

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

NGUYỄN TRỌNG THUẬN - 18521471

NGUYỄN QUANG THUẬN – 18521470

ĐINH THANH TOÀN - 18521504

ĐỒ ÁN CUỐI KỲ
MÔN HỌC: MÁY HỌC
ĐỀ TÀI: PHÁT HIỆN Ồ GÀ TRÊN ĐƯỜNG

TP. HỒ CHÍ MINH, 2020

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

NGUYỄN TRỌNG THUẬN - 18521471

NGUYỄN QUANG THUẬN – 18521470

ĐINH THANH TOÀN – 18521504

ĐỒ ÁN CUỐI KỲ
MÔN HỌC: MÁY HỌC
ĐỀ TÀI: PHÁT HIỆN Ồ GÀ TRÊN ĐƯỜNG

GIẢNG VIÊN HƯỚNG DẪN

PGS.TS. LÊ ĐÌNH DUY
THS. PHẠM NGUYỄN TRƯỜNG AN

TP. HỒ CHÍ MINH, 2020

MỤC LỤC

DANH MỤC HÌNH

DANH MỤC BẢNG

Chương 1. Giới thiệu bài toán

Dân Trí > Xã hội > Giao thông

Thứ tư, 12/12/2018 - 10:36

Phú Yên:

Tai nạn thương tâm vì “ổ gà” bao giờ mới chấm dứt?

Hình 1. Nguy hiểm từ “ổ gà”

Cùng với sự phát triển của nền kinh tế, nhu cầu của vận tải đường bộ không ngừng gia tăng. Để đáp ứng nhu cầu trên cần số lượng lớn xe tải đầu kéo tham gia giao thông, dẫn đến kết quả là số lượng trực xe cũng như tải trọng trực xe thực tế lớn hơn nhiều so với dự tính ban đầu, tạo nên một áp lực không nhỏ lên mặt đường. Đây chính là nguyên nhân làm cho kết cấu mặt đường nhanh chóng hư hỏng mà thường gặp nhất là hiện tượng mặt đường xuất hiện các “ổ gà”. Sự xuất hiện của các “ổ gà” trên mặt đường ít nhiều đã gây ra những nguy hiểm cho người tham giao thông. Từ đó, bài toán được đặt ra là xây dựng một thuật toán máy học có thể nhận dạng đường trên đường có xuất hiện của “ổ gà” hay không.

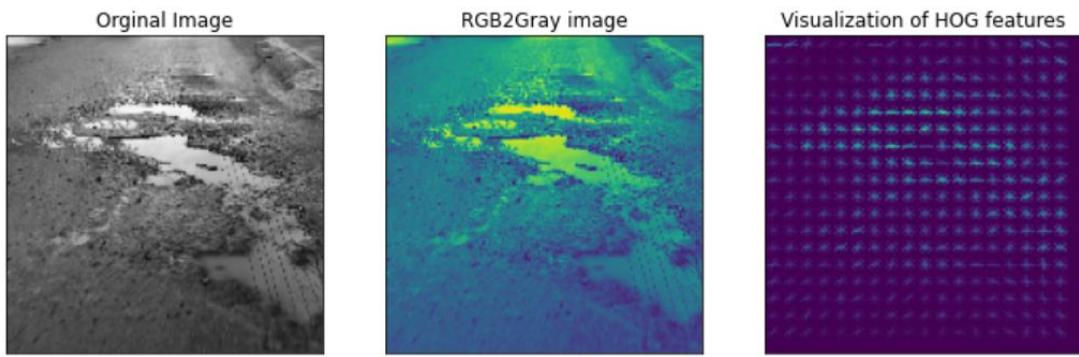
Chương 2. Cơ sở lý thuyết

2.1. Trích xuất đặc trưng: HOG

HOG là viết tắt của Histogram of Oriented Gradient - một loại trích xuất đặc trưng. Mục đích của trích xuất đặc trưng là trừu tượng hóa đối tượng bằng cách trích xuất ra những đặc trưng của đối tượng đó và bỏ đi những thông tin không hữu ích. Vì vậy, HOG được sử dụng chủ yếu để mô tả hình dạng và sự xuất hiện của một đối tượng trong ảnh.

