

ỨNG DỤNG HỆ SUY DIỄN MỜ PHỨC TRONG HỖ TRỢ CHẨN ĐOÁN BỆNH XƠ GAN

Trần Thị Ngân^{1,2,3*}, Nguyễn Thị Dung⁴, Nguyễn Long Giang², Trần Mạnh Tuấn³

¹Học viện Khoa học và Công nghệ - Viện hàn lâm Khoa học và Công nghệ Việt Nam,

²Viện Công nghệ thông tin - Viện hàn lâm Khoa học và Công nghệ Việt Nam,

³Trường Đại học Thủy lợi, ⁴Trường Đại học Công nghệ thông tin và truyền thông - ĐH Thái Nguyên

TÓM TẮT

Bài toán chẩn đoán bệnh là bài toán phổ biến trong y học. Việc chẩn đoán đúng và chính xác có ý nghĩa quan trọng trong việc điều trị của bệnh nhân. Chẩn đoán sớm và chính xác giúp việc điều trị có hiệu quả cao với chi phí thấp hơn rất nhiều. Có nhiều nghiên cứu đưa ra các phương pháp chẩn đoán bệnh sử dụng trí tuệ nhân tạo và học máy. Ngoài ra, lý thuyết tập mờ và logic mờ cũng có vai trò to lớn trong việc giải quyết bài toán chẩn đoán bệnh. Trong bài báo này, chúng tôi giới thiệu một phương pháp dựa trên lý thuyết về tập mờ phức để xây dựng hệ hỗ trợ ra quyết định. Từ đó xây dựng ứng dụng để giải bài toán chẩn đoán bệnh xơ gan trên bộ dữ liệu thực tế được thu thập từ các bệnh viện ở Thái Nguyên. Kết quả thực nghiệm chỉ ra rằng, mô hình đề xuất có kết quả hỗ trợ chẩn đoán cao hơn các phương pháp FMNN, SVM, FIS, FLT được so sánh.

Từ khóa: Tập mờ phức; hỗ trợ chẩn đoán bệnh; tập mờ; hệ hỗ trợ ra quyết định; học máy; trí tuệ nhân tạo.

Ngày nhận bài: 06/02/2020; Ngày hoàn thiện: 29/4/2020; Ngày đăng: 11/5/2020

USING COMPLEX FUZZY INFERENCE SYSTEM IN LIVER DISEASE DIAGNOSIS SUPPORT

Tran Thi Ngan^{1,2,3*}, Nguyen Thi Dung⁴, Nguyen Long Giang², Tran Manh Tuan³

¹Graduate University of Science and Technology – VAST,

²Institution of Information Technology – VAST,

³Thuyloi University, ⁴TNU - University of Information and Communication Technology

ABSTRACT

Disease diagnosis problem is a very popular problem in medicine. The early and accurate diagnosis will reduce the treatment cost and increase the probability of success for patients. In recent years, there were many researches related to medical support via machine learning methods. In this paper, we introduce the integration model including transfer learning and complex fuzzy set in order to solve this problem. Our proposed model is applied in a real data set related to liver diseases. This data set was collected from hospitals in Thai Nguyen to compare with different methods. The experimental results show that our model gets the best performance.

Keywords: Complex fuzzy set; Disease diagnosis support; Fuzzy set; Decision making support; Machine learning; Artificial intelligence.

Received: 06/02/2020; Revised: 29/4/2020; Published: 11/5/2020

* Corresponding author. Email: nganttt@tlu.edu.vn

1. Giới thiệu

Logic mờ đã phát triển khá hoàn chỉnh và kết hợp với một số ngành khoa học khác tạo nên cơ sở để hình thành các công cụ dựa trên lý thuyết mờ. Đóng góp của logic mờ rất quan trọng cho lý thuyết về tập mờ, hệ mờ. Trong năm 2014, Kantesh Kumar OAD và Xu Dezhi [1] đề xuất một phương pháp tiếp cận dựa trên nguyên tắc mờ để dự đoán mức độ rủi ro của các bệnh về tim mạch. Sutton [2] sử dụng thuật toán K-láng giềng gần nhất mờ (Fuzzy K-nearest neighbor - FKNN) cho các bài toán về y tế khác nhau bao gồm chẩn đoán nha khoa. Trong năm 2018, Hamido Fujita và cộng sự [3] đã đề xuất một phương pháp hỗ trợ chẩn đoán dựa trên trích chọn các đặc trưng của ảnh nha khoa. Lý thuyết tập mờ đã được sử dụng trong một số hệ chuyên gia y tế [4].

