



ÁP DỤNG PHƯƠNG PHÁP KIỂM TRA DỮ LIỆU NGOẠI LAI VÀO VIỆC PHÁT HIỆN Ổ GÀ BẰNG ĐIỆN THOẠI THÔNG MINH

Phan Minh Đức*

Trường Đại học Sư phạm Huế, Đại học Huế, 34 Lê Lợi, Huế, Việt Nam

Tóm tắt: Trong bối cảnh ngày càng gia tăng hệ thống đường bộ cũng như nhu cầu ngày càng tăng về vận tải đường bộ thì vai trò của các hoạt động bảo trì, cảnh báo và sửa chữa hệ thống đường bộ ngày càng đặc biệt quan trọng. Cùng với sự phổ biến của điện thoại thông minh bao gồm nhiều cảm biến, nhiều con đường đã được hệ thống phát hiện bất thường đã được đề xuất. Bài báo này nhằm mục đích phân tích một số phát hiện bất thường trên hệ thống đường, tôi đề xuất một số phương pháp phát hiện các bất thường trên hệ thống đường bằng các phương pháp phát hiện các giá trị ngoại lai phù hợp cho điện thoại thông minh.

Từ khóa: Ổ gà, đường bất thường, điện thoại thông minh, mạng cảm biến

The application of ourliers to detecting potholes by using smartphones

Phan Minh Đức*

University of Education, Hue University, 34 Le Loi St., Hue, Vietnam

Abstract. In the context of increasing road system as well as increasing demand for road transport, the role of road system maintenance, warning and repair is becoming increasingly important. Along with the popularity of smartphones that include multiple sensors, many methods used to detect the anomalies on the roads have been proposed. This paper aims to analyze some abnormal occurrences on the roads and propose some methods to find out anomalies by detecting outliers suitable for smartphones.

Keywords: Road anomaly, pothole, sensor network, smartphones

* Liên hệ: phanminhduc@dhsphue.edu.vn

Ngày nhận bài: 13-12-2021; Ngày nhận đăng: 20-08-2023

1 Giới thiệu

Trong bối cảnh ngày càng gia tăng hệ thống đường bộ cũng như nhu cầu ngày càng tăng về vận tải đường bộ thì vai trò của các hoạt động bảo trì đường bộ ngày càng quan trọng, đặc biệt là việc tập trung vào công tác bảo trì và sửa chữa hệ thống đường bộ. Dưới tác động trực tiếp của các phương tiện giao thông đi lại như: xe máy, xe hơi, xe đạp và đặc biệt là các xe có trọng tải cực kỳ lớn.... và các điều kiện về thời tiết nên mức độ phục vụ của hệ thống đường sẽ giảm dần đi theo thời gian cùng với quá trình hóa các kết cấu công trình đường bộ như mặt đường, mặt cầu... đặc biệt trong điều kiện thiếu công tác bảo trì phù hợp sẽ ảnh hưởng trực tiếp đến an toàn giao thông khi người điều khiển các phương tiện giao thông cũng như đảm bảo các điều kiện khác như: hư hỏng xe, cảm giác thoải mái khi điều khiển phương tiện giao thông... Việc xác định chất lượng hệ thống đường bộ đặc biệt là việc xác định các ổ gà và tìm các ổ gà để sửa chữa là một thử thách rất khó khăn và rất cần nhiều nỗ lực. Do đó, nhiều hệ thống giám sát đường và xác định ổ gà được hình thành, phát triển và được đề xuất như: Hệ thống phát hiện ổ gà Pothole [1], Hệ thống phát hiện ổ gà Nericell [4], Hệ thống phát hiện ổ gà của Artis Mednis [5], Hệ thống phát hiện ổ gà của Astarita Vittorio [6] và rất nhiều hệ thống phát hiện ổ gà khác.... Việc phát hiện ổ gà dựa trên nhiều thông tin thu thập được như âm thanh, hình ảnh, video, laser 3D hoặc rung. Cảm biến được sử dụng để xác định ổ gà có thể là phần cứng chuyên dụng gắn trên xe hoặc cảm biến trên điện thoại thông minh.

Cùng với sự phát triển của điện thoại thông minh, Internet với chi phí thấp và nhiều nghiên cứu về phát hiện ổ gà dựa trên gia tốc kế và GPS của điện thoại thông minh thì việc phát hiện bất thường có thể được xử lý trên điện thoại thông minh hoặc tại một máy trung tâm hoặc cả hai. Nếu xử lý được thực hiện trong một trung tâm dữ liệu thì một lượng lớn dữ liệu có xử lý được thực hiện trong điện thoại thông minh và trung tâm dữ liệu, trong khi nếu quá trình xử lý được thực hiện trong điện thoại thông minh thì chương trình phát hiện phải được thiết kế để nó không tiêu thụ nhiều tài nguyên như CPU, bộ nhớ và năng lượng.

