

ĐỀ XUẤT MÔ HÌNH DỰ BÁO ỚNG VỠ TRÊN MẠNG LƯỚI CẤP NƯỚC

Phạm Thị Minh Lành^{1,3}, Vũ Thị Vân Anh², Phạm Hà Hải³

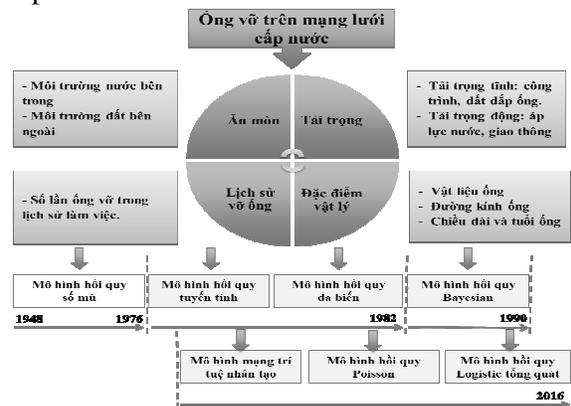
Tóm tắt: Đa số các đường ống cấp nước trong đô thị đều được đặt dưới mặt đường hoặc vỉa hè nên khi xảy ra vỡ một lượng nước đáng kể di chuyển ra khỏi đường ống mà không được phát hiện kịp thời. Đồng thời, khi có điểm vỡ trên đường ống các chất ô nhiễm tồn tại bên ngoài đường ống có điều kiện đi vào bên trong ống dễ hơn làm tăng nguy cơ nhiễm độc hàng loạt cho những người tiêu thụ nước. Để giảm thiểu chi phí do thất thoát nước đồng thời đảm bảo an toàn cho sức khỏe của người sử dụng nước, cần xác định được khả năng ống vỡ trên mạng lưới cấp nước từ đó có kế hoạch bảo dưỡng và thay thế kịp thời. Trong bài báo này sẽ phân tích định lượng các yếu tố liên quan đến sự kiện ống vỡ và đề xuất sử dụng mô hình cây quyết định ước lượng khả năng vỡ ống từ các yếu tố liên quan này. Mô hình đề xuất sẽ được kiểm chứng cho số liệu thống kê thực tế và đánh giá chất lượng bằng chỉ số AUC.

Từ khóa: mạng lưới cấp nước; thất thoát nước; ống cấp nước vỡ; chỉ số AUC; mô hình cây quyết định; phần mềm R.

1. ĐẶT VẤN ĐỀ

Các sự cố vỡ ống đặc biệt là các tuyến ống dịch vụ thường xuyên xảy ra trên mạng lưới cấp nước (MLCN), nguyên nhân có thể do động đất thiên tai, các hoạt động thi công ở khu vực đặt ống, khả năng chịu tải của mặt đường phía trên ống dẫn nước thấp hơn tải trọng xe chạy.... Mặc dù đã có những cải tiến để tăng khả năng chịu lực của vật liệu và các biện pháp bảo vệ nhưng vẫn không tránh khỏi hư hỏng trong những trường hợp này. Bên cạnh những sai sót trên sản phẩm cũng như quy trình thi công thì nguyên nhân dẫn đến vỡ ống chủ yếu là do ăn mòn (Sheikh, Boah and Hansen, 1990) hoặc lực tác động từ môi trường trong và ngoài (Gómez-Martínez *et al.*, 2017), một số đoạn ống thường xuyên xảy ra vỡ thì cần xem xét cả hai nguyên nhân này. Đôi khi chi phí để sửa chữa đoạn ống nhiều lần sẽ cao hơn so với việc lắp đặt ống mới (Pelliccia, 1981). Bên cạnh đó hoạt động sửa chữa cũng làm ảnh hưởng đáng kể đến chất lượng phục vụ của hệ thống và khả năng làm

gián đoạn giao thông của khu vực đặt ống. Bỏ qua đặc điểm hình học của điểm vỡ, những sự cố bất thường làm vỡ ống, tác giả xác định các yếu tố liên quan tới sự kiện ống vỡ từ đó đề xuất mô hình dự báo khả năng ống vỡ trên mạng lưới cấp nước.



