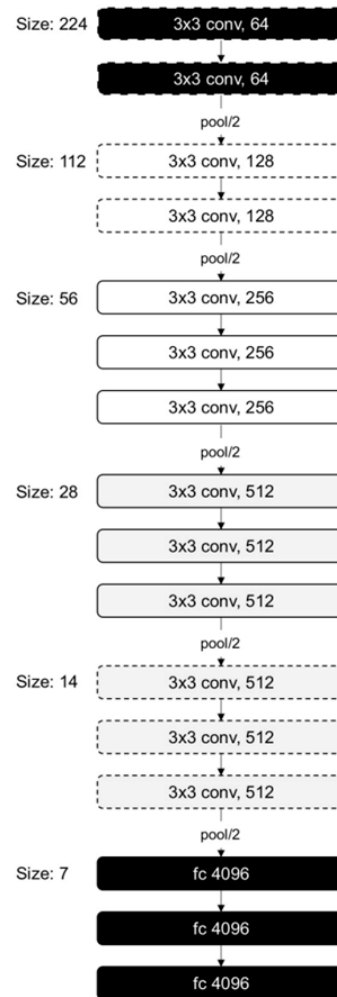
$$P(y_n = c|x_n) = \frac{\exp(q_{n,c})}{\sum_{j=1}^k \exp(q_{n,j})} \quad (10)$$

Tiếp theo sẽ tối ưu hóa các tham số của mô hình sao cho khả năng xuất hiện dữ liệu quan sát được là cao nhất. Sau đó, mô hình sẽ đưa ra dự đoán bằng cách đặt ngưỡng xác suất, ví dụ dự đoán nhãn đúng là nhãn có xác suất cao nhất.

D. Mô hình VGG16

Kể từ AlexNet, các kiến trúc CNN ngày càng sâu hơn, như là VGG [10] ra đời với một số cải tiến, trước tiên là mô hình VGG sẽ sâu hơn, tiếp theo là thay đổi trong thứ tự tích chập. Tuy nhiên, tăng độ sâu mạng không chỉ đơn giản là xếp chồng các lớp lại với nhau. Mạng sâu rất khó huấn luyện vì vấn đề mất mát đạo hàm, vì độ dốc được truyền ngược trở lại các lớp trước đó, phép nhân lặp đi lặp lại có thể làm cho độ dốc cực nhỏ. Kết quả là, hiệu suất của mạng bị bão hòa hoặc giảm hiệu suất nhanh chóng.

Trước AlexNet đều sử dụng tích chập kết hợp gộp cực đại còn VGG thì sử dụng 1 chuỗi tích chập liên tiếp ở giữa và cuối của kiến trúc VGG. Việc này sẽ làm cho việc tính toán trở nên lâu hơn nhưng những đặc trưng sẽ vẫn được giữ lại nhiều hơn so với việc sử dụng gộp cực đại sau mỗi tích chập. Hơn nữa hiện nay với sự ra đời của GPU giúp tốc độ tính toán trở nên nhanh hơn rất nhiều lần thì vấn đề này không còn đáng lo ngại. VGG cho sai số nhỏ hơn AlexNet trong cuộc thi ILSVRC năm 2014. VGG có 2 phiên bản là VGG16 và VGG19. Kiến trúc VGG16 [10] được biểu diễn ở Hình 2 dưới đây.



Hình 2. Kiến trúc mô hình VGG16

Kiến trúc của VGG16 bao gồm 16 lớp: 13 lớp tích chập (2 lớp conv-conv, 3 lớp conv-conv-conv) đều có kernel 3x3, sau mỗi lớp conv là gộp cực đại giảm kích thước ảnh xuống 0.5, và 3 lớp kết nối hoàn chỉnh. VGG19 tương tự như VGG16 nhưng có thêm 3 lớp tích chập.

IV. KẾT QUẢ THỰC NGHIỆM

Nghiên cứu tiến hành thử nghiệm các thuật toán phân loại đã được đề cập ở phần trên, bộ dữ liệu hình ảnh TrashNet được dùng để huấn luyện và đánh giá kết quả. Bộ dữ liệu TrashNet có các hình ảnh rác thải sinh hoạt do G. Thung và đồng sự tập hợp [11]. Bộ dữ liệu tổng cộng có 2527 bức hình chia làm 6 loại rác, trong đó có 5 loại tái chế được, gồm 501 hình rác thủy tinh, 594 hình giấy thải, 403 hình bìa các tông, 410 hình rác kim loại và 137 rác sinh hoạt. Bộ ảnh đã được chuẩn hóa kích thước là 512x384px. Hình ảnh ví dụ các loại rác thải trong Hình 3.