Bản chất của phương pháp HOG là sử dụng thông tin về sự phân bố của các cường độ gradient (intensity gradient) hoặc của hướng biên (edge directins) để mô tả các đối tượng cục bộ trong ảnh. Các toán tử HOG được cài đặt bằng cách chia nhỏ

một bức ảnh thành các vùng con, được gọi là “tế bào” (cells) và với mỗi cell, ta sẽ tính toán một histogram về các hướng của gradients cho các điểm nằm trong cell. Ghép các histogram lại với nhau ta sẽ có một biểu diễn cho bức ảnh ban đầu. Để tăng cường hiệu năng nhận dạng, các histogram cục bộ có thể được chuẩn hóa về độ tương phản bằng cách tính một ngưỡng cường độ trong một vùng lớn hơn cell, gọi là các khối (blocks) và sử dụng giá trị ngưỡng đó để chuẩn hóa tất cả các cell trong khối. Kết quả sau bước chuẩn hóa sẽ là một vector đặc trưng có tính bất biến cao hơn đối với các thay đổi về điều kiện ánh sáng.



Hình 2. Trích xuất đặc trưng với HOG.

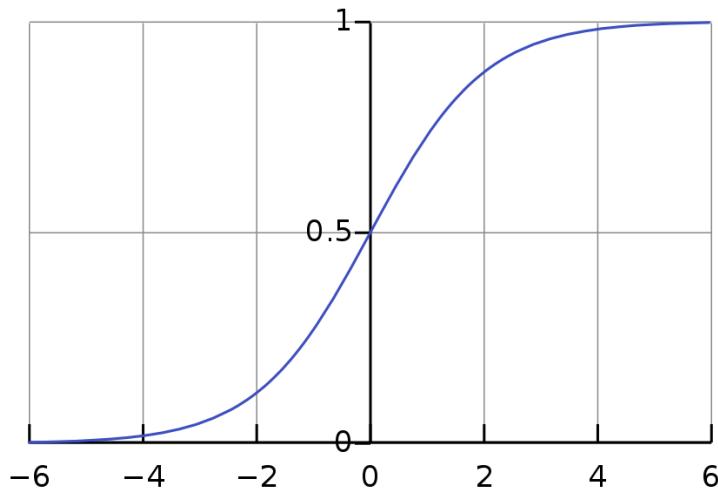
2.2. Hồi quy Logistic (Logistic Regression)

Logistic Regression là một mô hình sử dụng cho các bài toán phân loại. Logistic Regression được biến đổi một chút từ Linear Regression (hồi quy tuyến tính đa biến), bằng cách cho kết quả của Linear Regression vào hàm *sigmoid*, cụ thể:

$$y = \text{sigmoid}(f(X)) = \text{sigmoid}(w_0 + w_1x_1 + \dots + w_nx_n)$$

Trong đó:

- w_0, w_1, \dots, w_n là các tham số của mô hình
- x_0, x_1, \dots, x_n là các biến (đặc trưng – biến độc lập) đầu vào
- y là kết quả đầu ra (biến phụ thuộc)
- $\text{sigmoid}(x) = \frac{1}{1+e^{-x}}$

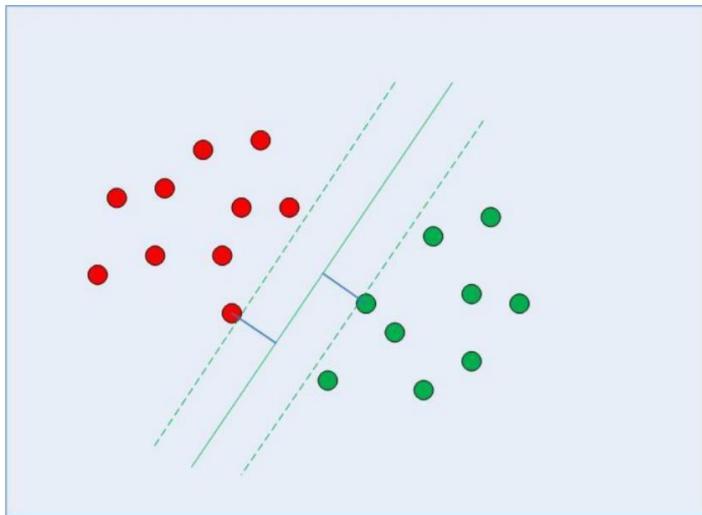


Hình 3. Hàm sigmoid.

Hàm sigmoid là bởi vì nó tồn tại trong khoảng giữa 0 và 1. Do đó, nó đặc biệt được sử dụng cho các mô hình mà phải dự đoán xác suất đầu ra.