Hình 1. Các yếu tố dẫn đến hiện tượng ống vỡ trên MLCN

Trong Hình 1 trình bày các nguyên nhân dẫn đến ống cấp nước bị vỡ là ăn mòn, tải trọng, đặc điểm vật lý và lịch sử vỡ ống. Đánh giá mức độ ảnh hưởng của các yếu tố này tới hiện tượng vỡ ống bằng phương pháp thống kê các tác giả thường sử dụng mô hình số mũ, hồi quy tuyến tính, mô hình Bayesian, Poisson. Mỗi mô hình

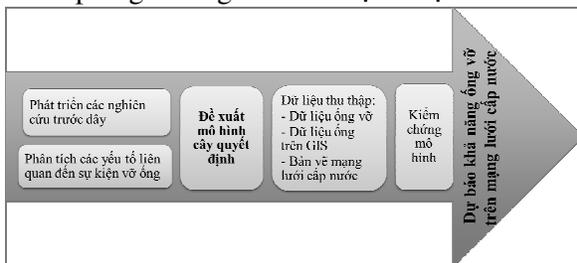
¹Khoa Kỹ thuật Xây dựng, ĐH Bách Khoa Tp. Hồ Chí Minh

²Khoa Khí tượng Thủy văn, ĐH Tài Nguyên và Môi Trường

³Khoa Kỹ thuật Đô thị, ĐH Kiến Trúc Tp. Hồ Chí Minh

có một ưu điểm riêng và phù hợp với số liệu khảo sát.

Nhận định ban đầu cho rằng khả năng ống vỡ phụ thuộc vào thời gian làm việc, tác giả Bubbis (Bubbis, 1948) đã đưa ra mô hình số mũ để mô tả hiện tượng vỡ ống và mô hình này tiếp tục được phát triển bởi Ossman năm 2011 (Bainbridge, 2011), từ lịch sử vỡ ống để ước lượng khả năng vỡ trong tương lai. Tuy nhiên, dữ liệu khảo sát chưa đề cập tới các yếu tố liên quan như môi trường làm việc của ống, điều kiện áp suất, môi trường trong và ngoài ống. Một số nghiên cứu tiếp cận theo phương pháp mạng trí tuệ nhân tạo-Artificial Neural Network ANN để xác định hiệu suất làm việc của hệ thống cũng như tỉ lệ vỡ ống trong thời gian làm việc (Al-barqawi and Zayed, 2008). Một trong những nhược điểm của mô hình ANN là thời gian chạy mô hình lớn và chi phí cho khảo sát số liệu đầu vào khá cao, vậy nên mô hình mới chỉ đáp ứng cho nghiên cứu học thuật.



Hình 2. Nội dung nghiên cứu ước lượng khả năng ống vỡ trên MLCN

Các mô hình hồi quy tuyến tính của tác giả Andreou (1987) đánh giá đạt kết quả chính xác đến 70% (Wengström, 1993), đồng thời tác giả cũng nhận định ống vỡ phụ thuộc vào đường kính, vị trí đặt ống, nguyên nhân gây vỡ và không phụ thuộc vào độ tuổi. Sau khi so sánh hiệu quả các mô hình hồi quy tuyến tính, hồi quy logistic tổng quát, Poisson, số mũ, tác giả Yamijala(2007) đề xuất mô hình số mũ thời gian và Poisson là mô hình dự báo chính xác nguy cơ vỡ, đặc biệt là đếm được số lần không vỡ. Hồi quy Bayesian là mô hình mới được áp dụng trong các nghiên cứu gần đây, với những ưu điểm của mô hình đã giúp bài toán có kết quả chính xác hơn các phương pháp trước.

Phát triển các kết quả nghiên cứu trước đây và dự báo ống vỡ trong điều kiện Việt Nam, nghiên cứu thực hiện theo trình tự như Hình 2. Các số liệu thống kê đầu vào trong mô hình được thu thập từ công ty cấp nước. Từ kết quả phân tích và xử lý số liệu thống kê nghiên cứu đề xuất sử dụng mô hình cây quyết định (DT) để dự báo khả năng ống vỡ.

