

Nghiên cứu ứng dụng thuật toán ước tính mật độ nhân trong xác định "điểm đen" tai nạn giao thông tại Việt Nam

■ TS. TRẦN QUANG HỌC - Trường Đại học Giao thông vận tải
 ■ ThS. NCS. LÊ KHÁNH GIANG - Feng Chia University, Taiwan

TÓM TẮT: Bài báo nghiên cứu xác định "điểm đen" TNGT bằng thuật toán ước tính mật độ nhân kết hợp với phần mềm mã nguồn mở QGIS. Kết quả nghiên cứu với phân tích dữ liệu tại khu vực Hà Nội cho phép xác định chính xác vị trí các "điểm đen", phân loại mức độ nghiêm trọng các "điểm đen" và thể hiện trực quan trên bản đồ. Điều đó giúp các nhà quản lý có các giải pháp đảm bảo an toàn cho người và các phương tiện tham gia giao thông

TỪ KHÓA: Tai nạn giao thông, QGIS, ước tính mật độ nhân, "điểm đen", mức độ tai nạn

ABSTRACT: The article's purpose aims to identify traffic accident black spots by using kernel density estimation algorithm combined with open-source software - quantum geographic information system QGIS. Research results with data analysis in Hanoi area enable us to accurately identify the locations of black spots, classify the severity of black spots, and visualize them on a map. That helps traffic authorities have solutions to ensure safety for traffic users.

KEYWORDS: Road traffic accident (RTA), quantum geographic information system (QGIS), kernel density estimation (KDE), black spots, accident severity

1. ĐẶT VẤN ĐỀ

Hệ thống giao thông đường bộ được coi là hệ thống phức tạp và nguy hiểm nhất mà con người phải đối mặt hàng ngày. Việt Nam là một quốc gia đang phát triển, do đó vấn đề TNGT cũng là một trong những mối quan tâm nhất của xã hội. Theo thống kê năm 2017 đã có 20.000 vụ TNGT, khoảng 8.200 người chết và 17.000 người bị thương trên mạng lưới các tuyến đường tại Việt Nam [1].

TNGT đường bộ và hệ thống cơ sở hạ tầng giao

thông (HTGT) có liên quan chặt chẽ với nhau. Một trong những nguyên nhân chính của TNGT là chất lượng HTGT thấp được phản ánh bởi một số yếu tố như đường xuống cấp, thiết kế hình học đường không phù hợp và thiếu các đèn tín hiệu, vạch sơn, dải phân cách, biển chỉ dẫn, cảnh báo...

"Điểm đen" TNGT là nơi mà tai nạn thường xảy ra TNGT [2]. Để xác định các vị trí "điểm đen", hiện nay thường sử dụng mô hình phi không gian, cụ thể là: phương pháp tần suất tai nạn (phân loại theo mức độ tổn thương) trong khoảng thời gian một năm [3]. Đây là phương pháp lâu đời nhất và đơn giản nhất để xác định các vị trí nguy hiểm. Tuy nhiên, phương pháp này có nhiều hạn chế như thiếu trực quan, kết nối giữa không gian và thời gian, xếp hạng mức độ ưu tiên của "điểm đen", không xem xét lưu lượng giao thông, không có mối quan hệ trực tiếp với tần suất sự cố. Do đó, kết quả đạt được chưa thực sự tin cậy và không trực quan [4]. Vì vậy, trong phạm vi bài báo, nhóm tác giả nghiên cứu ứng dụng thuật toán ước tính mật độ nhân KDE kết hợp phần mềm mã nguồn mở QGIS trong xác định "điểm đen" TNGT tại Việt Nam.