Hình 3. Hình mẫu của Bộ dữ liệu TrashNet (a) giấy bìa các tông (cardboard) (b) thủy tinh (glass) (c) kim loại (metal) (d) giấy (paper) (e) nhựa (plastic) (f) rác sinh hoạt (trash)

Nghiên cứu này sử dụng phần cứng như sau: CPU core i7 3.6Ghz, bộ nhớ 16GB Ram, card đồ họa là GTX 1070 và được cài đặt Windows 10. Mã nguồn cho thực nghiệm cài đặt Python 3.7 và sử dụng thư viện TensorFlow2.

Độ chính xác trong nhiệm vụ phân loại rác thải và thời gian huấn luyện mô hình là tham số chính được so sánh giữa các thuật toán; đồng thời những mặt hạn chế của nó cũng được xem xét.

Bộ dữ liệu được chia theo tỉ lệ 8:2, với 8 phần cho mục đích huấn luyện và 2 phần cho kiểm thử đánh giá độ chính xác.

A. Hiệu suất của các thuật toán học máy

Các thuật toán đại diện trong học máy như cây quyết định, rừng ngẫu nhiên, SVMs và PCA được sử dụng để huấn luyện và phân loại bộ dữ liệu. Đầu tiên, khi xử lý dữ liệu hình ảnh sẽ chuyển ma trận ảnh 2 chiều thành ma trận 1 chiều đầu vào cho quá trình huấn luyện và phân loại. Khi sử dụng thuật toán liên quan cây quyết định thì kích thước

ma trận 1 chiều được điều chỉnh thích hợp sẽ gia tăng được độ chính xác và tốc độ phân loại. Kích thước hình ảnh của bộ dữ liệu TrashNet là 512x384px, khi được giảm xuống còn 28x28px để làm đầu vào, thì thời gian huấn luyện của cây quyết định là 2.6s và độ chính xác phân loại đạt được là 49.3%. Tuy nhiên khi kích thước ảnh đầu vào là 128x128px thì thời gian huấn luyện tăng lên tới 48.2s nhưng độ chính xác chỉ còn 47.3%.

Khi cùng là kích thước hình ảnh đầu vào giảm xuống 128x128px thì thuật toán rừng ngẫu nhiên cho kết quả là tốt nhất khi độ chính xác phân loại đạt 68.2% với thời gian huấn luyện là 41.3s. Ngược lại là mô hình SVM cho hiệu suất thấp, thời gian huấn luyện dài và độ chính xác là thấp nhất trong các mô hình. Do chiều dữ liệu lớn nên vấn đề SVM gặp phải là gia tăng khối lượng tính toán. Do đó, ở đây nếu kết hợp với PCA để giảm chiều dữ liệu mà vẫn giữ các đặc trưng quan trọng thì phân loại sử dụng SVM chứng tỏ có hiệu quả hơn khi thời gian huấn luyện giảm xuống chỉ còn xấp xỉ 29s và độ chính xác được cải thiện ở mức 60.8%.

BẢNG 1. HIỆU SUẤT CÁC THUẬT TOÁN HỌC MÁY

Thuật toán	Độ chính xác (%)	Thời gian huấn luyện (s)
Cây quyết định	47.3	48.2
Rừng ngẫu nhiên	68.2	41.3
SVM	41.6	215.4
SVM+PCA	60.8	28.7

B. Hiệu suất của các mô hình học sâu

Trong trường hợp học sâu, mô hình CNN và VGG16 được sử dụng để huấn luyện và phân loại bộ dữ liệu. Hình ảnh đầu vào được giữ nguyên kích thước đã được chuẩn hóa là 384x512px với 3 kênh màu. CNN cho kết quả tốt hơn với độ chính xác cao hơn là 60.2% với số lần huấn luyện lặp lại là 20 lần trong khoảng thời gian huấn luyện là 362.8s. Một điều lưu ý là khi số lần huấn luyện lớn hơn, mô hình bị hiện tượng quá khớp (overfitting). Hiện tượng quá khớp này diễn ra trên mô hình VGG16 sớm hơn dẫn đến tình trạng độ chính xác trong phân loại các bức ảnh thấp khi cùng là 20 lần lặp huấn luyện. Điều này chứng tỏ bộ dữ liệu có số lượng mẫu nhỏ không thích hợp cho mô hình có độ phức tạp tương đối cao. Bài báo này thêm một lớp phân loại softmax vào mạng VGG16 thì kết quả thực nghiệm cho kết quả phân loại khả quan hơn với độ chính xác đạt 71.1% trong thời gian huấn luyện 545.2s với số lần huấn luyện là 20.