2 Một số nghiên cứu liên quan

Một số phương pháp dựa trên ngưỡng đã được đề xuất cho việc phát hiện ổ gà. Đối với hệ thống phát hiện ổ gà của Jakob Eriksson và cộng sự [1] đề xuất một hệ thống được gọi là hệ thống phát hiện ổ gà P². Hệ thống này thu thập thông tin từ máy đo gia tốc và cảm biến GPS được lắp trong xe taxi dưới hình thức:

<thời gian, vị trí, tốc độ, gia tốc 3 trục>

Việc phát hiện ổ gà được xử lý trên một thiết bị được nhúng vào máy vi tính. Kết quả được gửi đến máy chủ như là trung tâm điều khiển thông qua mạng Wifi. Thuật toán phát hiện ổ gà P² bao gồm 5 bộ lọc. Dữ liệu nhận được từ cảm biến là dữ liệu được phân đoạn thành cửa sổ gồm

256 mẫu. Sau đó mỗi cửa sổ đi qua năm bộ lọc và được kiểm tra bởi một bộ lọc để từ chối các cửa sổ không phải là ổ gà: tốc độ, tốc độ vượt quá cao, đỉnh z , tỷ lệ xz và tốc độ so với tỷ lệ z . Tuy nhiên, đối với bộ lọc Xz , chương trình tìm kiếm đầu tiên với cực đại trên trục Z của gia tốc và sau đó đối với mỗi đỉnh của gia tốc nó tìm kiếm một đỉnh trên trục X của gia tốc trong $4w$. Vì kích thước $4w$ chỉ bao gồm 32 mẫu nên độ phức tạp của bộ lọc đỉnh gia tốc Z là $O(k)$. Do đó, độ phức tạp của thuật toán trở thành $O(n)$.

Phương pháp được đề xuất bởi Artis Mednis và cộng sự [5] đề xuất bốn thuật toán cho một kiến trúc phát hiện bất thường trên hệ thống đường. Các thuật toán này sử dụng trục 3 dữ liệu gia tốc để thực hiện phát hiện: thuật toán Z - Thresh phát hiện đường bất thường khi biên độ của thành phần Z của gia tốc vượt quá một ngưỡng T cụ thể, thuật toán Z - Diff phát hiện đường bất thường khi sự khác biệt giữa hai giá trị liên tiếp lớn hơn một giá trị cụ thể, thuật toán $STDdev$ phát hiện đường bất thường khi độ lệch chuẩn lớn hơn một ngưỡng T cụ thể, thuật toán G - Zero phát hiện đường bất thường khi các giá trị của cả ba trục đều nhỏ hơn một ngưỡng T cụ thể. Đối với thuật toán Z - Thresh, Z - Diff, G - Zero thì độ phức tạp của thuật toán là $O(n)$. Đối với thuật toán $STDev$, độ phức tạp của thuật toán là $O(n \times k)$ trong đó k là kích thước cửa sổ.

Phương pháp được đề xuất bởi Astarita Vittorio và cộng sự [6] áp dụng phát hiện đường bất thường trên điện thoại thông minh sử dụng gia tốc của cả tín hiệu gia tốc và tín hiệu GPS. Kể từ khi dữ liệu GPS tần số là 1 Hz và tần số của gia tốc ít nhất là 5 Hz, tác giả xử lý trước dữ liệu gia tốc kể bằng cách tính toán a_{z_min} , a_{z_max} và a_{z_avg} . Sự phát hiện dựa trên xung gia tốc thẳng đứng (DVA) được xác định bởi $DVA = a_{z_max} - a_{z_min}$. Thuật toán được đề xuất bởi Astarita Vittorio giống như thuật toán Z - Thresh, độ phức tạp của thuật toán là $O(n)$.

Phần trình bày ở trên đã phân tích một số phương pháp được phát triển để mục đích kiểm tra mức độ phù hợp của các phương pháp khi được cài đặt trên điện thoại thông minh. So sánh được thực hiện về yêu cầu nguồn lực, khả năng thích ứng, các điều kiện đường, các loại xe và tốc độ khác nhau.

Bảng 1. Tóm tắt các phương pháp sử dụng các tính năng phù hợp với điện thoại thông minh

Phương pháp dựa trên ngưỡng	Độ phức tạp	Bộ nhớ	Điều kiện	Định hướng
Pothle Patrol	Thấp	Thấp	Thấp	Thấp
Z - Thresh, Z - Diff, $Stdev$	Thấp	Thấp	Thấp	Tốt
G - Zero	Thấp	Thấp	Thấp	Vừa
Method of Astarita Vittorio and al.	Thấp	Thấp	Thấp	Tốt

