

Tạp chí

# Giáo chức

Việt Nam

HỘI CỤU GIÁO CHỨC VIỆT NAM \* TẠP CHÍ KHOA HỌC \* THE VIETNAM TEACHERS' SCIENTIFIC MAGAZINE

ISSN 1859 - 2902

Năm thứ

**17**

Số

**198**

Tháng 10/2023

Tap chí  
**Giáo chức**  
Việt Nam



THE VIETNAM TEACHERS' SCIENTIFIC MAGAZINE

**Tổng Biên tập**  
PGS.TS. NGUYỄN GIA CẦU  
**Phó Tổng Biên tập**  
PGS.TS. TÔ BÁ TRƯƠNG

**Editor in chief**  
Assoc.Prof.Dr. NGUYEN GIA CAU  
**Assoc.Editor in chief**  
Assoc.Prof.Dr. TO BA TRUONG

**HỘI ĐỒNG BIÊN TẬP**

GS.VS. PHAM MINH HAC (Chủ tịch)

Hội Cựu Giáo chức Việt Nam

GS.TSKH. NGUYỄN MẬU BÀNH

Hội Cựu giáo chức Việt Nam

GS.TS. NGUYỄN THANH HÙNG

Trưởng Đại học Sư phạm Hà Nội

PGS.TS. PHAM MINH HÙNG

Trưởng Đại học Vinh

GS.TSKH. TRẦN VĂN NHUNG

Hội đồng chức danh Giáo sư nhà nước

GS.TS. TRẦN QUỐC THÀNH

Trưởng Đại học Sư phạm Hà Nội

TS. NINH VĂN BÌNH

Trưởng Đại học Sài Gòn

PGS.TS. LÊ VĂN TẤN

Viện Hàn lâm Khoa học Xã hội Việt Nam

GS.TS. THAI VĂN THÀNH

Sở GD&ĐT Nghệ An

TS. PHẠM XUÂN LÝ

Trưởng Đại học Nguyễn Huệ

TS. NGUYỄN THỊ HƯƠNG THÚY

Trưởng Đại học Thủ Dầu Một

PGS.TS. TRẦN ĐÌNH TUẤN

Học viện Chính trị, Bộ Quốc phòng

**EDITORIAL BOARD**

Prof.Acad. PHAM MINH HAC (Chairman)

Vietnam Association of former Teachers

Prof.Dr.Sc. NGUYEN MAU BANH

Vietnam Association of former Teachers

Prof.Dr. NGUYEN THANH HUNG

Hanoi Pedagogical University

Assoc.Prof.Dr. PHAM MINH HUNG

Vinh University

Prof.Dr.Sc. TRAN VAN NHUNG

The State Council for Professor Title of Vietnam

Prof.Dr. TRAN QUOC THANH

Hanoi Pedagogical University

Dr. NINH VAN BINH

Saigon University

Assoc.Prof.Dr. LE VAN TAN

Vietnam Academy of Social and Sciences

Prof.Dr. THAI VAN THANH

Department of Nghe An Education and Training

Dr. PHAM XUAN LY

Nguyen Hue University

Dr. NGUYEN THI HUONG THUY

Thu Dau Mot University

Assoc.Prof.Dr. TRAN DINH TUAN

Political Academy, Ministry of National Defence

**Ban Thư ký:** ĐT: 0837371388

**Ban Biên tập:** ĐT: 0913091153

**Ban Trị sự:** ĐT: 0826 662 268

**Secretariat:** Phone: 0837 371388

**Editorial Board:** Phone: 0913 091 153

**Board of Trustees:** Phone: 0826662268

**Tòa soạn trị sự**

Phòng 201, số 25 Tạ Quang Bửu,

quận Hai Bà Trưng, Hà Nội

Điện thoại: (024) 3868 4680

Email: tapchigiaochuc@yahoo.com.vn

**Tài khoản:** 1100 104010 Ngân hàng

Thương mại cổ phần Sài Gòn - Hà Nội

**Head Office**

201, Home 25 Tạ Quang Bửu,

Hai Bà Trưng, Hà Nội.

Tel: (04). (024) 3868 4680

Email: [tapchigiaochuc@yahoo.com.vn](mailto:tapchigiaochuc@yahoo.com.vn)

**Account No:** 1100 104010

SHB Bank

**Giấy phép xuất bản:** Số 470/GP-BTTTT,

ngày 28/9/2022 của Bộ Thông tin

và Truyền thông

**In và chế bản tại:**

Công ty CP In Công đoàn Việt Nam,

167 Tây Sơn, Đống Đa, Hà Nội.

**Publishing licence,** No. 470/GP-BTTTT,

28/9/2022 by Ministry of Information

and Communication

**Designed and Printed by:**

Vietnam Trade Union Printing JSC.