Diễn tiến bệnh gan mang tính chất lịch sử (mang tính chu kỳ, định kỳ) và không xác định (đối với mỗi đối tượng thì diễn tiến bệnh lại khác nhau). Ngoài ra các thông tin về bệnh có mối liên hệ tương hỗ nhau, do đó không chỉ dựa vào một chỉ số nào đó mà có thể kết luận bệnh. Lý thuyết mờ phức là công cụ hữu hiệu để giải quyết các vấn đề kể trên.

Bài toán chẩn đoán nói riêng và bài toán hỗ trợ ra quyết định nói chung có thể được giải quyết bằng các cách tiếp cận khác nhau. Trong bài báo này, nhóm tác giả đề xuất một ứng dụng của tập mờ phức trong bài toán chẩn đoán bệnh. Đối với các phương pháp khác, cách tiếp cận giải quyết bài toán là hoàn toàn khác nhau. Phần thực nghiệm của bài báo so sánh độ chính xác, sai số bình phương trung bình và sai số tuyệt đối trung bình nhận được từ mô hình đề xuất với các cách tiếp cận đã có gồm FIS, FMNN, SVM, FTL. Sau khi xây dựng mô hình, nhóm tác giả thực hiện cài đặt mô hình đề xuất trên bộ dữ liệu cụ thể và có đánh giá hiệu năng thông qua việc so sánh với các phương pháp tương tự khác.

Phần tiếp theo của bài báo được bố cục như sau: các kiến thức lý thuyết nền tảng sẽ được trình bày trong phần 2. Phần 3 trình bày cụ

thể về mô hình được đề xuất trong bài toán chẩn đoán bệnh trên bộ dữ liệu cụ thể. Các kết quả đánh giá thực nghiệm để so sánh hiệu năng của mô hình với các mô hình đã có khác được trình bày trong phần 4. Cuối cùng là một số kết luận được trao đổi ở phần 5 của bài báo.

2. Tập mờ phức và các khái niệm cơ bản

Logic mờ, tập mờ phức và suy diễn mờ phức

Logic mờ là một cách mới để biểu diễn xác suất: Logic mờ và xác suất nói đến các loại không chắc chắn khác nhau. Logic mờ được thiết kế để làm việc với các sự kiện **không chính xác** (các mệnh đề Logic mờ), trong khi xác suất làm việc với các **khả năng** sự kiện đó xảy ra (*nhưng vẫn coi kết quả là chính xác*).

Tập mờ phức và ứng dụng

Khái niệm về **tập mờ phức** (Complex Fuzzy Set - CFS) và **logic mờ phức** (Complex Fuzzy Logic - CFL) [5], [6] đã được đề xuất bởi Ramot và các cộng sự như là một phần mở rộng của lý thuyết tập mờ và logic mờ. Một tập mờ phức được đặc trưng bởi một hàm thuộc giá trị phức $\mu_s(x)$ mà phạm vi giá trị của nó là đường tròn đơn vị trong không gian phức, và được biểu diễn có dạng:

$$\mu_s(x) = r_s(x) \cdot e^{j\omega_s(x)}, \quad j = \sqrt{-1}$$

Trong đó: $r_s(x)$ là biên độ và $\omega_s(x)$ pha, và cả 2 đều là các hàm có giá trị thực với điều kiện $r_s(x) \in [0, 1]$. Pha không phải là 1 hàm mờ có thể nhận bất kỳ giá trị thực nào nên nó không ảnh hưởng tới mức độ thuộc. Pha bổ sung thêm một số thông tin mở rộng liên quan tới chu kỳ không gian và thời gian trong tập mờ đã được xác định bởi biên độ. Tuy nhiên khái niệm về tập mờ phức này khác với các khái niệm do Buckley [7] và Zhang [8] đưa ra. Các tập mờ phức giữ lại các đặc điểm về sự không chắc chắn dưới dạng biên độ, trong khi thêm vào thành phần pha để chỉ ra các thuộc tính dạng sóng. Các tập mờ phức có một vài ứng dụng trong các lĩnh vực khác nhau như vật lý, xử lý tín hiệu [9], ra quyết định [10], chứng khoán [11]. Suy diễn là cơ

chế liên kết các tri thức đã có để suy dẫn ra các tri thức mới. Cơ chế suy diễn phụ thuộc rất nhiều vào phương thức biểu diễn tri thức và không có một phương pháp suy diễn duy nhất cho mọi loại tri thức.

Hệ suy diễn mờ [12] là một cơ chế suy diễn thường xuyên được áp dụng khi xây dựng các hệ chuyên gia. Hệ suy diễn mờ tỏ ra hiệu quả trong trường hợp tri thức không đầy đủ, bất định hoặc không chính xác.



Hình 1. Sơ đồ hệ suy diễn mờ

Hệ suy diễn mờ (hình 1) gồm các thành phần sau:

- Giao diện mờ hóa: chuyển đổi các lớp đầu vào vào các biến độ phù hợp với các giá trị ngôn ngữ.
- Cơ sở tri thức bao gồm 2 phần:
 - Cơ sở dữ liệu: định nghĩa các hàm thuộc của các tập mờ được sử dụng trong các luật mờ
 - Bộ luật: gồm các luật mờ IF – THEN
- Đơn vị thực thi: thực hiện các hoạt động suy diễn trong các luật
- Giao diện giải mờ: chuyển đổi các giá trị kết quả mờ của hệ suy diễn ra các lớp đầu ra

Các bước suy diễn mờ:

- Mờ hóa các biến vào: ta cần mờ hóa những giá trị rõ để tham gia vào quá trình suy diễn
- Áp dụng các toán tử mờ (AND hoặc OR) cho các giả thiết của từng luật.
- Áp dụng phép kéo theo để tính toán giá trị các giá trị từ giả thiết đến kết luận của từng luật.
- Áp dụng toán tử gộp để kết hợp các kết quả trong từng luật thành một kết quả duy nhất cho cả hệ.
- Giải mờ kết quả tìm được cho ta một kết quả rõ.

Hệ suy diễn mờ phức (Complex Fuzzy Inference System – CFIS) theo Mamdani được nhóm tác giả Sơn và cộng sự đưa ra [13].

Các nghiên cứu trên tập mờ phức tập trung vào việc xây dựng các hệ logic mờ phức, mạng Neural mờ phức và mạng ARIMA mờ phức kết hợp với một số thuật toán học kinh điển để nâng cao hiệu năng của hệ thống trong các ứng dụng thực tế.

3. Mô hình hỗ trợ chẩn đoán dùng tập mờ phức

Mô hình hỗ trợ chẩn đoán được xây dựng thành 2 pha:

Pha 1: Xây dựng hệ luật mờ phức từ một phần dữ liệu ban đầu (hình 2). Pha 1 bao gồm 3 bước chính:

- Mờ phức hoá dữ liệu huấn luyện (training) bằng cách xác định phần thực, phần ảo tương ứng của dữ liệu đầu vào.
- Áp dụng hệ suy diễn mờ phức CFIS vào bộ dữ liệu đã được mờ phức hoá
- Xây dựng hệ luật mờ từ kết quả của quá trình áp dụng CFIS trên.

Quá trình mờ phức hoá dữ liệu nhằm xác định giá trị phần thực và phần ảo. Trong đó, phần thực được xác định là giá trị của dữ liệu đầu vào. Phần ảo được xác định là phương sai. Sau đó, mờ hóa giá trị phần thực và phần ảo nhận được để xác định ma trận độ thuộc cho phần thực và phần ảo.