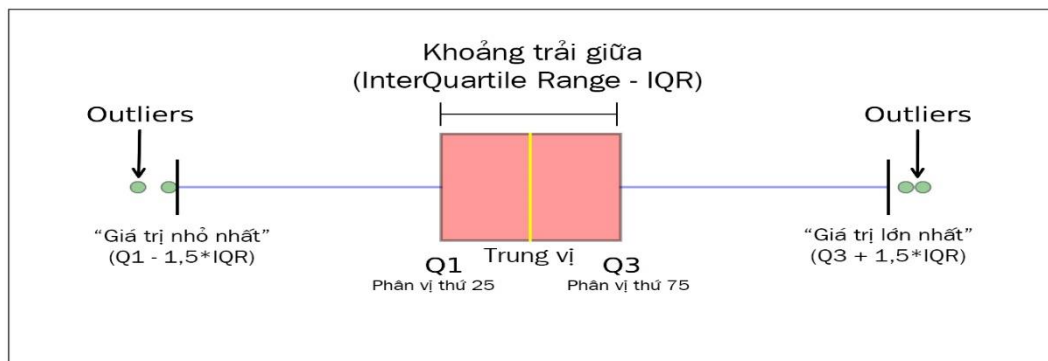
3 Phương pháp đề xuất

Một trong những thách thức lớn đối với việc phát hiện sự bất thường trên hệ thống đường là nhịp điệu dựa trên dao động, biên độ và dao động tần suất cả hai phụ thuộc vào điều kiện đường xá, tốc độ, loại xe, chất lượng hệ thống treo... Trong hầu hết các nghiên cứu thực tế, các tác giả thực hiện các thử nghiệm trên một số đoạn đường với loại xe chuyển động với tốc độ xác định. Do đó, các thông số thu được sau những thí nghiệm không phải là thông số chung đủ để áp dụng cho nhiều trường hợp khác trong thực tế đời sống.

Đối với trường hợp cụ thể là phương pháp dựa trên ngưỡng, khả năng thích ứng với điều kiện thực tế rất thấp vì ngưỡng T là cố định. Để cải thiện đáng kể phương pháp này, tôi đã triển khai bằng cách sử dụng một ngưỡng động. Việc xác định giá trị ngưỡng T có thể thu được bằng cách sử dụng một số hoạt động nhằm xác định các giá trị ngoại lai. Dữ liệu gia tốc được chia theo từng đoạn, một đoạn là một mẫu. Sau đó tôi sử dụng thử nghiệm đối với phương pháp Q - Test [8] và phương pháp IQR [3] để phát hiện ổ gà.

Thử nghiệm đối với phương pháp Q - Test [8] là một thử nghiệm thống kê của Dixon hay đơn giản là bài kiểm tra Q được sử dụng để xác định và loại bỏ các giá trị ngoại lai. Điều này giả định phân phối chuẩn và theo Robert Dean, Wilfrid Dixon và cộng sự, thử nghiệm này nên được sử dụng một cách tiết kiệm và không bao giờ nhiều hơn một lần trong một tập dữ liệu. Một bài kiểm tra Q đơn giản được hiểu như sau: Tính khoảng cách của một quan sát đáng ngờ từ điểm gần nhất là hàng xóm, sau đó chia khoảng cách này cho phạm vi của nó. Nếu Q vượt quá các giá trị được lập trong bảng Q_{crit} của Dixon Table thì giá trị đó được tính là giá trị ngoại lai.

Thử nghiệm đối với phương pháp IQR [3] hay còn gọi là phương pháp tứ phân vị. Tứ phân vị có 3 giá trị, đó là tứ phân vị thứ nhất Q_1 , thứ hai Q_2 hay còn gọi là trung bình, và thứ ba Q_3 . Ba giá trị này chia một tập hợp dữ liệu (đã sắp xếp dữ liệu theo trật tự từ bé đến lớn) thành 4 phần có số lượng quan sát đều nhau. Tứ phân vị được xác định như trong Hình 1. Giá trị IQR là sự khác biệt giữa tứ phân vị thứ nhất Q_1 và tứ phân vị thứ ba Q_3 : $IQR = Q_3 - Q_1$ (1). Giá trị IQR có thể sử dụng để xác định các giá trị ngoại lai bằng cách thiết lập giá trị biên như sau: Nếu chúng ta trừ đi $k \times IQR$ từ tứ phân vị thứ nhất Q_1 , bất kỳ giá trị dữ liệu nào nhỏ hơn con số này được coi là giá trị ngoại lai. Tương tự như vậy, nếu chúng ta thêm $k \times IQR$ đến tứ phân vị thứ ba Q_3 , bất kỳ giá trị dữ liệu nào lớn hơn con số này được coi là giá trị ngoại lai. Giá trị K thường được chọn là 1,5.



Hình 1. Phương pháp IQR phát hiện các giá trị ngoại lai

4 Thực nghiệm và phân tích kết quả

A. Thu thập dữ liệu

Một trong những thách thức lớn đối với việc phát hiện các điểm bất thường trên đường là nhịp điệu dựa trên dao động biên độ và dao động tần suất cả hai phụ thuộc vào điều kiện đường xá, tốc độ, loại xe, chất lượng hệ thống treo... Trong hầu hết các nghiên cứu thực tế, các tác giả thực hiện các thử nghiệm trên một số đoạn đường với loại xe chuyển động với tốc độ xác định. Do đó, các thông số thu được sau những thí nghiệm này không phải là thông số chung đủ để áp dụng cho nhiều trường hợp khác nhau trong thực tế.

