

Dùng logic mờ dự đoán kết quả thi của sinh viên

Lê Thị Kim Anh¹, Đinh Phước Vinh²

¹ Trường Đại học Ngân hàng Thành phố Hồ Chí Minh
56 Hoàng Diệu 2, quận Thủ Đức,
Thành phố Hồ Chí Minh, Việt Nam
Email: anhkt@buh.edu.vn

² Trường Đại học FPT cơ sở Thành phố Hồ Chí Minh
Lô E2a-7, Đường D1, khu Công Nghệ Cao,
Quận 9, Thành phố Hồ Chí Minh, Việt Nam
Email: vinhdp2@fpt.edu.vn

TÓM TẮT: Bài viết trình bày việc sử dụng logic mờ để dự đoán kết quả thi của sinh viên nhằm giúp giảng viên dùng lớp có cơ sở đưa ra những tác động sư phạm phù hợp để nâng cao chất lượng dạy và học. Mô hình đánh giá nhận hai biến đầu vào là điểm kiểm tra giữa kì và số buổi sinh viên nghỉ học sau nửa thời gian học tập. Các biến được mờ hóa thành ba mức để đưa vào mô hình suy diễn với chỉ sáu luật suy diễn. Điểm thi khi tính toán bằng mô hình được so sánh với điểm thi thực tế để đánh giá độ chính xác của mô hình. Với dữ liệu 86 sinh viên học môn Toán rời rạc tại Trường Đại học FPT Thành phố Hồ Chí Minh, mô hình cho độ chính xác 79.9% tương đồng với các nghiên cứu trước sử dụng nhiều biến và nhiều luật hơn.

TỪ KHÓA: Dự đoán; logic mờ; phương pháp đánh giá; phương pháp giảng dạy; suy diễn mờ.

→ Nhận bài 12/12/2019 → Nhận kết quả phản biện và chỉnh sửa 26/1/2020 → Duyệt đăng 25/02/2020.

1. Đặt vấn đề

Việc đảm bảo chất lượng giảng dạy luôn là điều được quan tâm không chỉ của sinh viên (SV), nhà trường mà trước hết là của bản thân giáo viên đứng lớp. Hiệu quả của hoạt động này phần nào được phản ánh bởi kết quả thi hết môn của SV trong điều kiện khách quan nhất. Đánh giá sớm khả năng học tập của SV một cách hiệu quả giúp người dạy sớm có những tác động phù hợp đến các đối tượng SV khác nhau nhằm nâng cao chất lượng dạy và học. Các nghiên cứu gần đây dùng nhiều biến và dùng nhiều phương pháp khác nhau trong khai phá dữ liệu và học máy để đánh giá năng lực của SV (Rusli, 2008). Trong các nghiên cứu này, nhiều thông tin có thể được dùng để dự đoán kết quả học tập của SV như thông qua điểm quá trình hay điểm giữa kì, thu nhập của phụ huynh, giới tính, loại hình học tập ở phổ thông.

Sau khi tập mờ được giới thiệu bởi Zadeh lần đầu tiên năm 1965, các mô hình suy diễn mờ cũng đã ra đời và đã có nhiều ứng dụng trong khoa học kĩ thuật như điều khiển mờ, ra quyết định trong kinh doanh (George J. Klir & Bo Yuan, 1995). Một số nghiên cứu đã dùng logic mờ để dự đoán kết quả học tập của SV và đạt được một số kết quả nhất định như của Yildiz (2013), Rao D. H. (2017) cùng cộng sự. Trong đó, nghiên cứu của Yildiz và các cộng sự (2013) đã dùng hệ chuyên gia mờ để xếp loại SV trong năm học và đạt kết quả với độ chính xác 78% trên 218 SV học thực tiễn. Nhóm tác giả sử dụng năm biến đầu vào, mờ hóa mỗi biến với năm mức và sử dụng 51 luật suy diễn. Trong mô hình tương tự, chúng tôi chỉ sử dụng hai biến đầu vào để phù hợp với dữ liệu thực tế của SV tại Trường Đại học FPT Thành phố Hồ Chí Minh. Với việc SV phải có mặt ở lớp tối thiểu 80% thời gian học tập mỗi môn học (đồng nghĩa SV được phép vắng mặt tối đa 20% số buổi học trên lớp) cùng với khối lượng kiến thức hàn lâm tương đối nhiều ở môi trường đại học, SV nghỉ học nhiều phần lớn được dự đoán

có kết quả thi không tốt. Tại Trường Đại học FPT, SV học mỗi môn học với thời lượng 30 slot (mỗi slot 90 phút). Do đó, SV được phép nghỉ tối đa 6 slot mỗi môn học. Trong bài này, người viết chọn biến đầu vào đầu tiên là số slot SV đã nghỉ. Biến đầu vào còn lại là điểm trung bình hai bài quiz (trắc nghiệm trên hệ thống lms của nhà trường) đầu tiên. Đầu ra của mô hình là điểm thi dự đoán của SV. Việc chỉ dùng hai biến đầu vào giúp loại bỏ một số khó khăn nhất định về mặt thu thập số liệu cũng như xây dựng các luật suy diễn (nhờ vào các chuyên gia). Kết quả đạt được tương đồng với các nghiên cứu trước đó về độ chính xác trong khi số biến đầu vào cũng như số luật suy diễn được sử dụng ít hơn rất nhiều.