2. NỘI DUNG

2.1. Khu vực và dữ liệu nghiên cứu

Nghiên cứu được thực hiện tại khu vực Hà Nội. Bộ dữ liệu được sử dụng cho nghiên cứu này bao gồm:

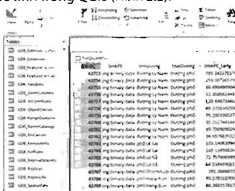
- Bản đồ mạng lưới các tuyến đường ở Hà Nội có định dạng shapefile được cung cấp bởi Sở Tài nguyên và Môi trường tại Hà Nội, bao gồm các thông số kỹ thuật của các con đường như chiều dài đường, chiều rộng đường, loại đường, tốc độ giới hạn...
- Dữ liệu TNGT trong 3 năm từ 2015 đến 2017 tại Hà Nội, được cung cấp bởi CSGT tại Hà Nội. Tập dữ liệu này chứa các thông tin quan trọng về các vụ tai nạn như ngày và thời gian xảy ra tai nạn, vị trí và phạm vi, loại sự cố, loại phương tiện, tuổi và giới tính của người bị thương...

2.2. Thiết kế mô hình hệ thống quản lý dữ liệu tai nạn

Hệ thống thông tin địa lý (GIS) là công nghệ cho quản lý, xử lý các thông tin về vị trí và các thông tin thuộc tính liên quan. Thêm vào đó, GIS còn có khả năng phân tích kết hợp giữa dữ liệu không gian và dữ liệu thuộc

tính. GIS cho phép hiển thị trực quan kết quả phân tích, do đó cho phép thực hiện các phân tích phức tạp và ra quyết định nhanh chóng [5,6,7]. Phát triển một hệ thống sử dụng GIS để phân tích TNGT nhằm nâng cao hiệu quả của các biện pháp đối phó với TNGT là điều cần thiết và cấp bách [5,7]. Do đó, phát triển một hệ thống như vậy trong đó các bản đồ số được liên kết với dữ liệu TNGT, HTGT, thời tiết, nhân khẩu, sử dụng đất và các loại dữ liệu khác cũng là mục đích của nghiên cứu.

Việc thiết lập hệ thống đòi hỏi một thiết kế cho phép các cơ sở dữ liệu (CSDL) xây dựng được sử dụng hiệu quả bởi GIS. Để phát triển một hệ thống linh hoạt đạt được mục tiêu này, chúng tôi đã kết hợp QGIS với Access - một ứng dụng quản lý quan hệ quản lý CSDL từ Microsoft như trong Hình 2.1. Do nhiều loại dữ liệu được xử lý, kết nối CSDL mở, một tiêu chuẩn của thế giới, được sử dụng để kết nối QGIS và Access. Có thể thay chính các menu bằng cách sử dụng Visual Basic hoặc Python để nâng cao tính thân thiện với người dùng. Ngoài ra, dữ liệu TNGT từ tệp Excel có thể được nhập vào phần mềm QGIS một cách dễ dàng và thông tin này sẽ trở thành dữ liệu thuộc tính trong QGIS (Hình 2.2).



Hình 2.1: CSDL không gian mạng lưới đường được lưu trong Access



Hình 2.2: Dữ liệu TNGT trong Excel trở thành dữ liệu thuộc tính trong QGIS



Hình 2.3: Sơ đồ mô hình CSDL của hệ thống phân tích TNGT dựa trên GIS

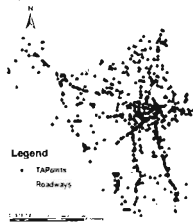
Hình 2.3 cho thấy hệ thống CSDL bao gồm: bản đồ các tuyến đường, dữ liệu TNGT, dữ liệu giao thông, dữ liệu thời tiết... Dữ liệu đường được trích xuất từ hệ thống dữ liệu quản lý đường đang hoạt động. Các mục dữ liệu bao gồm thông tin tuyến đường (chiều dài, chiều rộng, làn đường, tốc độ...) và các tín hiệu giao thông như đèn đường, biển báo... Dữ liệu về nhân khẩu học và đất sử dụng hai bên đường cũng như vị trí các trường học, bệnh viện, trung tâm thương mại... Dữ liệu thời tiết là từ cơ quan khí tượng của từng khu vực và chúng bao gồm nhiệt độ, lượng mưa và giờ nắng...