Để cải thiện đáng kể khả năng thích ứng thấp theo một ngưỡng T cố định, tôi đã triển khai bằng cách sử dụng hệ thống sử dụng một ngưỡng động. Việc xác định giá trị ngưỡng có thể thu được bằng cách sử dụng một số thống kê hoạt động nhằm xác định các giá trị ngoại lai. Ví dụ, một số thành phần của dữ liệu tăng tốc xuất hiện bất thường khi xe đi vào đường bất thường.

Để thu thập dữ liệu cho các phương pháp phát hiện các giá trị ngoại lai, tôi đã phát triển một phần mềm dành cho điện thoại thông minh để đọc dữ liệu từ hai cảm biến đó là gia tốc kế và GPS. Phần mềm đã được cài đặt trên điện thoại thông minh Samsung A02 sử dụng hệ điều hành Android được cố định ở bên phải định hướng trên các phương tiện giao thông được thể hiện trong Hình 2 và Hình 3. Ứng dụng liên tục lưu ba loại dữ liệu khác nhau: gia tốc, tốc độ và vị trí. Cả tốc độ và vị trí đều được cung cấp bởi cảm biến GPS. Các thông tin thu thập được cung cấp dưới dạng bộ ba chuỗi:

<thời gian, gia tốc 3 trục>

<thời gian, địa điểm>

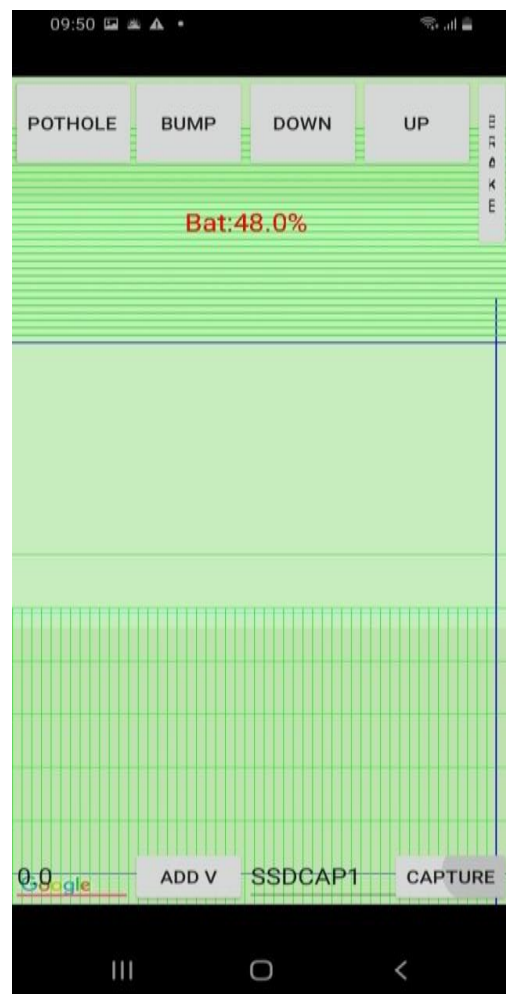
<thời gian, tốc độ>

Dữ liệu thu thập từ một đoạn đường khoảng 40km trong thành phố Huế. Tôi lái xe tại tốc độ từ 10 km/h đến 50 km/h. Tôi sử dụng xe ô tô mang nhãn hiệu Mazda và xe máy mang nhãn hiệu Lead.

Hơn nữa để có thể đánh giá hiệu quả của các thuật toán, thông tin bổ sung cần thiết để thiết lập sự thật nền tảng, tức là đánh dấu vị trí thật trên một con đường bất thường để so sánh, một nút đã được thêm vào công cụ thu thập dữ liệu nhằm cho mục đích này. Khi xe đi vào ổ gà, người điều khiển bấm nút để đánh dấu đó là vị trí ổ gà thật sự. Vị trí bất thường được đánh dấu bằng dấu thời gian thực và được lưu cùng với các thông tin khác. Trong thời gian thực nghiệm trên đoạn đường có 217 điểm là ổ gà thật của đường đã được đánh dấu đối với xe ô tô và 219 điểm là ổ gà thật của đường đã được đánh dấu đối với xe máy.