2.3. Support Vector Machine (SVM)

Support Vector Machine là một trong các thuật toán phân loại được sử dụng phổ biến nhất trong Machine Learning. Đây là một thuật toán phân loại tuyến tính. Tức là sau khi huấn luyện, ta thu được các siêu phẳng phân chia các lớp dữ liệu với nhau, giống như thuật toán Logistic Regression. Mục tiêu của thuật toán này không những phân chia được các lớp dữ liệu với nhau, mà còn tìm cách để tối đa khoảng cách giữa đường phân chia với các điểm dữ liệu giữa các lớp (maximum margin).



Hình 4. Mô hình SVM

Thuật toán SVM sẽ tìm một số vector đặc biệt (gọi là support vectors). Mô hình (Model) dự đoán (predict) kết quả đầu ra của những điểm dữ liệu mới dựa trên các vector đặc biệt này. Điểm đặc biệt của Support Vector Machine:

Hầu hết các thuật toán Machine Learning khác đều phân chia dữ liệu dựa trên các điểm dữ liệu đặc trưng nhất của lớp dữ liệu đó.

Trong khi đó, Support Vector Machine phân chia dữ liệu dựa trên các điểm dữ liệu dễ gây nhầm lẫn nhất giữa các lớp dữ liệu.

2.4. Các độ đo đánh giá mô hình

Confusion matrix

		Dự đoán	
		1	0
Thực tế	1	TP	FN
	0	FP	TN

Quan sát confusion matrix, có được các thông tin sau

- TP (true positive) – điểm dữ liệu có nhãn thật sự là 1 được dự đoán là 1.
- FN (false negative) – điểm dữ liệu có nhãn thật sự là 1 được dự đoán là 0.

- FP (false positive) – điểm dữ liệu có nhãn thật sự là 0 được dự đoán là 1.
- TN (true negative) – điểm dữ liệu có nhãn thật sự là 0 được dự đoán là 1.

Từ đó, tính được độ chính xác của mô hình:

$$Accuracy = \frac{TP + TN}{(TP + FN + FP + TN)}$$

Tỉ lệ điểm dữ liệu TP trong số những điểm được phân loại là positive:

$$Precision = \frac{TP}{(TP + FP)}$$

Tỉ lệ điểm dữ liệu TP trong số những điểm thực sự là positive:

$$Recall = \frac{TP}{TP + FN}$$

Precision và Recall đều là các số không âm nhỏ hơn hoặc bằng một. Precision cao đồng nghĩa với việc độ chính xác của các điểm tìm được là cao. Recall cao đồng nghĩa tỉ lệ bỏ sót các điểm thực sự positive là thấp.

Khi Precision = 1, mọi điểm tìm được đều thực sự là positive, tức không có điểm negative nào lẫn vào kết quả. Tuy nhiên, Precision = 1 không đảm bảo mô hình là tốt, vì câu hỏi đặt ra là liệu mô hình đã tìm được tất cả các điểm positive hay chưa. Nếu một mô hình chỉ tìm được đúng một điểm positive mà nó chắc chắn nhất thì ta không thể gọi nó là một mô hình tốt.

Khi Recall = 1, mọi điểm positive đều được tìm thấy. Tuy nhiên, đại lượng này lại không đo liệu có bao nhiêu điểm negative bị lẫn trong đó. Nếu mô hình phân loại mọi điểm là positive thì chắc chắn Recall = 1.

Một mô hình phân lớp tốt là mô hình có cả Precision và Recall đều cao, tức càng gần 1 càng tốt.

F1-score, là harmonic mean của precision và recall (giả sử hai đại lượng này khác 0):

$$F_1 = 2 \frac{precision \times recall}{precision + recall}$$

F1-score có giá trị nằm trong nửa khoảng [0,1]. F_1 càng cao, bộ phân lớp càng tốt. Khi cả recall và precision đều bằng 1 (tốt nhất có thể).

Chương 3. Thực nghiệm

3.1. Thu thập dữ liệu

Dữ liệu từ thu thập bằng cách chụp lại tình trạng mặt đường khi đang di chuyển trên đường. Chụp lại hình ảnh có “ổ gà” hoặc không, nếu có “ổ gà” thì ổ gà cách xa 5 – 7m, góc chụp là dưới ngực. Toàn bộ ảnh được chụp vào ban ngày trong điều kiện không âm u hay mờ tối.

Sau khi đã chụp được hình ảnh cần thiết, tiến hành gán nhãn cho ảnh. Bộ dữ liệu gồm có 2 lớp dữ liệu: 0 (đại diện cho mặt đường không có xuất hiện “ổ gà”) và 1 (đại diện cho hình ảnh có tồn tại ổ gà).