Hệ luật kết quả của Pha 1 bao gồm các luật có dạng:

$$\text{If } x \text{ is } A \text{ then } y \text{ is } B$$

Trong đó x và y là các biến được lấy từ hai vũ trụ U và V (khác nhau) tương ứng. A và B là các tập mờ phức được xác định trên U, V ;

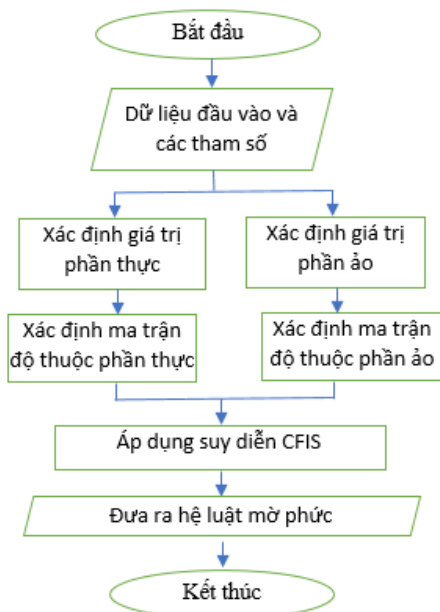
Trong sơ đồ ở hình 2, dữ liệu sử dụng là dữ liệu training.

Pha 2: Áp dụng hệ luật mờ phức nhận được từ pha 1 trên phần dữ liệu còn lại để nhận được kết quả chẩn đoán (hình 3).

- Mờ phức hoá dữ liệu testing bằng cách xác định phần thực, phần ảo tương ứng của dữ liệu đầu vào.

- Áp dụng hệ suy diễn mờ phức CFIS vào bộ dữ liệu testing đã được mờ phức hoá.

Đưa ra kết quả chẩn đoán từ hệ suy diễn.



Hình 2. Sơ đồ xây dựng hệ luật mờ phức



Hình 3. Sơ đồ hỗ trợ chẩn đoán

Trong sơ đồ hỗ trợ chẩn đoán (hình 3) dữ liệu sử dụng là testing. Ta xác định giá trị phần thực và phần ảo, từ đó làm mờ hóa giá trị phần thực và phần ảo để xác định ma trận độ thuộc cho phần thực và phần ảo. Sử dụng hệ luật mờ phức xác định trên đưa ra kết quả hỗ trợ chẩn đoán.

Kết quả hỗ trợ chẩn đoán được đưa ra sau khi thực nghiệm là kết luận: dữ liệu đầu vào là của bệnh nhân mắc bệnh xơ gan (gán nhãn 1) hay không mắc bệnh xơ gan (gán nhãn -1).

4. Kết quả thực nghiệm

Trong thực nghiệm với bộ dữ liệu từ các bệnh nhân được bác sĩ chỉ định cận lâm sàng với các xét nghiệm công thức máu và sinh hóa máu để chẩn đoán xơ gan. Từ đó có thể đánh giá xơ hóa gan một cách rộng rãi, lặp lại nhiều lần đối với viêm gan mạn, do đó rất có ý nghĩa trong việc phát hiện và theo dõi diễn tiến bệnh, nhất là theo dõi đáp ứng về mặt giảm mức độ xơ hóa sau điều trị viêm gan mạn do vi-rút, bệnh gan nhiễm mỡ không do rượu. Tập dữ liệu này gồm 320 bệnh nhân đến khám và điều trị bệnh do rối loạn men gan tại Bệnh viện Gang thép Thái Nguyên và Bệnh Viện Đa khoa Trung ương Thái Nguyên. Mỗi hồ sơ bệnh nhân chứa các thông tin liên quan tới bệnh nhân đến khám và điều trị các bệnh do rối loạn men gan. Trong số 320 hồ sơ bệnh nhân gồm 2 nhóm: nhóm 1 gồm 150 hồ sơ bệnh nhân được cho là không bị xơ gan; nhóm 2 gồm 170 hồ sơ bệnh nhân được chẩn đoán là xơ gan. Các thông tin sau được trích từ hồ sơ bệnh nhân, có liên quan đến chẩn đoán bệnh của bác sĩ (một số thông tin khác được bảo vệ vì lý do bảo mật). Các thông tin bao gồm: tuổi được tính đến ngày làm xét nghiệm (tuổi lớn hơn 90 được coi là 90 tuổi); men AST hay còn gọi là SGOT; men ALT và tiểu cầu. Các thông tin trên được cấu thành 4 thuộc tính đầu vào cho thực nghiệm.