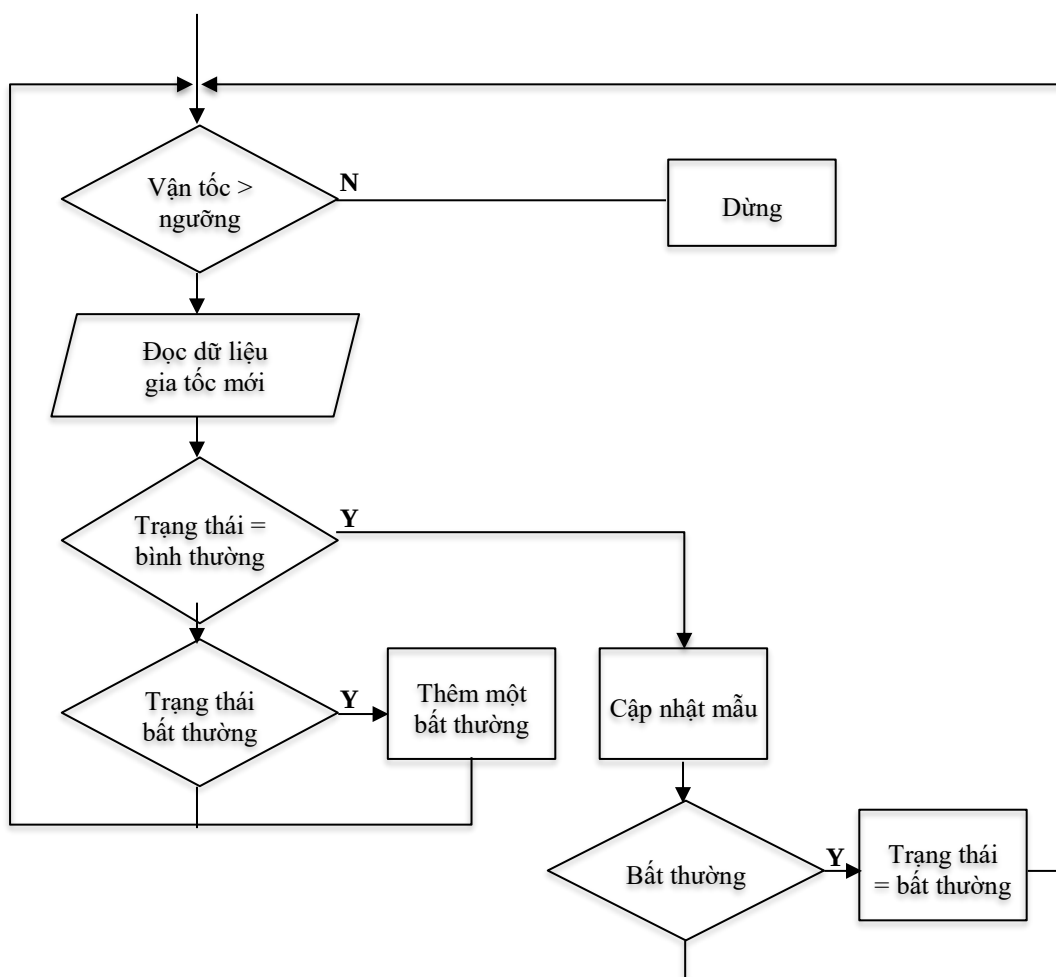
Hình 2. Giao diện phần mềm thu thập dữ liệu



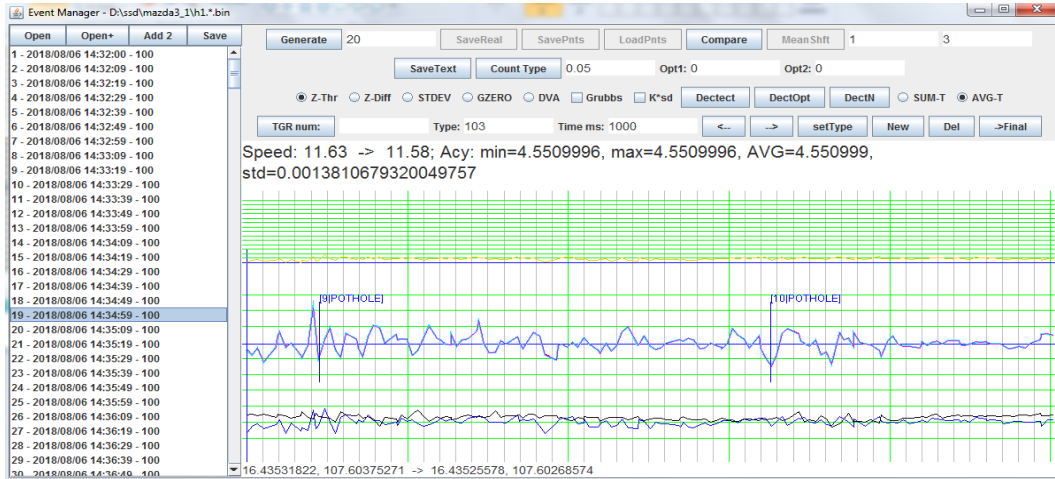
Hình 3. Giao diện phần mềm thu thập dữ liệu

Trong thử nghiệm này, chỉ có dữ liệu về gia tốc và tốc độ được sử dụng. Dữ liệu tăng tốc đã được xử lý liên tục trong chuỗi dữ liệu, để mô phỏng một phát hiện bất thường trong thời gian thực. Đầu tiên, một điểm bất thường đã được phát hiện, bước tiếp theo là tách phân đoạn dữ liệu tăng tốc tương ứng được liên kết với ổ gà trong Hình 4. Nếu phân đoạn dữ liệu quá dài, hệ thống sẽ cắt đến độ dài tối đa là 2 giây cộng với một thời gian di chuyển 10 mét. Các giá trị bổ sung này đảm bảo rằng dữ liệu không được cắt quá ngắn so với ổ gà thật.

Một phần mềm được lập trình bằng ngôn ngữ Java để xử lý thông tin được thu thập từ điện thoại thông minh để chuyển lên máy chủ trung tâm được thể hiện trong Hình 5.