167 Tây Sơn, Dong Da, Hanoi.

**SỨ MỆNH, MỤC ĐÍCH  
VÀ TÔN CHỈ HOẠT ĐỘNG**

Là tạp chí khoa học của Hội Cựu Giáo chức Việt Nam, *Sứ mệnh của Tạp chí là công bố công trình nghiên cứu của các nhà khoa học, các nhà nghiên cứu, các nhà quản lý, chỉ đạo giáo dục, các thầy cô giáo, nghiên cứu sinh, học viên cao học, học sinh, sinh viên trong các lĩnh vực Chính trị - Tư tưởng; Văn hóa - Xã hội; Khoa học - Công nghệ; Giáo dục và Đào tạo theo chỉ trương định hướng của Đảng, của nhà nước và của ngành. Tạp chí thực hiện chức năng là sự đồng hành về tri thức và niềm tin với đồng đảo bạn đọc. Tạp chí Giáo chức Việt Nam được xuất bản từ năm 2007, với mã số ISSN 1859 - 2902, phát hành định kì ngày 10 hằng tháng.*

Tạp chí Giáo chức Việt Nam luôn không ngừng cải tiến, đổi mới về nội dung và hình thức theo hướng nâng cao tính khoa học, tính chuyên nghiệp phù hợp với sự phát triển của báo chí khoa học hiện đại trong xu thế hội nhập, phát triển.

Để đảm bảo tính chuẩn mực và nâng cao hơn nữa chất lượng khoa học, các bài viết trước khi đăng Tạp chí Giáo chức Việt Nam đều được phân biện kín theo một quy trình chặt chẽ, khách quan bởi các nhà khoa học có uy tín trong nước và quốc tế.

**MISSION, PURPOSE AND PRINCIPLE  
OF OPERATION**

To be the scientific magazine of the Association of the former Vietnamese teachers, its Mission is to publish the research works of the scientists, researchers, managers, educational leaders, teachers, postgraduates, graduate students, students in the fields of Politics - Thought, Socioculture, Science and Technology, Education and Training according to the orientation guidelines of the Party, the State and the Ministry. The magazine's function serves as a companion of knowledge and beliefs to a large number of readers. The Vietnam Teachers Magazine has been published since 2007, with the code ISSN 1859-2902, released periodically on the 10th of every month.

The Vietnam Teachers Magazine is constantly improving and innovating in content and form towards improving its scientificity and professionalism in accordance with the development of modern scientific journalism in the trend of integration and development.

To ensure the standardization and further improve the scientific quality, all articles before being published in the Vietnam Teachers Magazine are critically reviewed according to a strict and objective process by domestic and international reputable scientists.

# MỤC LỤC

## VẤN ĐỀ HIỆN NAY

- Ngô Thị Huyền.** Tư tưởng Hồ Chí Minh về vai trò của trí thức và những gợi mở cho công tác phát triển đội ngũ trí thức ở Đồng Nai trong giai đoạn hiện nay - Ho Chi Minh's Thought on the role of intellectuals and suggestions for the development of intellectuals in Dong Nai in the current period. 1
- Nguyễn Thị Huyền Trang.** Đổi mới phương thức bảo vệ nền tảng tư tưởng của Đảng, đấu tranh phản bác các quan điểm sai trái, thù địch trên không gian mạng - innovating the method of protecting the Party's ideological foundation, fighting against the false and hostile views in cyberspace. 4
- Nguyễn Thị Phương Nga - Bùi Quỳnh Anh.** Nâng cao chất lượng công tác đào tạo bồi dưỡng lý luận chính trị cho đội ngũ cán bộ cơ sở ở tỉnh Hòa Bình hiện nay - Improving the quality of political theory training for the grassroots cadres in Hoa Binh province today. 8
- Thần Văn Thương.** Giảng dạy nội dung công nghiệp hóa, hiện đại hóa ở Việt Nam trong học phần kinh tế chính trị Mác - Lênin cho sinh viên theo tinh thần Nghị quyết Đại hội XIII của Đảng - Teaching the content of industrialization and modernization