2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Các yếu tố liên quan tới sự kiện vỡ ống

Trong quá trình hoạt động, đường ống chịu tác động của ăn mòn cũng như tác động của tải trọng đất xung quanh ống và trọng lượng nước mà ống truyền tải. Đại lượng đặc trưng cho tỉ lệ ăn mòn được nghiên cứu lựa chọn là độ tuổi ống (A), cho tải trọng nước là áp suất làm việc trung bình của ống (P), vị trí tuyến đường đặt ống (R). Đại lượng thứ ba để xác định khả năng ống vỡ là đặc trưng vật lí của đường ống bao gồm đường kính (D), chiều dài (L) và vật liệu (Mat). Các nghiên cứu trước đây tập trung chủ yếu vào vật liệu ống gang và thép nói chung, trong nghiên cứu này đề xuất xem xét cụ thể các vật liệu thép trắng (SS), thép đen (ST), gang dẻo (DI), gang xám (CI), nhựa Poly Vinyl Clorua (PVC), nhựa Poly Etilen cao phân tử (HDPE) và xi măng amiăng (AC).. Ngoài ra các đại lượng đặc trưng cho đặc điểm hình học của mạng lưới là vị trí tuyến đường đặt ống (R), số đoạn ống kết nối trên một tuyến ống (N0), tổng chiều dài ống trong một tiêu vùng cấp nước-District Metered Area (DMA) và số lần vỡ trước đây (Prior) cũng được khảo sát để đưa vào mô hình thống kê.

2.2. Giới thiệu phần mềm R và lý thuyết mô hình cây quyết định (DT)

Được phát triển từ năm 1996 bởi hai tác giả Ross Ihaka và Robert Gentleman, phần mềm R có mã nguồn mở sử dụng cho thống kê và biểu diễn biểu đồ. Phần mềm có thể đọc dữ liệu trực tiếp hoặc từ Excel (.csv); SAS; SPSS; Stata; text;...và linh hoạt trong quá trình xử lý số liệu thống kê bằng cách phân loại thành các nhóm, thay thế biến từ kí tự sang số. Với các ưu điểm trên, nghiên cứu lựa chọn ngôn ngữ R để xây

dựng mô hình cây quyết định, từ đó ước lượng khả năng ống vỡ và áp dụng cho dữ liệu khảo sát. Mô hình cây quyết định phát triển từ những năm 80 bởi các tác giả Breiman, Friedman, Olshen, Stone (Leo Breiman; Jerome H. Friedman; Richard A. Olshen; Charles J. Stone, 1984). Mô hình có khả năng xử lý dữ liệu lớn trong thời gian ngắn bằng cách phân loại dữ liệu thống kê thành các lớp từ đó xây dựng nên mô hình cây quyết định- DT. Ưu điểm của DT là có thể loại bỏ các giá trị ngoại vi ra khỏi hệ thống, mỗi giá trị này bị cô lập tại các nút riêng lẻ. Mô hình DT được gọi là mô hình phi tham số nên khi sử dụng không có những ràng buộc giữa các biến số và với cả những dữ liệu có giá trị rỗng. Theo tài liệu (Witten, Frank and Hall, 2011) thiết lập mô hình DT qua các bước: xác định nút, xây dựng và lựa chọn DT tối ưu.

2.2.1. Xác định nút trên cây quyết định

Nút là vị trí phân chia cây quyết định thành một phân lớp mới, phương pháp phân chia có thể áp dụng cho những tập dữ liệu khác nhau. Mỗi nút được xác định dựa trên giá trị nguy cơ thuộc về nút đó nhỏ nhất, nếu $P(A_j)$ là xác suất dự báo nút A thì giá trị nguy cơ nhỏ nhất $R(A_j)$ khi đó cây quyết định T với j nút A sẽ có giá trị nguy cơ $R(T)$ là:

$$R(T) = \sum_{j=1}^k P(A_j)R(A_j) \quad (1)$$

2.2.2. Xây dựng cây quyết định

Để xây dựng cây quyết định cần phải tạo ra các nút và phân chia thành lớp (*nsplit*). Phương pháp phân chia đã trình bày ở trên nhưng phân chia dữ liệu để cây quyết định đạt giá trị chính xác và ra kết quả nhanh thì cần tiêu chuẩn để so sánh. Nếu nút A chia là hai nút A_L và A_R với $P(A)$ là xác suất dự báo tại nút A và $r(A)$ là nguy cơ tại nút A thì:

$$P(A_L).r(A_L)+P(A_R).r(A_R) \leq P(A).r(A) \quad (2)$$

Giá trị tạo ra nhánh mới là Δr lớn nhất để giảm nguy cơ gây ra ống vỡ. Bài toán sẽ hội tụ nhanh hơn khi đa dạng hóa nút A bằng cách sử dụng hàm f – hàm hỗn hợp:

$$I(A) = \sum_{i=1}^C f(p_{iA}) \quad (3)$$

Trong đó p_{iA} là phần trăm trong nút A thuộc về nút i của mẫu tương lai. Nếu C_1, C_2 là một phần của lớp C trong 2 nút mới tạo ra thì đồ thị hàm f là đường cong có đáy:

$$I(A) = \min_{C_1, C_2} [f(p_{C_1}) + f(p_{C_2})] \quad (4)$$

2.2.3. Lựa chọn cây quyết định

Một cây quyết định hoàn chỉnh bao gồm nhiều nhánh và số lượng nút khá lớn nên cần loại bỏ một phần các biến không quan trọng trong quá trình ra quyết định để có một kết quả tối ưu. Để cắt gọn cây quyết định cần tối thiểu hàm nguy cơ của nút:

$$T = R(T) = \sum_{i=1}^k P(T_i).R(T_i) \quad (5)$$

Sử dụng thuật toán tối thiểu hàm T cho toàn bộ các nhánh có giá trị nguy cơ nhỏ nhất sẽ xác định được cây quyết định tối ưu cho mô hình.

2.2.4. Áp dụng mô hình cây quyết định cho dự báo

Trong cây quyết định tối ưu, tập hợp các số liệu thống kê thuộc nút A sẽ có xác suất dự báo là $P(A_L)$ và $P(A_R)$ trong công thức (2) là giá trị dự báo khả năng ống không vỡ và ống vỡ trên mạng lưới cấp nước.

2.3. Kiểm chứng mô hình DT

Đường ống cấp nước quận Hải Châu có độ tuổi lớn nhất (hơn 36 tuổi) trong MLCN của thành phố Đà Nẵng, số lượng ống dẫn đa dạng, hồ sơ ống vỡ lưu trữ đầy đủ, nên tác giả lựa chọn làm khu vực nghiên cứu. Mục tiêu thu thập dữ liệu là xác định xác suất ống vỡ trên mạng lưới cấp nước nên tác giả tiến hành thu thập dữ liệu từ hồ sơ quản lý của công ty và tham vấn các nhân viên phụ trách kỹ thuật cũng như công nhân vận hành, sửa chữa hệ thống để hiểu rõ hơn về số liệu khảo sát cũng như các yếu tố liên quan tới sự kiện ống vỡ trong quá trình làm việc. Các thông tin trong hồ sơ sửa chữa ống vỡ của công ty bao gồm vị trí vỡ ống, đường kính (D), vật liệu (Mat) và nguyên nhân vỡ, hồ sơ ghi nhận tất cả các trường hợp sửa chữa trên mạng lưới từ hỏng van, thay đồng hồ cho đến các tác động từ bên ngoài gây ra vỡ ống và số lượng ống vỡ trên đường ống dịch vụ là lớn nhất mà nguyên nhân xảy ra hiện tượng này chủ yếu là do khách hàng sửa chữa thi công nên các số liệu

ống vỡ được chọn lọc trên các ống đường kính lớn hơn 100mm và nguyên nhân vỡ do bản thân vật liệu ống.