2. Nội dung nghiên cứu

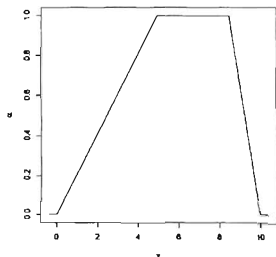
2.1. Tập mờ và suy diễn sử dụng logic mờ

Ngôn ngữ tự nhiên hàng ngày chứa đựng nhiều từ mang thông tin thiếu chính xác và mơ hồ, hoặc mơ nhạt. Chẳng hạn, "rất giỏi" là một khái niệm mờ, do không mang thông tin chính xác về điểm số, "rất giỏi" có thể là 9 hoặc 10 trong thang điểm 10. Khái niệm tập mờ do Zadeh (1965) đưa ra nhằm mục đích thao tác và rút ra những suy luận từ các thông tin thiếu chắc chắn. Có thể biểu diễn khái niệm "rất giỏi" bằng đồ thị của hàm thành viên như Hình 1. Trong đó, nếu điểm số từ 8.5 đến 10 thì nói "rất giỏi" là hoàn toàn đúng (100% tương ứng với giá trị 1 của hàm thành viên). Với 8 điểm trên thang điểm 10, giá trị hàm thành viên (giá trị μ trong hình) là 0.857 hay ta nói đây là điểm số "rất giỏi" với độ đúng đắn 85.7%. Tương tự, nếu điểm số là 5 trở xuống thì hàm thành viên có giá trị 0, tức hoàn toàn không chính xác khi nói điểm 5 trở xuống là "rất giỏi".

Tập mờ A trong U được định nghĩa là một hàm số xác định trên U và nhận giá trị trong tập $[0, 1]$. Hàm số này được gọi là hàm thành viên, và giá trị của hàm tại mỗi x thuộc U cho biết mức độ nhiều hay ít x thuộc về A. Điểm

x là “rất giỏi” là một phát biểu mờ mà độ chính xác (hay phần trăm đúng) bằng giá trị hàm thành viên có được tại điểm số x .

Từ định nghĩa cho thấy, đồ thị hàm thành viên của tập mờ có thể có dạng bất kỳ thay vì có dạng hình thang như ví dụ trên (xem Hình 1).



Hình 1: Hàm thành viên mô tả điểm số “rất giỏi”

Người ta cũng xây dựng các phép toán trên các tập mờ tương tự như các tập hợp thông thường (Timothy, 2010). Trên tập mờ, các phép toán hợp, giao, bù, kéo theo dạng nếu-thì được xây dựng ở mức độ tổng quát hơn mà ở đó tập hợp thông thường cũng chính là một tập mờ với hàm thành viên đặc biệt. Nghiên cứu này sử dụng các luật suy diễn dạng “nếu x là A và y là B, thì z là C” trong đó các phát biểu “ x là A”, “ y là B” và “ z là C” đều là các phát biểu mờ gắn liền với các tập hợp mờ nào đó. Để xây dựng hệ suy diễn mờ chúng tôi sử dụng ngôn ngữ R với gói lệnh sets (David Meyer, Kurt Hornik, Christian Buchta, 2017). Các ngôn ngữ khác cũng có thể được sử dụng thay thế như Python hay Matlab.

2.2. Dữ liệu khảo sát và các luật suy diễn mờ

Với chưa tới 30 SV một lớp học, thông tin về số slot nghỉ và điểm giữa kì của 4 lớp học môn toán rời rạc (MAD101) được thu thập với tổng số 86 SV sau khi đã loại bỏ các dữ liệu khuyết. Dữ liệu nghiên cứu với 86 SV, quan sát lúc này được ghi nhận có cấu trúc như ở Bảng 1.