## NHÀ TRƯỞNG HIỆN ĐẠI

- Ngô Thị Kim Liên.** Phương thức tổ chức hoạt động lễ hội cho trẻ ở các trường mầm non tại thành phố Thủ Đức Một - The method of organizing festive activities for the children in the preschools in Thu Dau Mot city. 27
- Nguyễn Thị Hoàng Yến - Trần Thị Kim Huệ.** Tích hợp các giải pháp công nghệ trong kế hoạch chiến lược phát triển giáo dục mầm non - Integrating technology solutions in the strategic plan for the preschool education development. 31
- Nguyễn Thị Ngọc Diệp.** Một số biện pháp cho trẻ 5-8 tuổi làm quen chữ cái ở trường mầm non - Some measures to help 5-8 years old children become familiar with letters in the preschool. 34
- Dương Thủy Linh.** Phát triển năng lực đọc cho học sinh lớp 3 theo hướng trải nghiệm - Developing reading ability for the 3rd grade students in the experiential direction. 37
- Bùi Thị Thảo.** Giáo dục kỹ năng sống cho học sinh tiểu học thông qua hoạt động trải nghiệm - Educating life skills for primary school students through experiential activities. 42
- Hoàng Thủy Lan - Trần Bùi Thị Phương - Nguyễn Hoàng Yến.** Quản lý hoạt động "xây dựng trường học hạnh phúc" Góp phần giảm áp lực học tập cho học sinh tiểu học tại Hà Nội - Managing the activity of "building a happy school", contributing to reducing the learning pressure for the primary school students in Hanoi. 45
- Nguyễn Thị Hằng.** Biện pháp giáo dục hòa nhập cho trẻ mù chữ ở trường tiểu học trên địa bàn thành phố Thủ Đức Một tỉnh Bình Dương - Inclusive education measures for the orphans in the primary schools in Thu Dau Mot city, Binh Duong province. 48
- Võ Nguyễn Du - Bùi Thanh Lâm.** Quản lý ứng dụng công nghệ thông tin trong dạy học ở các trường tiểu học thành phố Kon Tum trong bối cảnh hiện nay - Managing the application of information technology in teaching at the primary schools in Kon Tum city in the current context. 53

## NGHIÊN CỨU - KINH NGHIỆM

- Hồ Thị Huyền Thương.** Nghiên cứu và đánh giá một số thuật toán phân lớp trên tập dữ liệu trò chơi Caro - Researching and evaluating some classification algorithms on the Caro game dataset. 88
- Phạm Ngọc Huy.** Nâng cao hiệu quả ứng dụng chuyển đổi số trong đổi mới giảng dạy môn Giáo dục Quốc phòng và An ninh ở các trường đại học hiện nay - Improving the effectiveness of digital transformation applications in teaching innovation in National Defense and Security Education at the universities today. 93
- Bồ Thu Hương, Hàn Thị Hạnh Thuý.** Thực trạng năng lực thực tiễn của đội ngũ cán bộ lãnh đạo xã, phường, thị trấn trên địa bàn tỉnh Tuyên Quang - The reality of practical capacity of the leadership team of the communes, wards and towns in Tuyen Quang province. 97
- Hồ Thị Tuyết Nhung.** Đổi mới nội dung, phương pháp dạy học chế nhạo xã hội khoa học theo hướng phát huy tính tích cực học tập của học viên - innovating the contents, methods in teaching the scientific socialism in the direction of promoting the students' active learning. 99
- Hồ Văn Nguyễn.** Quản lý xây dựng trường tiểu học và trung học cơ sở đạt chuẩn quốc gia trên địa bàn huyện Sa Thầy, tỉnh Kon Tum - Managing the construction of primary and junior high schools meeting the national standards Sa Thay district, Kon Tum province. 93
- Lê Thị Hồng Hạnh, Nguyễn Khắc Hùng.** Nhận thức về hoạt động của tổ chuyên môn theo định hướng nghiên cứu bài học ở một số quốc gia - The awareness of the

in Vietnam in the Marxist-Leninist political economy module for the students according to the Resolution of the 13th Party Congress. 11

- Phan Thị Thanh Hương.** Nét độc đáo và giá trị của Cửu đỉnh Triều Nguyễn trong kho tàng văn hóa Việt Nam - The uniqueness and value of the Nine Caldrons of the Nguyen Dynasty in Vietnam's cultural treasure. 15
- Huỳnh Trọng Nhân.** Tăng cường giáo dục quốc phòng và an ninh cho sinh viên các trường đại học đáp ứng yêu cầu xây dựng và bảo vệ Tổ quốc trong tình hình mới - Strengthening the defense and security education for the university students to meet the requirements of building and protecting the Fatherland in the new situation. 19
- Nguyễn Đức Hoàng - Hoàng Hồng Quang.** Giải pháp bồi dưỡng phẩm chất nhân cách theo tinh thần Nghị quyết 847 của Quân ủy Trung ương cho học viên ở Học viện Chính trị hiện nay - The solutions for fostering the personality qualities according to the spirit of the Resolution 847 of the Central Military Party Committee for the students at the Academy of Politics today. 23