Hình 4. Quá trình xử lý gia tốc



Hình 5. Phần mềm chạy trên máy trung tâm

B. Cơ sở đánh giá

Trong thử nghiệm này, chỉ có dữ liệu gia tốc và tốc độ được sử dụng. Dữ liệu tăng tốc đã được xử lý liên tục trong chuỗi dữ liệu, để mô phỏng phát hiện bất thường trong thời gian thực. Một lần là một điểm bất thường đã được phát hiện, bước tiếp theo là tách phân đoạn dữ liệu tăng tốc tương ứng được liên kết với ổ gà (Xem Hình 4). Nếu phân đoạn dữ liệu quá dài, thì phân đoạn sẽ được cắt đến độ dài tối đa là 2 giây cộng với một thời gian di chuyển 10m. Các giá trị bổ sung này đảm bảo rằng dữ liệu không được cắt quá ngắn so với ổ gà thật.

Việc so sánh giữa các bất thường được phát hiện và sự thật trên mặt đất dựa trên dấu thời gian nhưng không phải trên vị trí dữ liệu. Khi phát hiện bất thường, phân đoạn dữ liệu tương ứng được so sánh với dữ liệu sự thật trên mặt đất. Nếu có bất thường ở đó, đây là True Positive (TP), nếu không thì đó là False Positive (FP). Một sự bất thường không được phát hiện trong sự thật thì đó là một False Negative (FN).

Lưu ý rằng đầu vào bao gồm một chuỗi liên tục dữ liệu và sự bất thường của đường chằng hạn như ổ gà. Do đó, số lượng đường “bình thường” là không đếm được còn được gọi là True Negative không thể xác định được.

Hai phương pháp phát hiện ngoại lai đã được áp dụng trên một quy trình phát hiện bất thường. Chương trình thử nghiệm đã ghi lại các giá trị TP, FN và FP để đánh giá. Tuy nhiên, vì không có True Negative (TN) nên tôi quyết định chỉ sử dụng độ đo Precision (P), độ đo Recall (R) và thước đo F₁ để đánh giá và so sánh hai phương pháp phát hiện giá trị ngoại lai.

Từ [9], Precision (P) và Recall (R) được tính như sau:

$$P = \frac{TP}{TP+FP} \quad (2) \quad \text{và} \quad R = \frac{TP}{TP+FN} \quad (3)$$

Điểm F với tham số β được định nghĩa như sau [9]:

$$F_{\beta} = (1 + \beta) \frac{P.R}{P+R} \quad (4)$$

Trong thực tế, giá trị của β được chọn tùy thuộc vào tầm quan trọng tương đối của cả P hoặc R. Khi β là bằng 1, hai đại lượng quan trọng như nhau và điểm F trở thành điểm F_1 . Điểm F_1 còn được gọi là số đo F là trung bình hài hòa của cả P và R.

$$F_1 = 2 \cdot \frac{P.R}{P+R} \quad (5)$$

Trong trường hợp thí nghiệm cụ thể của tôi, tôi đã áp dụng F_1 để đánh giá độ chính xác của các thí nghiệm. Tuy nhiên, một biện pháp khác đáng lưu ý và tính toán trong phạm vi của phân tích này. Khu vực dưới P và R còn được gọi là độ chính xác trung bình AP [9] rất hữu ích để xác định hiệu quả của một thuật toán:

$$AP = \int_0^1 P(R) dR \quad (6)$$

Nếu các ngưỡng T là các giá trị rời rạc thì độ chính xác trung bình có thể được rút gọn thành một tổng đơn giản. Gọi R_i (tương ứng P_i) là phần thu hồi của ngưỡng T thứ i trong số N lần liên tiếp giá trị ngưỡng AP trở thành:

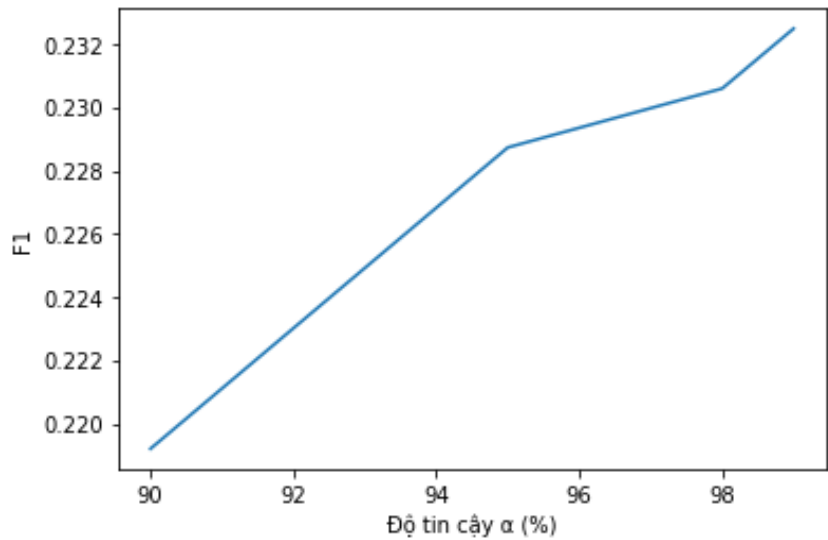
$$AP = \sum_{i=1}^N (R_i - R_{i-1}) \cdot P_i \quad (7)$$