2.3. Phương pháp luận

Phương pháp thực hiện được tiến hành theo các bước sau:

2.3.1. Mã hóa dữ liệu địa lý tại nạn và tổng hợp dữ liệu tại nạn tại mỗi địa điểm

Bản đồ mạng lưới đường và dữ liệu tại nạn được mã hóa địa lý và được tải lên trong phần mềm QGIS (Hình 2.4). Mỗi điểm dữ liệu có thể đại diện cho một vụ va chạm hoặc nhiều vụ va chạm vì một số va chạm có thể được báo cáo tại cùng một vị trí nhưng giao lộ. Do đó, dữ liệu TNGT được tổng hợp để tạo ra số lượng cho từng loại va chạm tại mỗi vị trí.



Hình 2.4: Khu vực nghiên cứu với sự phân bố của tất cả các vụ va chạm tại Hà Nội (2015 - 2017)

2.3.2. Thuật toán ước tính mật độ nhân (KDE) và bản đồ mật độ (Heat Map)

KDE là một trong những phương pháp hiệu quả nhất để phân tích xác định các mô hình không gian của TNGT. Mật độ của các sự kiện (TNGT) được tính toán trong phạm vi bán kính nghiên cứu đã được xác định trước trong các khu vực nghiên cứu để tạo ra một bề mặt nhẵn mịn. Hàm nhân được sử dụng để gán trọng số cho khu vực xung quanh các sự kiện (TNGT) tỷ lệ thuận với khoảng cách của nó đến sự kiện điểm. Từ đó, giá trị cao nhất là tại trung tâm vị trí sự kiện điểm (TNGT) và giảm dần xuống giá trị 0 tại giá trị biên của bán kính của vòng tròn nghiên cứu. Cuối cùng, một bề mặt mật độ mịn liên tục được tạo ra bằng cách thêm vào các nhân riêng lẻ trong khu vực nghiên cứu [8]. Mật độ tại một vị trí cụ thể được tính bằng công thức (1):

$$f(s) = \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{d_i}{h}\right) \quad (1)$$

Trong đó: $f(s)$ - Mật độ tại vị trí s , n - Số lượng quan sát; h - Băng thông hoặc bán kính nghiên cứu; K - Hàm nhân; d_i - Khoảng cách giữa vị trí s và vị trí của quan sát thứ i .

Bản đồ mật độ (Heat map) trong QGIS 2.18 là một công cụ dựa trên thuật toán KDE được sử dụng để tạo bản đồ mật độ của lớp vector điểm đầu vào. Nó được sử dụng để tính toán mật độ tại nan trong phạm vi một băng thông tìm kiếm cần thiết. Băng thông lớn hơn dẫn đến vùng lớn hơn và mịn hơn trong khi băng thông nhỏ hơn sẽ chỉ ra mức độ chi tiết hơn. Trong nghiên cứu này, băng thông tìm kiếm 1.000m được kiến nghị xem xét để phân tích.