Kết quả cuối cùng tạo được một bộ dữ liệu gồm 300 hình ảnh, trong đó 150 hình ảnh là được gán nhãn là 1 và 150 hình ảnh được gán nhãn là 0.

Bộ dữ liệu sẽ được chia làm hai phần theo tỉ lệ 80% nhằm sử dụng vào mục đích huấn luyện (bao gồm 120 hình ảnh có nhãn là 1 và 120 hình ảnh có nhãn là 0), 20% được sử dụng vào mục đích kiểm tra/ đánh giá (bao gồm 30 hình ảnh từng nhãn tương ứng).

3.2. Tiền xử lý dữ liệu

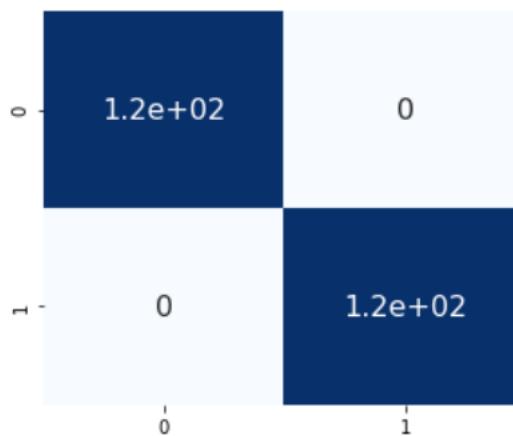
Toàn bộ ảnh sẽ được resize về cùng một kích thước là (150, 150), đồng thời chuyển toàn bộ hình ảnh về định dạng thang độ xám (grayscale).

Hình ảnh thang độ xám (grayscale) là giá trị của mỗi pixel là một mẫu duy nhất chỉ đại diện cho một lượng ánh sáng, tức là nó chỉ mang thông tin cường độ ảnh sáng tại điểm đó. Hình ảnh thang độ xám, một loại đơn sắc đen trắng hoặc xám, được tạo thành từ các sắc thái của màu xám. Độ tương phản nằm trong khoảng từ đen ở cường độ yếu nhất đến trắng ở mức mạnh nhất. Hình ảnh thang độ xám khác với hình ảnh đen trắng là hình ảnh chỉ có hai màu: đen và

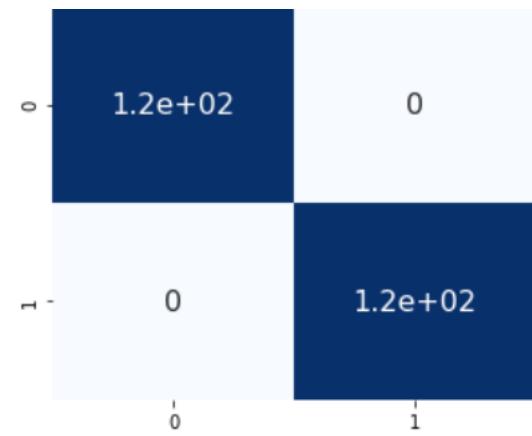
trắng (còn được gọi là hình ảnh nhị phân). Hình ảnh thang độ xám có nhiều sắc độ xám ở giữa. Hình ảnh thang độ xám có thể là kết quả của việc đo cường độ ánh sáng ở mỗi pixel theo sự kết hợp tần số (hoặc bước sóng) có trọng số riêng, chúng là đơn sắc phù hợp khi chỉ có một tần số duy nhất. Về nguyên tắc, các tần số có thể từ bất kỳ nơi nào trong phổ điện từ (ví dụ: hồng ngoại, ánh sáng nhìn thấy, tia cực tím, v.v.). Hình ảnh thang độ xám (hay cụ thể hơn là trắc quang) là hình ảnh có không gian màu thang độ xác định, ánh xạ các giá trị mẫu số được lưu trữ vào kênh màu sắc của không gian màu tiêu chuẩn, dựa trên các đặc tính đo được của thị giác con người.