Trong thực nghiệm này, dữ liệu training chiếm 2/3 tập dữ liệu ban đầu và phần còn lại của dữ liệu được dùng cho testing.

Ngôn ngữ sử dụng để cài đặt thực nghiệm là Matlab 2014. Các độ đo dùng để đánh giá và so sánh hiệu năng của các thuật toán được cài đặt trong bài báo này gồm Accuracy (Acc) [14], MSE [14], MAE [14]. Các phương pháp sử dụng để so sánh: suy diễn mờ (FIS – Fuzzy Inference System) [12], Máy vector hỗ trợ (SVM-Support Vector Machine) [15], Mạng

ơ-ron min-max mờ (FMNN – Fuzzy Min-max Neural Network) [16] và học chuyển giao mờ (FTL - Fuzzy Transfer Learning) [17].

Kết quả thực hiện chương trình được thể hiện trong hình 4 và hình 5 dưới đây. Trong đó, hình 4 thể hiện kết quả gán nhãn trên bộ dữ liệu testing với nhãn 1 (tương ứng là với chẩn đoán “Có bệnh”) và nhãn -1 (tương ứng là với đầu ra “Không có bệnh”).



Hình 4. Kết quả gán nhãn cho dữ liệu trong testing

Hình 5 thể hiện kết quả khi tính toán độ đo Accuracy, MSE, MAE trên bộ dữ liệu testing của mô hình đề xuất (đối với các mô hình dùng để so sánh, giao diện thực hiện là tương tự).



Hình 5. Kết quả tính toán giá trị các độ đo khi áp dụng mô hình đề xuất

Các độ đo Accuracy, MSE, MAE được tính toán và biểu diễn kết quả ở bảng 1.

Bảng 1. Kết quả thực nghiệm.

Thuật toán	Acc	MSE	MAE
FMNN	85,94	0,928	0,895
SVM	83,56	1,437	1,234
FIS	87,00	1,132	1,038
FTL	89,31	0,892	0,873
CFIS	90,41	0,853	0,856

Trong bảng 1, các giá trị ở cả 3 độ đo đương dùng để đánh giá khi áp dụng mô hình dựa trên tập mờ phức tốt hơn các phương pháp SVM, FIS, FLT.

5. Kết luận

Trong bài báo này, chúng tôi đã trình bày việc sử dụng suy diễn mờ phức trong hỗ trợ chẩn

đoán. Bài báo có một số đóng góp chính như sau: (i) đã vận dụng mô hình suy diễn mờ phức trong bài toán hỗ trợ chẩn đoán bệnh; (ii) đã cài đặt thực nghiệm mô hình suy diễn mờ phức cho hỗ trợ chẩn đoán bệnh dựa trên bộ dữ liệu thu thập thực tế tại bệnh viện Đa khoa Gang thép Thái Nguyên và Bệnh viện Đa khoa Trung ương Thái Nguyên; (iii) Kết quả thực nghiệm dựa trên 3 độ đo MSE, Accuracy, MAE cũng đã cho thấy suy diễn mờ phức cho kết quả tốt hơn so với một số phương pháp khác.

Nghiên cứu này tạo tiền đề cho các nghiên cứu tiếp theo về việc giải quyết một số bài toán hỗ trợ chẩn đoán y tế.