2.3.3. Tính toán chỉ số mức độ nghiêm trọng tại mỗi vị trí tai nạn

Trong nhiều nghiên cứu trước đây, số lượng và phạm vi thường được sử dụng để đánh giá vấn đề an toàn tại một địa điểm. Các nghiên cứu trước đây cũng chỉ ra rằng, nếu không có dữ liệu trong số rất khó để biết liệu phạm vi là cao hay thấp đang tồn tại. Có một sự tin tưởng rằng các vụ va chạm nghiêm trọng hơn sẽ có trọng số lớn hơn trong việc xác định các vị trí nguy hiểm trên cơ sở thiệt hại kinh tế do va chạm gây ra. Vì vậy, để xác định các vị trí không an toàn, các sự cố phải được gán trọng số theo mức độ nghiêm trọng. Ngoài ra, kết quả của phương pháp mức độ nghiêm trọng và phạm vi rất nhạy cảm với các hệ thống trọng số khác nhau [9]. Mặc dù không có sự đồng thuận về cách phát triển hệ thống trọng số tối ưu, một cách tiếp cận thỏa hiệp đang ngày càng trở nên phổ biến. Cách tiếp cận là đưa ra trong số cho các va chạm nghiêm trọng hơn, nhưng không phải với các giá trị cực cao mà được tính theo tỷ lệ trực tiếp với chi phí va chạm. Nghiên cứu này sử dụng một hệ thống trọng số mức độ và phạm vi của Cảnh phủ Bỉ. Theo hệ thống này, các trọng số riêng biệt 5, 3 và 1 được cung cấp cho các tai nạn chết người, nghiêm trọng và nhẹ tương ứng. Chỉ số mức độ nghiêm trọng cho từng vị trí có thể được tính theo công thức (2):

$$B = L + 3S + SD \quad (2)$$

Trong đó:

L - Tổng số thương tích nhẹ; S - Tổng số thương tích nghiêm trọng; D - Tổng số người chết.

Định nghĩa của BI về "điểm đen" này thể hiện tính kinh tế rất cao. Tương tự như mức độ thiệt hại kinh tế đã được xác định. Một người chết tương đương với 5 người bị thương nhẹ và 01 người bị thương nặng, tương đương với 3 người bị thương nhẹ. Đây là một phương pháp được xác định rất khoa học và chứng minh tác động kinh tế của TNGT. Từ đây có thể ước tính mức độ thiệt hại được coi là một "điểm đen" [10]. Để tính toán mức độ nghiêm trọng của tai nạn, trọng số theo hệ thống của BI được chỉ định cho mỗi tai nạn được biểu thị dưới dạng chỉ số nhân dạng (ID). Điều này tạo thuận lợi cho việc đếm các tai nạn theo trọng số được gán.

2.3.4. Phân loại "điểm đen"

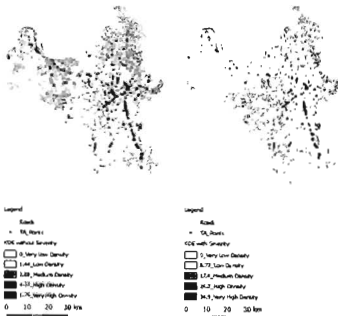
Trong trường hợp phân tích bằng KDE, vì không có chỉ số liên quan đến ý nghĩa thống kê, các "điểm đen"

được phân loại theo các khoảng bằng nhau. Trong nghiên cứu này, việc phân loại được thực hiện theo năm loại bao gồm: mật độ rất thấp, thấp, trung bình, cao và rất cao dựa trên mật độ tại nan liên quan của chúng.

2.4. Kết quả và thảo luận

2.4.1. Bản đồ mật độ trong trường hợp không có và có tính đến chỉ số mức độ tai nạn

Kết quả xác định "điểm đen" TNGT được thực hiện với sự trợ giúp của thuật toán KDE, trong hai trường hợp có và không tính đến mức độ nghiêm trọng của tai nạn được hiển thị trong Hình 2.5 và 2.6. Các vị trí nguy hiểm với mật độ TNGT cao được hiển thị trong khu vực nghiên cứu của Hà Nội (trong vùng đen sẫm).