Sau 20 lần lặp, độ chính xác không có xu hướng tăng nữa, đồng thời kiểm tra sai số cũng không giảm và nó duy trì cùng số liệu.

BẢNG II. HIỆU SUẤT CÁC MÔ HÌNH HỌC SÂU

Mô hình	Độ chính xác (%)	Thời gian huấn luyện (s)
CNN	60.2	362.8
VGG16	23.6	1248.5
VGG16+softmax	71.1	545.2

V. KẾT LUẬN

Trong bài báo này, các thuật toán trí tuệ nhân tạo được nghiên cứu và thử nghiệm trong bài toán phân loại rác thải. Kết quả được đánh giá so sánh thông qua độ chính xác trong phân loại và thời gian huấn luyện của thuật toán. Thông qua kết quả thực nghiệm, các mô hình học sâu hứa hẹn trong tương lai gần có thể đáp ứng được các tác vụ công nghiệp thời gian thực. Từ nghiên cứu này, việc cài đặt nâng cấp mô hình và huấn luyện bộ dữ liệu lớn hơn để đạt được mục tiêu phân loại chính xác và tin cậy hơn, mang tính thực tiễn nhằm áp dụng cho ngành công nghiệp xử lý rác thải, đem lại môi trường sống tốt đẹp hơn.

TÀI LIỆU THAM KHẢO

- [1] J.R. Quinlan, "Induction of decision trees", Mach Learn 1, 81–106 (1986), <https://doi.org/10.1007/BF0011625>
- [2] S.L. Salzberg, "C4.5: Programs for Machine Learning" by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. Mach Learn 16, 235–240 (1994), <https://doi.org/10.1007/BF00993309>
- [3] L. Breiman, "Random Forests", Machine Learning 45, 5–32 (2001), <https://doi.org/10.1023/A:1010933404324>
- [4] C. Cortes, V. Vapnik, "Support-vector networks", Mach Learn 20, 273–297 (1995), <https://doi.org/10.1007/BF00994018>
- [5] Karl Pearson F.R.S. (1901), "LIII. On lines and planes of closest fit to systems of points in space", The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2:11, 559-572, DOI: 10.1080/14786440109462720
- [6] B. Schölkopf, A. Smola, K.R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", Neural Computation 1998 10:5, 1299-1319, DOI: 10.1162/089976698300017467
- [7] R. Garg, H. Aggarwal, P. Centobelli, R. Cerchione, "Extracting Knowledge from Big Data for Sustainability: A Comparison of Machine Learning Techniques", Sustainability 2019, 11, 6669, DOI: 10.3390/su11236669
- [8] A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", In Advances in Neural Information Processing Systems 25, edited by F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger, 1097–1105, Curran Associates, Inc., 2012.
- [9] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, L. D. Jackel, "Backpropagation Applied to Handwritten Zip Code Recognition", Neural Computation 1989 1:4, 541-551, <https://doi.org/10.1162/neco.1989.1.4.541>
- [10] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", ICLR 2015, arXiv:1409.1556
- [11] G. Thung, "Trashnet," GitHub repository, 2016
- [12] LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. Nature, 521(7553), pp.436-444
- [13] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning." MIT Press, 2016

PERFORMANCE EVALUATION ON ARTIFICIAL INTELLIGENT MODELS IN WASTE CLASSIFICATION

Abstract: Waste classification is a big problem in computer vision and nowadays there are many approaches to offer solutions, in which the direction of using artificial intelligence reaches a significant level of accuracy. In this paper, the classification algorithms in machine learning such as decision trees, random forest, SVM, PCA and the typical deep learning model VGG16 are studied to compare the efficiency in the classification. This study proposes to add a Softmax classification after VGG16 to give higher accuracy results. The deep learning models studied here use the Convolutional Neural Network (CNN) core which is one of the most advanced networks in computer vision, demonstrating the ability to effectively classify images. The results from the proposed model have been improved with 71.1% accuracy compared to using the CNN model under the condition utilizing dataset with small sample quantity. In the near future, deep learning models will mechanically assist in automatic garbage classification and without necessarily human intervention.

Keywords: CNN, Decision tree, Random forest, PCA, waste classification, SVM, VGG16



Lê Minh Hóa, tốt nghiệp Thạc sỹ Máy tính, chuyên ngành Đa phương tiện năm 2011, tại trường Soongsil, Hàn Quốc. Hiện là giảng viên Học Viện Công Nghệ Bưu Chính Viễn Thông, cơ sở TP. Hồ Chí Minh. Lĩnh vực nghiên cứu: HCI, trí tuệ nhân tạo, học sâu.