C. Kết quả và phân tích

Tôi đã chạy thử nghiệm phương pháp Q - Test với Q_{crit} 90%, Q_{crit} 95%, Q_{crit} 98% và Q_{crit} 99 % cho bài toán phát hiện ổ gà trên hai phương tiện là xe máy mang nhãn hiệu Lead và xe ô tô mang nhãn hiệu Mazda. Kết quả phương pháp Q - Test được thể hiện trong Bảng 1 với xe máy mang nhãn hiệu Lead và Bảng 2 thể hiện kết quả xe ô tô mang nhãn hiệu Mazda.

Bảng 2. Bảng giá trị của các độ tin cậy α trên phương pháp Q - Test sử dụng cho xe ô tô mang nhãn hiệu Mazda

Độ tin cậy Q	TP	FP	FN	P	R	F_1
90%	117	704	110	0.139334	0.513636	0.219205
95%	113	655	104	0.147135	0.520737	0.229442
98%	113	647	104	0.148684	0.520737	0.23132
99%	113	639	104	0.150266	0.520737	0.23323



Hình 6. Đồ thị F_1 với xe ô tô mang nhãn hiệu Mazda

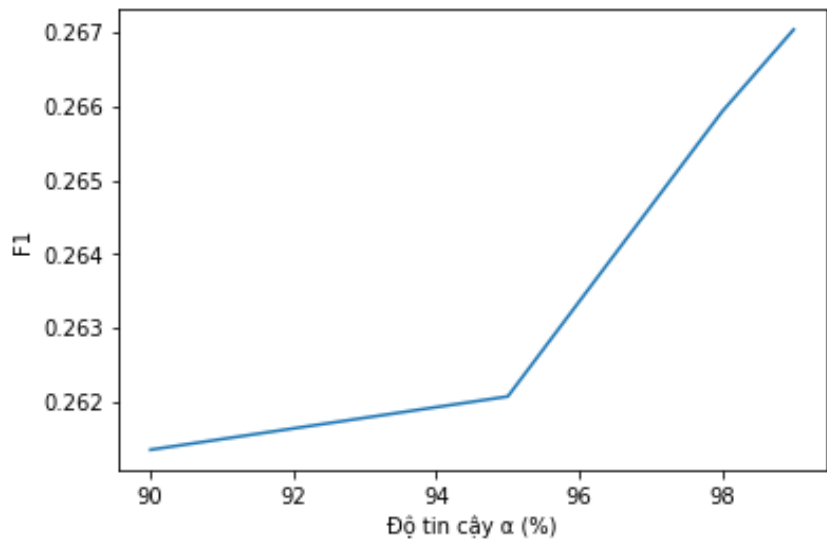
Qua Bảng 2 và biểu đồ trong Hình 6 cho thấy rằng việc sử dụng phương pháp Q - Test để phát hiện ổ gà có phép đo F_1 đạt cực đại ở các giá trị α (độ tin cậy) bằng 99% đối với mọi dữ liệu điều cho kết quả tốt. Tỷ lệ phát hiện ổ gà True Positive đối với các độ tin cậy 90%, 95%, 98%, 99% lần lượt là 53,92 %, 52,07 %, 52,07 %, 52,07 %.

Bảng 3. Bảng giá trị của các độ tin cậy α trên phương pháp Q - Test sử dụng cho xe máy mang nhãn hiệu Lead

Độ tin cậy Q	TP	FP	FN	P	R	F ₁
90%	95	411	124	0.187747	0.43379	0.262069
95%	95	412	124	0.187377	0.43379	0.263158
98%	96	408	123	0.190476	0.434389	0.264828
99%	96	405	123	0.191617	0.434389	0.265928

Qua Bảng 3 và biểu đồ trong Hình 7 cho thấy rằng việc sử dụng phương pháp Q - Test để phát hiện ổ gà có phép đo F_1 đạt cực đại ở các giá trị α (độ tin cậy) bằng 99% đối với mọi dữ liệu điều cho kết quả tốt. Tỷ lệ phát hiện ổ gà True Positive đối với các độ tin cậy 90%, 95%, 98%, 99% lần lượt là 43,38 %, 43,38 %, 43,84 %, 43,84 %.

Tôi đã chạy thử nghiệm phương pháp IQR cho bài toán phát hiện ổ gà trên hai phương tiện là xe máy nhãn hiệu Lead và xe ô tô nhãn hiệu Mazda. Kết quả phương pháp IQR được thể hiện trong Bảng 4 với xe máy mang nhãn hiệu Lead và Bảng 5 với xe ô tô mang nhãn hiệu Mazda.