Bảng 1: Cấu trúc dữ liệu nghiên cứu

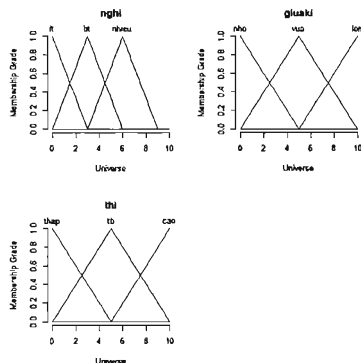
Số slot nghỉ sau 1/2 thời gian (nghỉ)	Kiểm tra giữa kì (Quiz 1 + Quiz 2)/2 (giuaki)	Điểm thi (thì)
3	6.5	5
1	6.2	5.8
5	5.9	4.8
5	7	5
1	5.3	5.8

Phân tích thống kê trên tập dữ liệu cho các thông tin cơ bản cho ở Bảng 2.

Bảng 2: Các thống kê cơ bản của dữ liệu nghiên cứu

	nghỉ	giuaki	thì
Nhỏ nhất	0	2.90	2.70
Trung vị	2	7.15	5.60
Trung bình	2.36	6.85	5.79
Lớn nhất	6	9.50	9.80
Độ lệch chuẩn	2.01	1.58	1.63

Để sử dụng logic mờ làm mô hình dự đoán, các biến liên quan được mờ hóa thành ba mức với các hàm thành viên như Hình 2. Theo đó, điểm giữa kì (giuaki) 5 được xem là vừa (vua) với độ chính xác 100%, 0% là nhỏ (nhỏ) và 0% là lớn (lớn) trong khi với điểm giữa kì 2.5 thì 50% nhỏ, 50% vừa và 0% lớn.



Hình 2: Các hàm thành viên

Ngoài ra, một số luật suy diễn ở dạng “Nếu - Thì” được sử dụng để tính toán giá trị đưa ra theo mô hình mờ ở Hình 3. Dựa vào kinh nghiệm đứng lớp của người viết cũng như tham khảo ý kiến các chuyên gia trong lĩnh vực giáo dục, 6 luật suy diễn như Bảng 3 được sử dụng. Yildiz (2013) trong công trình của mình cũng chỉ ra rằng việc xây dựng các luật suy diễn dựa vào kiến thức các chuyên gia có kinh nghiệm góp phần làm giảm sai số dự đoán của hệ thống. Có thể hiểu các luật 1 như sau: Nếu điểm giữa kì lớn và số ngày nghỉ học là không nhiều thì điểm thi của SV được mô hình dự đoán là cao (xem Bảng 3).



Hình 3: Cấu trúc của mô hình suy diễn bằng logic mờ

Bảng 3: Các luật suy diễn

Luật	Điều kiện giữa kì	Và nếu số ngày nghỉ học	Thi dự đoán điểm thi
1	Lớn	Không nhiều	Cao
2	Lớn	Nhiều	Trung bình
3	Vừa	Nhiều	Thấp
4	Nhỏ	Ít	Trung bình
5	Nhỏ	Không ít	Thấp
6	Vừa	Không nhiều	Trung bình

Kết quả thực nghiệm:

Sử dụng mô hình ta có thể dự đoán điểm thi của một SV có điểm kiểm tra giữa kì 6.5 và số slot nghỉ học là 3. Hàm thành viên mô tả điểm thi của SV này được cho như ở Hình 4. Mô hình cũng cho kết quả dự đoán SV sẽ đạt khoảng 5.23 trong kì thi cuối khóa. Kết quả thực nghiệm trên tập dữ liệu 86 SV cho độ chính xác 79.9% thông qua việc tính sai số của mô hình. Điểm thi ban đầu có giá trị mờ, sau được tính toán thành giá trị thực để trở thành điểm thi dự đoán của SV. Sai số tương đối của từng trường hợp được tính theo công thức:

$$\frac{\text{điểm thi dự đoán} - \text{điểm thi thực tế}}{\text{điểm thi thực tế}}$$

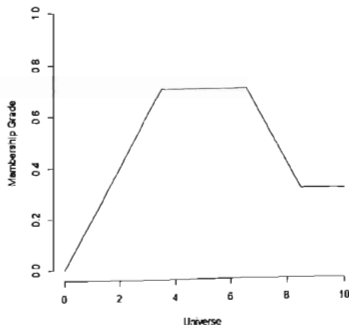
Sai số của mô hình cũng chính là trung bình sai số tương đối của tất cả các trường hợp (xem Hình 4) và được tính theo công thức:

$$\frac{1}{n} \sum_{i=1}^n \frac{|x_i^* - x_i|}{x_i}$$

Trong đó: n: tổng số SV,
 x_i^* : điểm thi dự đoán của SV i,
 x_i : điểm thi thực tế của SV i.

Bảng 4: Một số kết quả dự đoán cùng với sai số tương đối

TT	nghi	giuaki	thi	Dự đoán thi	Sai số tương đối
1	3	6.5	5	5.23	0.046
2	1	6.2	5.8	5.14	0.112



Hình 4: Hàm thành viên mô tả điểm thi của SV nghi 3 slot và có điểm giữa kì 6.5

TT	nghi	giuaki	thi	Dự đoán thi	Sai số tương đối
3	5	5.9	4.8	3.99	0.169
4	5	7	5	4.45	0.110
5	1	5.3	5.8	5.01	0.136
...