Lời cảm ơn

Nghiên cứu này được thực hiện dưới sự tài trợ của đề tài sau tiến sĩ, mã số: GUST.STS.ĐT2017- TT02 từ Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Ngoài ra, nhóm tác giả xin chân thành cảm ơn sự hỗ trợ và hợp tác từ đơn vị phối hợp, Viện Công nghệ thông tin, Viện hàn lâm Khoa học và Công nghệ Việt Nam.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1]. K. K. Oad, X. DeZhi, and P. K. Butt, “A Fuzzy Rule Based Approach to Predict Risk Level of Heart Disease,” *Global Journal of Computer Science and Technology*, vol. 14, no. 3, pp. 16-22, 2014.
- [2]. E. Ramírez, O. Castillo, and J. Soria, *Hybrid System for Cardiac Arrhythmia Classification with Fuzzy K-Nearest Neighbors and Neural Networks Combined by a Fuzzy Inference System*, In Soft Computing for Recognition Based on Biometrics, Springer Berlin Heidelberg, pp. 37-55, 2010.
- [3]. L. H. Son, T. M. Tuan, H. Fujita, N. Dey, A. S. Ashour, V. T. N. Ngoc, and D. T. Chu, “Dental diagnosis from X-Ray images: An expert system based on fuzzy computing,” *Biomedical Signal Processing and Control*, vol. 39, pp. 64-73, 2018.
- [4]. J. Shell, and S. Coupland, “Fuzzy transfer learning: methodology and application,” *Information Sciences*, vol. 293, pp. 59-79, 2015.

- [5]. D. Ramot, R. Milo, M. Friedman, and A. Kandel, "Complex fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 171-186, 2002.
- [6]. D. Ramot, M. Friedman, G. Langholz, and A. Kandel, "Complex fuzzy logic," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 4, pp. 450-461, 2003.
- [7]. J. J. Buckley, "Fuzzy complex analysis II: integration," *Fuzzy Sets and Systems*, vol. 49, no. 2, pp. 171-179, 1992.
- [8]. Z. Guang-Quan, "Fuzzy limit theory of fuzzy complex numbers," *Fuzzy Sets and Systems*, vol. 46, no. 2, pp. 227-235, 1992.
- [9]. X. Ma, J. Zhan, M. Khan, M. Zeeshan, S. Anis, and A. S. Awan, "Complex fuzzy sets with applications in signals," *Computational and Applied Mathematics*, vol. 38, no. 4, p. 150, 2019.
- [10]. H. Garg, and D. Rani, "A robust correlation coefficient measure of complex intuitionistic fuzzy sets and their applications in decision-making," *Applied Intelligence*, vol. 49, no. 2, pp. 496-512, 2019.
- [11]. L. Y. Wei, T. L. Chen, and T. H. Ho, "A hybrid model based on adaptive-network-based fuzzy inference system to forecast Taiwan stock market," *Expert Systems with Applications*, vol. 38, no. 11, pp. 13625-13631, 2011.
- [12]. J. S. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE transactions on systems, man, and cybernetics*, vol. 23, no. 3, pp. 665-685, 1993.
- [13]. G. Selvachandran, S. G. Quek, L. T. H. Lan, N. L. Giang, W. Ding, M. Abdel-Basset, and V. H. C. Albuquerque, "A New Design of Mamdani Complex Fuzzy Inference System for Multi-attribute Decision Making Problems," *IEEE Trans. Fuzzy Syst.*, 2019, doi:10.1109/TFUZZ.2019.2961350
- [14]. R. J. Hyndman, and A. B. Koehler, "Another look at measures of forecast accuracy," *International journal of forecasting*, vol. 22, no. 4, pp. 679-688, 2006.
- [15]. C. Cortes, and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.
- [16]. T. N. Tran, D. M. Vu, M. T. Tran, and B. D. Le, "The Combination of Fuzzy Min-Max Neural Network and Semi-supervised Learning in Solving Liver Disease Diagnosis Support Problem," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 2933-2944, 2018.
- [17]. T. N. Tran, T. D. Nguyen, M. T. Tran, T. H. L. Luong, "Fuzzy transfer learning model in cirrhosis diagnosis support," *Journal of Science and Technology*, vol. 189, no. 13, pp. 93-98, 2018.