Hồ sơ sửa chữa ống vỡ chưa cung cấp các thuộc tính liên quan tới đường ống như chiều dài tuyến ống (L), khu vực đặt ống (DMA), năm cài đặt cũng như giá trị áp lực làm việc trung bình (P) của từng đoạn ống. Bên cạnh đó, mô hình đề xuất dùng để ước lượng khả năng vỡ cho tất cả các đoạn ống trên hệ thống phân phối nước bao gồm cả những đoạn ống đã vỡ và những đoạn ống chưa vỡ, thậm chí cả những ống mới được lắp đặt, vậy nên số liệu thống kê không chỉ thu thập từ hồ sơ ống vỡ trên mạng lưới mà còn cần các thông tin liên quan tới các đoạn ống chưa vỡ. Các thông tin này được lấy từ hệ thống GIS do phòng kỹ thuật của công ty cung cấp đồng thời các dữ liệu GIS sẽ được sử dụng để kiểm chứng lại một lần nữa các thông tin lấy được trong hồ sơ sửa chữa ống vỡ ở trên. Số liệu thu thập được tổng hợp bằng bảng tính Excel và xử lý các dữ liệu thiếu bằng khảo sát thực địa, phỏng vấn nhân viên công ty cũng như phân tích các hồ sơ thiết kế, dữ liệu hình ảnh Google Earth và hệ thống GIS.

Bảng 1 tổng hợp dữ liệu khảo sát của 9 yếu tố ảnh hưởng tới sự kiện vỡ F của quận Hải

Bảng 1. Dữ liệu mẫu sử dụng cho mô hình DT

F	A	N0	Mat	D	P	L	R	La	Prior
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
0	16,07	8	PVC	100	8,9	0,002	1299	31,85	No
0	10,38	19	PVC	100	8,9	0,001	1374	31,85	No
...
1	2,32	4	PVC	150	17,96	0,161	1317	31,85	1st
1	16,84	15	PVC	300	11,83	0,137	1140	22,94	1st

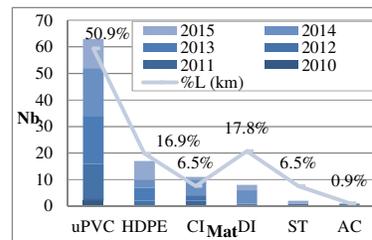
3. KẾT QUẢ VÀ THẢO LUẬN

3.1 Kết quả

Bảng 2 biểu diễn kết quả mô hình DT trong phần mềm R, chỉ số ống vỡ ban đầu `Root node error:60/1385` cho biết tổng số dữ liệu ống vỡ trên mô hình là 60 trên tổng 1385 ống, các đại lượng CP (tham số hỗn hợp),

Châu với 1979 hàng x10 cột. Dữ liệu này được chia thành hai phần, 70% để huấn luyện, 30% còn lại được sử dụng để kiểm tra và đưa vào mô hình DT bằng phần mềm R với hàm `rpart`.

Tổng chiều dài MLCN quận HC là 158,43 km ống với các loại vật liệu và tỉ lệ như Hình 3. Thống kê số lần vỡ có xu thế tăng theo chiều dài ống, trong đó ống PVC có chiều dài lớn nhất là 80,6 km với số lần vỡ trong 6 năm khảo sát là 63 lần. Vật liệu AC có số lần vỡ ít nhất là 1 lần, vật liệu DI có tỉ lệ vỡ lớn hơn, kết quả này cũng phù hợp với kết quả các nghiên cứu trước đây (Kabir, Tesfamariam and Sadiq, 2016)