2.4. Hạn chế của nghiên cứu và dự kiến nghiên cứu tiếp theo

Việc chỉ đạt được độ chính xác gần 80% cho thấy việc dùng logic mờ chưa thật sự hiệu quả so với đánh giá thông thường của các giáo viên giàu kinh nghiệm đứng lớp. Ngoài ra, việc chọn các khoảng chia mức (Ví dụ, điểm cao là 8 đến 10, thấp là 0 đến 4) cho các biến cũng ảnh hưởng đến độ chính xác của mô hình. Tiến hành thử nhiều hàm thành viên khác nhau tương ứng nhiều khoảng chia khác nhau đòi hỏi nhiều công sức. Trong tương lai, các tác giả sẽ áp dụng thuật toán phát sinh mờ (genetic fuzzy) để tìm khoảng chia tối ưu nhằm giảm sai số dự đoán đến mức thấp nhất.

3. Kết luận

Logic có thể được vận dụng vào bài toán dự đoán kết quả thi kết thúc học phần của SV ngay tại thời điểm SV hoàn thành được nửa thời gian học tập với độ chính xác xấp xỉ 80%. Mô hình được sử dụng tương đối đơn giản với chỉ hai biến đầu vào là điểm giữa kì và số buổi nghỉ học của SV cùng với một số rất ít các luật suy diễn mờ được đưa ra với ý kiến chuyên gia. Từ góc độ sư phạm, giảng viên đứng lớp có thể tham khảo kết quả dự đoán để đưa ra các tác động thích hợp trong quá trình giảng dạy nhằm nâng cao chất lượng dạy và học.

Tài liệu tham khảo

- [1] David Meyer, Kurt Hornik, & Christian Buchta, (2017), Package 'sets', Version 1.0-18, URL: <https://CRAN.R-project.org/package=sets>.
- [2] George J. Klir & Bo Yuan, (1995), *Fuzzy sets and fuzzy logic Theory and Applications*, Prentice Hall PTR.
- [3] L. A. Zadeh, (1965), *Fuzzy sets, Information and Control*, 8, pp.338-353.
- [4] Le Hoang Son - Hamido Fujita, (2019), *Neural-fuzzy with representative sets for prediction of student performance*, Applied Intelligence, Volume 49, Issue 1, pp.172-187.
- [5] Rao D. H., Mangalwede S. R., & Deshmukh V. B, (2017), *Student performance evaluation model based on scoring rubric tool for network analysis subject using fuzzy logic*, International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECOT).
- [6] Ravi Kumar Rathore and J. Jayanthi, (2017), *Student prediction system for placement training using fuzzy inference system*, ICTACT Journal on soft computing, Volume 07, Issue 03.
- [7] Rusli N. M., Ibrahim Z., & Janor R. M, (2008), *Predicting students' academic achievement. Comparison between logistic regression, artificial neural network, and Neuro-fuzzy*, International Symposium on Information Technology
- [8] Timothy, (2010), *Fuzzy Logic with Engineering Applications*, Third Edition, A John Wiley and Sons, Ltd., Publication.
- [9] Yıldız O., Bal A., & Gulsecen S, (2013), *Improved fuzzy modelling to predict the academic performance of distance education students*, The International Review of Research in Open and Distributed Learning, 14(5).

USING FUZZY LOGIC TO PREDICT STUDENTS' MARKS ON FINAL EXAM

Lê Thị Kim Anh¹, Đinh Phước Vinh²

¹ Banking University - Ho Chi Minh City
56 Hoàng Diệu 2, Thủ Đức district,
Ho Chi Minh City, Vietnam
Email: anhlik@buih.edu.vn

² FPT University Ho Chi Minh City
Block E2a-7, D1 Street, Saigon Hi-tech Park,
District 9, Ho Chi Minh City, Vietnam
Email: vnhdtp2@fpt.edu.vn

ABSTRACT: *The goal of this paper is to present a fuzzy rule-based model to predict students' marks on the final exam so that teachers can give appropriate pedagogical guidelines to their students in order to improve the quality of teaching and learning. The input variables of the model are students' marks on mid-term test and the number of absent slots after the first half of a semester. After the fuzzification phase, three-level variables will be put into the fuzzy inference model which has only six rules. Accuracy of the model is calculated by comparing predicted marks and actual marks of all students in the data. Based on data of 86 students studying discrete mathematics at FPT University, the model gave 79.9% accuracy similar to previous research using more variables and more rules.*

KEYWORDS: Prediction; fuzzy logic; evaluation methods; teaching methods; fuzzy inference.