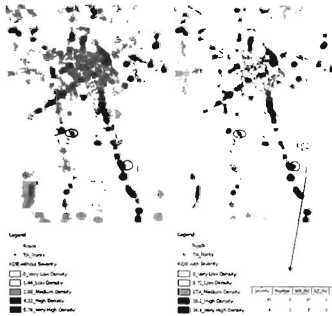
Hình 2.5: Bản đồ các điểm đen tại Hà Nội (2015 - 2017) không tính đến chỉ số mức độ nghiêm trọng

Hình 2.6: Bản đồ các điểm đen tại Hà Nội (2015 - 2017) có tính đến chỉ số mức độ nghiêm trọng

2.4.2. Phân loại "điểm đen" trong trường hợp không có và có tính đến chỉ số mức độ tai nạn

Cả hai phân tích đều xác định các "điểm đen" tương đối giống nhau, nhưng thứ hạng của một số "điểm đen" khá khác nhau do sự tích hợp của chỉ số mức độ nghiêm trọng của tai nạn (Hình 2.7 và 2.8). Do không có chỉ số liên quan đến ý nghĩa thống kê, các "điểm đen" được phân loại theo các khoảng bằng nhau. Trong nghiên cứu này, việc phân loại được thực hiện theo 5 loại bao gồm: mật độ rất thấp, thấp, trung bình, cao và rất cao dựa trên mật độ tại nan liên quan của chúng. Sau khi phân tích xếp hạng, có thể dễ dàng quan sát thấy ảnh hưởng của mức độ nghiêm trọng của tai nạn đối với kết quả. Cụ thể, vị trí 1 (bao quanh bởi vòng tròn màu đen) (Hình 2.7) được xếp hạng với mật độ thấp trong trường hợp không tính đến mức độ nghiêm trọng trong khi nó được xếp hạng với mật độ rất cao trong trường hợp có tính đến mức độ nghiêm trọng (Hình 2.8). Tính hướng tương tự có thể được quan sát ở vị trí 2. Vị trí 2 (Hình 2.7) được xếp hạng mật độ trung bình trong trường hợp không tính đến mức độ

nghiêm trọng trong khi nó được xếp hạng mật độ cao trong trường hợp có tính đến mức độ nghiêm trọng (Hình 2.8). Điều này là do tỷ lệ lớn hơn của các vụ tai nạn nghiêm trọng và chết người ở những địa điểm này, nó cho thấy tác động của chúng đối với kết quả.



Hình 2.7: Bản đồ "điểm đen" trong trường hợp không tính đến chỉ số mức độ

Hình 2.8: Bản đồ "điểm đen" trong trường hợp có tính đến chỉ số mức độ

2.4.3. Bản đồ "điểm đen" với mức độ ưu tiên cao

Với các vị trí của TNGT trong Hình 2.4, không thể tìm ra các "điểm đen" TNGT. Hình 2.8 cho thấy các vùng màu sẫm là các "điểm đen" TNGT. Tuy nhiên, QGIS có một chức năng cho phép phân tách các mức độ ưu tiên khác nhau. Trong trường hợp này, các "điểm đen" TNGT với mật độ cao và rất cao đã được giữ lại trong khi các điểm khác sẽ bị loại bỏ. Điều này cho phép chúng ta dễ dàng xem xét và phân tích các "điểm đen" mật độ cao (Hình 2.9). Hình 2.9 cho thấy các "điểm đen" với mật độ cao (màu đỏ) chủ yếu tập trung ở đoạn QL1A như gần ga Văn Điển, ga Chợ Tia, ngã tư Quang Trung - Nguyễn Trãi, vị trí các đầu cầu Chương Dương, Vĩnh Tuy...



Hình 2.9: Bản đồ "điểm đen" với mức độ ưu tiên cao ở Hà Nội (2015 - 2017)

3. KẾT LUẬN

Bài báo đã nghiên cứu xây dựng mô hình quản lý dữ liệu các "điểm đen" TNGT và thực nghiệm quản lý dữ liệu tại Hà Nội. Từ mô hình quản lý dữ liệu, ứng dụng thuật toán ước tính mật độ nhân KDE kết hợp với phần mềm mã nguồn mở QGIS để phân tích xác định "điểm đen" TNGT.