Hình 7. Đồ thị F_1 với xe máy mang nhãn hiệu Lead

Bảng 4. Bảng giá trị phát hiện ổ gà dựa trên phương pháp IQR đối với xe ô tô mang nhãn hiệu Mazda

Phương pháp IQR	TP	FP	FN	P	R	F_1
	198	3598	19	0.05216	0.912442	0.098679

Bảng 5. Bảng giá trị phát hiện ổ gà dựa trên phương pháp IQR đối với xe máy mang nhãn hiệu Lead

Phương pháp IQR	TP	FP	FN	P	R	F_1
	193	2172	26	0.081607	0.88127	0.149381

Qua Bảng 4 và Bảng 5 đã cho thấy rằng phương pháp IQR áp dụng để phát hiện ổ gà trên hai phương tiện là xe máy mang nhãn hiệu Lead và xe ô tô mang nhãn hiệu Mazda đã cho kết quả tốt, tỷ lệ phát hiện ổ gà là True Positive tương đối cao đối với xe máy là 88,13% và đối với ô tô là 91,24%. Tuy nhiên phương pháp này phát hiện các điểm ổ gà là False Positive tương đối nhiều dẫn đến điểm đánh giá F_1 cho giá trị khá thấp.

5 Kết luận

Mục tiêu của bài viết này là xác định ổ gà trong thời gian thực bằng các phương pháp phát hiện các giá trị ngoại lai có thể thích ứng với nhiều điều kiện trong cuộc sống và đáp ứng yêu cầu càng sử dụng ít tài nguyên càng tốt trên điện thoại thông minh. Tôi đề xuất áp dụng phương pháp phát hiện ngoại lai bằng bài kiểm tra Q của Dixon và phương pháp IQR để xác định vị trí

đường bất thường bằng cách tìm ra các giá trị ngoại lai trong tập dữ liệu cảm biến. Bằng cách này, ổ gà được coi là bất thường ở khu vực xung quanh. Do đó, các phương pháp này hứa hẹn sẽ thích ứng tốt với các điều kiện thực tế. Việc xác nhận thử nghiệm được tiến hành dựa trên dữ liệu thu thập được từ cả xe ô tô mang nhãn hiệu Mazda và xe máy mang nhãn hiệu Lead trên hệ thống đường thành phố Huế, Việt Nam. Kết quả được phân tích dựa trên độ đo Percision - Recall và chỉ số đo F₁ cho thấy tầm quan trọng của các phương pháp.

Tài liệu tham khảo

1. Eriksson, J., Girod, L., Hull, B., Newton, R., Madden, S., Balakrishnan (2008), The pothole patrol: Using a mobile sensor network for road surface monitoring. In: Proceedings of the 6th international conference on Mobile system, applications, and service. pp. 29-39. ACM.
2. Charu C. Aggarwal (2016), *Outlier Analysis*, New York.
3. Jiawei Han, Micheline Kamber, Jian Pei (2012), "Data mining: Concepts and Techniques".
4. Mohan, N. Padmanabhan, R. Ramjee (2008), Nerice: Rich monitoring of road and traffic conditions using mobile smartphone. In: Proceedings of the 6th ACM conference On Embedded network sensor systems. pp. 323 - 336. ACM.
5. Mednis, A., Strazdins, G., Zviedris, R., Kanonirs, G., Selavo, L (2001), Real time pothole detection using android smartphone with accelerometers. In: Distributed Computing in Sensor Systems and Workshops (DCOSS), International Conference on. pp. 1-6 IEEE.
6. Vittorio, A., Rosolino, V., Teresa, I., Vittoria, C.M., Vincenzo, P.G (2014), et al.: Automated sensing system for monitoring of road surface quality by mobile devices. *Procedia-Social and Behavioral Sciences* 111, 242-251.
7. R. B. Dean and W. J. Dixon (1951), Simplified Statistics for Small Numbers of Observations, *Anal. Chem*, 23(4), 636 - 638.
8. R. B. Dean and W. J. Dixon (1951), "Dixon's Q Test", https://en.wikipedia.org/wiki/Dixon%27s_Q_test
9. M. Zhu (2004), "Recall, precision and average precision, department of statistics and actuarial science, university of waterloo", Waterloo Working paper, Tech. Rep.
10. Mohan, N Padmanabhan, R. Ramjee (2008), Trafficsense: Rich monitoring of road and traffic conditions using mobile smartphones, Microsoft Research, Tech. Rep. MSR-TR-2008-59, April.
11. Perttunen, M., Mazhelis, O., Cong, F., Kauppila, M., Leppanen, T., Kantola, J., Collin, J., Pirttikangas, S., Haverinen, J., Ristaniemi, T (2011), et al: Distributed road surface condition monitoring using mobile phones. In: International Conference on Ubiquitous Intelligence and Computing. pp. 64-78. Springer.
12. V. Vanpanik, Boser, Guyon (1982), "Support vector machine", http://en.wikipedia.org/wiki/Support_vector_machine.
13. V. Barnett, T. Lewis (1994). *Outliers in Statistical Data*, John Wiley, 3rd edition.
14. John M. Cimballa (2011), *Outliers*, Penn State University, Latest Revision.