Hình 3. Số liệu ống vỡ trên MLCN quận HC

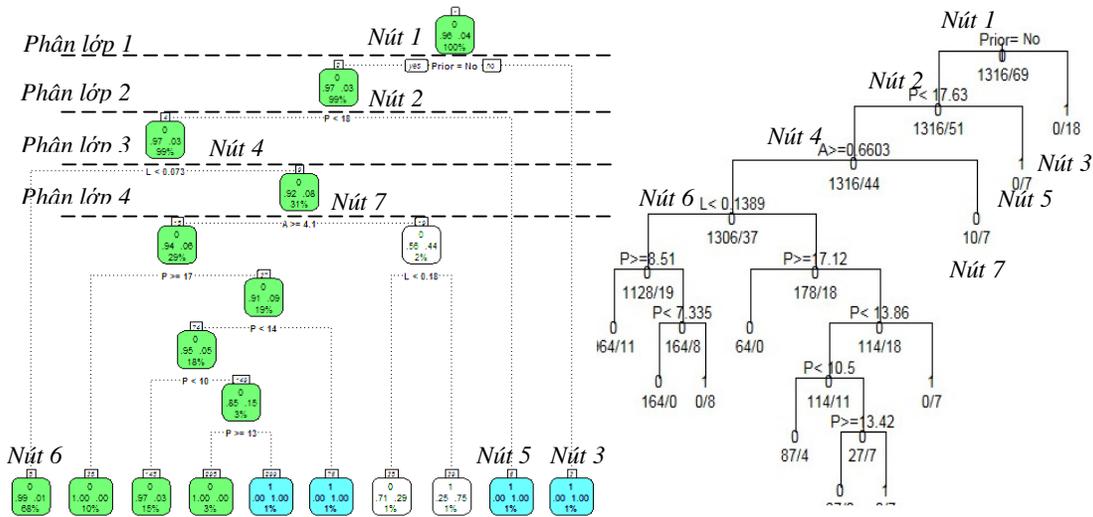
`nsplit` (số lớp được phân chia), `rel error` (chỉ số tương đối của số lượng ống vỡ trên tổng), `xerror` (chỉ số thống kê ống vỡ), `xstd` (độ lệch chuẩn) là các giá trị thể hiện quá trình cắt gọn DT.

Kết quả phân tích dữ liệu MLCN quận Hải Châu bằng mô hình DT cho thấy dự báo sự kiện

ống vỡ có 4 biến tham gia chính bao gồm độ tuổi (A), chiều dài ống dẫn (L), áp lực làm việc trung bình của từng đoạn ống (P) và lịch sử vỡ (Prior). Mô hình DT ban đầu xây dựng với đầy đủ các số liệu trong tập thống kê, sau đó bằng phương pháp chi phí tối thiểu, DT được giảm lược các biến không quan trọng và đưa ra cây tối ưu như Hình 4. Giải thích ý nghĩa của Hình 4 bằng Bảng 3 cho thấy phân lớp 1 là tập hợp dữ liệu đưa vào với 1 nút có 1325 hàng dữ liệu ống không vỡ và 60 hàng dữ liệu ống vỡ. Phân lớp 2 chia làm hai nút bằng tiêu chí lịch sử bể ống (Prior), nút 3 chứa 1% ống vỡ có lịch sử vỡ ống ≥ 1 .

Bảng 2. Kết quả mô hình DT quận Hải Châu

Variables actually used in tree construction:					
	A	L	P	Prior	
Root node error: 60/1385 = 0.043321					
n= 1385					
xstd	CP	nsplit	rel	error	xerror
1	0.200000	0	1.00000	1.00000	1.00000
0.12627					
2	0.116667	1	0.80000	0.85000	0.85000
0.11681					
3	0.046667	2	0.68333	0.85000	0.85000
0.11681					
4	0.010000	9	0.33333	0.73333	0.73333
0.10878					



Hình 4. Mô hình cây quyết định tối ưu cho MLCN quận Hải Châu

Bảng 3. Giải thích mô hình cây quyết định

Phân lớp 1: 1325/60	Nút 1 có 1325(F=0) + 60(F=1) = 1385(F) dòng dữ liệu trong đó: 1325 số liệu khảo sát có F=0; 60 hàng có F=1
Phân lớp 2: 1325/48 0/12	Nút 2 có 1325(F=0) + 48(F=1)thỏa mãn Prior<1 Nút 3 có 0(F =0)+12(F=1) có giá trị Prior >1.
Phân lớp 3: 1325/41 0/7	Nút 4 có 1325(F =0)+41(F=1) thỏa mãn P<17,63& Prior<1 Nút 5 có 0 (F =0)+ 7(F=1) thỏa mãn P>=17,63&Prior<1

So sánh với áp lực $P < 17,63$ để đưa vào phân lớp 3 thì 100% (F=0) trong tập số liệu ống

không vỡ thỏa mãn điều kiện được đưa vào nút 4. Tương tự cho các phân lớp tiếp theo, số ống không vỡ ($F=0$) ban đầu sẽ được chia nhỏ vào các nút đến phân lớp cuối cùng. Giá trị % trong mỗi nút cho biết số lượng biến trong tập không vỡ và vỡ thỏa mãn điều kiện phân chia nút.