Kết quả phân tích được phân loại thực hiện theo 5 loại bao gồm: mật độ rất thấp, thấp, trung bình, cao và rất cao dựa trên mật độ tai nạn liên quan giữa các "điểm đen". Khi sử dụng phần mềm QGIS, các kết quả phân tích được hiển thị trực quan trên bản đồ, đồng thời QGIS cho phép phân tách các mức độ ưu tiên khác nhau. Nếu chỉ quan tâm đến các "điểm đen" TNGT với mật độ cao và rất cao thì có thể loại bỏ các "điểm đen" với mật độ rất thấp, thấp hoặc trung bình. Điều này giúp cho các nhà quản lý có giải pháp nhanh chóng để làm giảm thiểu các TNGT cho người và phương tiện khi tham gia giao thông trong trường hợp ngân sách hạn hẹp.

Trong phần phân tích, nhóm tác giả đã sử dụng phần mềm mã nguồn mở QGIS và nhận thấy kết quả phân tích tương đương với kết quả từ phần mềm thương mại ArcGIS. Từ kết quả nghiên cứu này cũng thêm khẳng định phần mềm mã nguồn mở QGIS là lựa chọn thay thế cho phần mềm ArcGIS, phù hợp cho các công tác nghiên cứu khoa học với nguồn kinh phí có hạn.

Tài liệu tham khảo

- [1]. T. Giang, Ủy ban ATGT Quốc gia tổng kết công tác năm 2017, <http://backantv.vn/tim-tuc-n17855/uy-ban-an-toan-giao-thong-quoc-gia-tong-ket-cong-tac-nam-2017.html>, truy cập ngày 20/6/2019.
- [2]. R. Elvik (2008), *A Survey of Operational Definitions of Hazardous Road Locations in Some European Countries*, Accident Analysis & Prevention, 40, 6 1830-1835, <https://doi.org/10.1016/j.aap.2008.08.001>.
- [3]. Thông tư số 26/2012/TT-BGTVT Quy định về việc xác định và xử lý vị trí nguy hiểm trên đường bộ đang khai thác, Bộ GTVT, Việt Nam, 2012.
- [4]. L. Li (2006), *A GIS-based Bayesian approach for analyzing spatial-temporal patterns of traffic crashes*, Doctoral dissertation, Texas A&M University.
- [5]. M. A. Dereli, S. Erdogan, *A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods*, Transportation Research Part A: Policy and Practice, 103 (2017) 106-117, <https://doi.org/10.1016/j.tra.2017.05.031>.
- [6]. C. D. Lloyd (2010), *Spatial Data Analysis: An Introduction for GIS users*, Oxford, England: Oxford University Press.
- [7]. J. M. Su, Y. M. Wang, C. Chang, P. J. Wu (2019), *Application of a Geographic Information System to Analyze Traffic Accidents Using Nantou County, Taiwan, as an Example*, Journal of the Indian Society of Remote Sensing 47, 1, 101-111, <https://doi.org/10.1007/s12524-018-0874-z>.
- [8]. T. K. Anderson, *Kernel density estimation and*

K-means clustering to profile road accident hotspots, Accident Analysis & Prevention, 41, 3 (2009) 359-364, <https://doi.org/10.1016/j.aap.2008.12.014>.

[9]. K. Geurts, G. Wets, T. Brijs, K. Vanhoof, *Identification and ranking of black spots: Sensitivity analysis*, Transportation Research Record 1897 (2004) 34-42, <https://doi.org/10.3141/1897-05>.

[10]. N. T. Phong, *Nghiên cứu tiêu chí xác định "điểm đen" trên mạng quốc lộ Việt Nam*, <http://www.vusta.vn/vi/news/Thong-tin-Su-kien-Thanh-tuu-KH-CN/Nghien-cuu-tieu-chi-xac-dinh-diem-den-tren-mang-quoc-lo-Viet-Nam-33167.html>, truy cập ngày 22/6/2019.

Ngày nhận bài: 16/02/2020

Ngày chấp nhận đăng: 25/3/2020

Người phản biện: PGS. TS. Nguyễn Việt Hà

TS. Lê Văn Hiến