3.3. Đánh giá kết quả

3.3.1. Đánh giá trên tập dữ liệu huấn luyện



Hình 5. Confusion matrix trên tập dữ liệu huấn luyện (HOG + Logistic Regression)



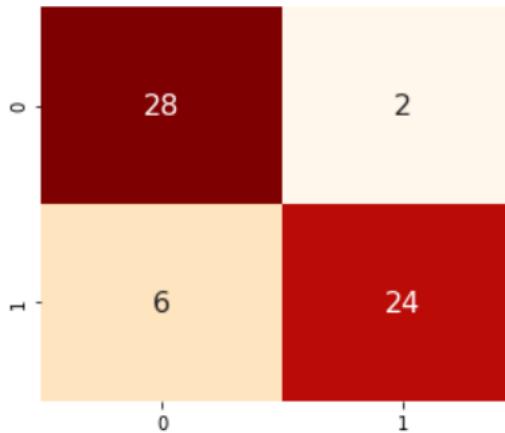
Hình 6. Confusion matrix trên tập dữ liệu huấn luyện (HOG + SVM)

Phương pháp Độ đo	HOG + Hồi quy Logistic	HOG + SVM
Accuracy	1.0	1.0
Precision	1.0	1.0

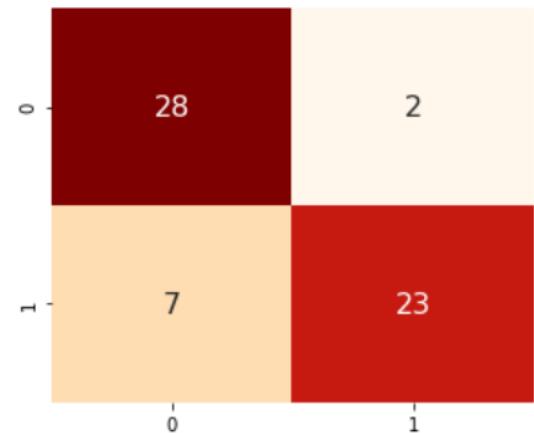
Recall	1.0	1.0
F1	1.0	1.0

Bảng 1. Đánh giá mô hình phân lớp trên tập dữ liệu huấn luyện.

3.3.2. Đánh giá trên tập dữ liệu kiểm tra



Hình 7. Confusion matrix trên tập dữ liệu kiểm tra (HOG + Logistic Regression)



Hình 8. Confusion matrix trên tập dữ liệu kiểm tra (HOG + SVM)

Phương pháp Độ đo	HOG + Hồi quy Logistic	HOG + SVM
Accuracy	0.87	0.85
Precision	0.92	0.92
Recall	0.8	0.76
F1	0.86	0.84

Bảng 2. Đánh giá mô hình phân lớp trên tập dữ liệu kiểm tra.

3.3.3. Đánh giá trên bộ dữ liệu thu thập ngẫu nhiên trên mạng

Phương pháp Độ đo	HOG + Hồi quy Logistic	HOG + SVM
Accuracy	0.60	0.65

Precision	0.56	0.59
Recall	0.90	1.00
F1	0.69	0.74

Chương 4. Nhận xét và hướng phát triển

4.1. Nhận xét.

- Cả 2 mô hình dự đoán khá tốt trên tập dữ liệu với độ chính xác 87% cho mô hình logistic regression và 85% cho mô hình SVM.
- Tuy nhiên, đối với những data thu thập ngẫu nhiên trên mạng, cả 2 mô hình lại rất dễ rơi vào trường hợp dương tính giả khi có tới 7/10 ảnh không có ổ gà được dự đoán thành có ổ gà.
- Bộ dữ liệu không được đa dạng về góc chụp cũng như điều kiện ánh sáng, nên khi gặp những bức ảnh có điều kiện sáng hoặc góc chụp khác thì dự đoán khá tệ.
- Quá trình trích xuất đặc trưng chưa hợp lý khi tỉ lệ dương tính giả quá cao ở những data thu thập ngẫu nhiên.

4.2. Hướng phát triển.

- Tăng cường dữ liệu cũng như sự đa dạng về góc chụp, ánh sáng, vật thể che khuất.
- Sử dụng mạng nơ-ron tích chập để tăng khả năng nhận diện ổ gà trong ảnh.
- Nâng cấp mô hình để có thể chạy real-time với camera hành trình.

TÀI LIỆU THAM KHẢO

<https://www.analyticsvidhya.com/blog/2019/09/feature-engineering-images-introduction-hog-feature-descriptor/>

<https://towardsdatascience.com/logistic-regression-detailed-overview-46c4da4303bc>

<https://machinelearningcoban.com/2017/04/09/smvy/>

<https://machinelearningcoban.com/2017/08/31/evaluation/>

<https://en.wikipedia.org/wiki/Grayscale>