Các ô vuông màu xanh trong Hình 4 thể hiện các nút A trong cây quyết định tối ưu, mỗi nút có ba hàng thông tin. Hàng thứ 1 là tiêu chuẩn phân chia nút ống vỡ (1) hay không vỡ (0); Hàng thứ 3 là % số liệu trong nút; Hàng thứ 2 là giá trị dự báo khả năng ống không vỡ hay vỡ theo tiêu chuẩn phân chia ở hàng thứ 1 ví dụ nút 1 có tiêu chuẩn phân chia là 0-ống không vỡ thì $P(A_L)=96\%$ nghĩa là 96% số liệu ống không vỡ nhưng nút 3 hàng thứ nhất là 1 thì $P(A_L)=0\%$ nghĩa là tập hợp số liệu trong nút có 0% khả năng ống vỡ. Các giá trị phân lớp trong cây DT này được áp dụng cho 30% số liệu còn lại, nghiên cứu sử dụng lệnh `predict` trong R lấy kết quả dự báo khả năng ống vỡ từ đó phân tích, so sánh với dữ liệu thực tế để kiểm chứng độ chính xác của mô hình đề xuất.

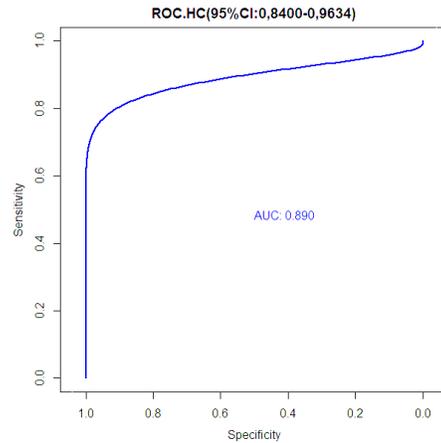
3.2 Thảo luận kết quả

Những ống được mô hình dự báo có khả năng vỡ lớn hơn hoặc bằng 95% được cho là vỡ ống có xảy ra ngoài thực tế như vậy kết quả mô hình dự báo là đúng. Bảng 4 đưa ra so sánh giữa kết quả mô hình DT và thực tế.

Kết quả cho thấy 13 ống vỡ ngoài thực tế nhưng mô hình DT lại dự báo không vỡ vậy sai số giữa dự báo và thực tế là 2,2%. Để đánh giá chất lượng của mô hình dự báo cần xem xét hai tiêu chí đường cong ROC và chỉ số AUC. Đường cong ROC là đường cong thể hiện độ nhạy (sensitivity) và độ đặc hiệu (specificity) của kết quả dự báo, một dự báo tốt khi diện tích dưới đường cong AUC là lớn nhất. Theo tài liệu (Sarah Boslaugh; and Louise-Anne McNutt, 2008) chỉ số AUC được đánh giá theo giới hạn từ 0,5-1 như Bảng 5. Sử dụng lệnh `roc` trong phần mềm R, kết quả như Hình 5 cho thấy mô hình cây quyết định đạt tiêu chuẩn $AUC = 0,890$ nằm trong giới hạn rất tốt.

Bảng 4. So sánh kết quả

Tiêu chí	Thực tế	Mô hình	Sai số
F=0	560	560	0
F=1	34	21	13
Tổng	594	581	2,2%



Hình 5. Đường cong ROC cho MLCN HC

Bảng 5. Diện tích dưới đường cong (AUC) để đánh giá hiệu suất của mô hình

AUC	Ý nghĩa diễn giải	AUC.HC
>0,9	Rất tốt	
0,8-0,9	Tốt	0,890
0,7-0,8	Trung bình	
0,6-0,7	Không tốt	
0,5-0,6	Kết quả không sử dụng được	

4. KẾT LUẬN

Mô hình cây quyết định là mô hình phi tham số và tìm kiếm kết quả tốt nhất trên các tập hợp biến độc lập nên thời gian chạy nhanh, phân lớp rõ ràng. Kết hợp giữa phương pháp tối thiểu nguy cơ và tối đa giá trị phân lớp, các nút không quan trọng đã được loại bỏ để đưa ra DT tốt nhất cho dữ liệu thống kê. Từ số liệu mạng lưới cấp nước quận Hải Châu đã kiểm chứng chất lượng của mô hình DT và cho chỉ số AUC của mô hình nằm trong giới hạn tốt.

TÀI LIỆU THAM KHẢO

- Al-barqawi, H. and Zayed, T. (2008) 'Infrastructure Management: Integrated AHP / ANN Model to Evaluate Municipal Water Mains ' Performance', (December), pp. 305–318.
- Bainbridge, H. O. and K. (2011) 'Comparison of Statistical Deterioration Models for Water Distribution Networks', Journal of Performance of Constructed Facilities, 25(May), pp. 259–266. doi: 10.1061/(ASCE)CF.1943-5509.
- Bubbis, N. S. (1948) 'Maintenance and Operating Problems of Winnipeg', Journal American Water Works Association, 41(5), pp. 429–436.
- Gómez-Martínez, P., Cubillo, F., Martín-Carrasco, F. and Garrote, L. (2017) 'Statistical Dependence of Pipe Breaks on Explanatory Variables', Water, 9(3), p. 158. doi: 10.3390/w9030158.
- Kabir, G., Tesfamariam, S. and Sadiq, R. (2016) 'Bayesian Model Averaging for the Prediction of Water Main Failure for Small to Large Canadian Municipalities', Canadian Journal of Civil Engineering, 43(3), pp. 233–240. doi: 10.1139/cjce-2015-0374.
- Leo Breiman; Jerome H. Friedman; Richard A. Olshen; Charles J. Stone (1984) *Classification and regression trees*. Chapman & Hall/CRC.
- Pelliccia, T. M. W. A. (1981) *Water main repair/replacement for Binghamton, N. Y.* Vicksburg, Miss. 39180.
- Sarah Boslaugh; and Louise-Anne McNutt (2008) *Encyclopedia of Epidemiology*. SAGE. Edited by M. Birdsall, Diana E. Axelsen, and Lisa Cuevas Shaw. California: Rolf A. Janke.
- Sheikh, A. K., Boah, J. K. and Hansen, D. A. (1990) 'Statistical modeling of pitting corrosion and pipeline reliability', Corrosion, pp. 190–197. doi: 10.5006/1.3585090.
- Wengström, T. (1993) 'Comparative analysis of Pipe Break Rates', Chalmers University of Technology.
- Witten, I. H., Frank, E. and Hall, M. a (2011) *Data Mining: Practical Machine Learning Tools and Techniques (Google eBook)*, Complementary literature None. doi: 0120884070, 9780120884070.
- Yamijala, S. (2007) *Statistical estimation of water distribution system pipe break risk*. Texas A&M University.

Abstract:

PROPOSING A MODEL FOR PIPE FAILURE ESTIMATION IN WATER SUPPLY NETWORK

In urban area, most of water supply pipes are located under roads or pavements. Therefore when the pipe failures occur, a significant amount of water is lost. At the same time, contaminants from outside are likely to enter the pipes, causing widespread water pollution in the water supply network, making possible risks of massive contamination for water consumers. In order to minimize water loss and to ensure water consumers' health, it is necessary to identify the probability of pipe failures in the water supply network, so that pipes maintenance and replacement planning is enabled to be made in time. This paper analyzes quantitatively the factors causing pipe failures in the water supply network and proposes the use of decision tree model to estimate the probability of pipe failures. The proposed model is validated in actual statistics and evaluated by AUC indicator.

Keywords: Water supply network; Pipe ruptures; AUC indicator; Decision tree model; R software

Ngày nhận bài: 09/10/2017

Ngày chấp nhận đăng: 08/1/2018