- Nguyễn Đình Nhân.** Vận dụng tiếp cận lịch sử nhằm phát triển năng lực tư duy phân tích cho học sinh trong dạy học chủ đề Di truyền học, Trang học phổ thông - Applying a historical approach to enhance the critical thinking skills for the high school students in teaching the topic of Genetics. 57
- Phạm Thị Bích Thủy.** Các yếu tố ảnh hưởng đến quản lý cảnh quan sư phạm ở các trường trung học cơ sở - The factors affecting the management of pedagogical landscape in the junior high schools. 60
- Phạm Thị Hằng - Hòa Thị Thủy - Phan Thị Xuân.** Một số giải pháp thiết kế các hoạt động dạy và học nhằm phát huy tính tích cực, chủ động của học sinh trong học tập môn Giáo dục Công dân lớp 12 ở trường THPT - Some solutions for designing teaching and learning activities to promote the students' activeness in learning grade 12 Civic education at the high school. 64
- Bồ Thị Hòa.** Giảng dạy môn học Giáo dục chính trị theo hướng phát triển năng lực nghề nghiệp cho sinh viên ngành tiếng Trung Quốc ở Trường Cao đẳng Sư phạm Cao Bằng - Teaching political education subject in the direction of developing professional capacity for the students majoring in Chinese language at Cao Bang College of Education. 68
- Có Thị Ánh Tuyết.** Đánh giá kỹ năng vẽ tay và vẽ máy của sinh viên ngành quy hoạch đô thị Trường Đại học Thủ Đức Một - Evaluating hand drawing and computer drawing skills of the urban planning students at Thu Dau Mot University. 71
- Trần Thị Kim Thanh.** Một số biện pháp nâng cao hiệu quả giảng dạy theo hình thức đào tạo kết hợp Blending trong học phần thống kê toán học tại Trường ĐH Kinh tế - Kỹ thuật Công nghiệp - Some measures to improve the teaching in the form of training combined with Blending in the mathematical statistics module at the University of Economic and Technical Industries. 75

- activities from the professional team following a lesson study orientation in some countries. 97
- Nguyễn Thanh Thủy.** Quản lý văn hoá giáo dục nâng cao nhân cách trong môi trường tiểu học sau đại dịch Covid - 19 - Managing the educational culture to enhance the personality in the primary school environment after the Covid - 19 pandemic. 101
- Nguyễn Thị Hiền.** Đặc điểm của dạy học Ngữ văn theo quan điểm dạy học phân hóa - The characteristics of teaching literature based on the perspective of differentiated teaching. 105
- Nguyễn Thị Nguyệt Trinh.** Sắc thái hoá ngôn ngữ trong tứ tuyệt Thơ mới - Nuanced language in the quatrain of new Poetry. 110
- Nguyễn Thị Tâm.** Vai trò của việc quan sát trực tiếp trong đánh giá hoạt động chơi của trẻ em - The role of direct observations in assessing the children's game activities. 113
- Trần Mỹ Linh.** Tổ chức hoạt động khởi động nhằm huy động tri thức nền trong dạy học đọc hiểu văn bản nghị luận - Organizing warm-up activities to mobilize the background knowledge in teaching reading comprehension of argumentative texts. 117
- Phạm Trà My - Nguyễn Thị Nhá.** Công nghệ Blockchain, hợp đồng thông minh và ứng dụng cho hệ thống chia sẻ xe an toàn - Blockchain technology, smart contracts and the application for safe car sharing systems. 121

# NGHIÊN CỨU VÀ ĐÁNH GIÁ MỘT SỐ THUẬT TOÁN PHÂN LỚP TRÊN TẬP DỮ LIỆU TRÒ CHƠI CARO

Hồ Thị Huyền Thương  
Trường Đại học Vinh  
Email: thuong.dhvinh@gmail.com

**Tóm tắt:** Bài viết này tôi trình bày tìm hiểu một số thuật toán phân lớp trên tập dữ liệu trò chơi Caro nhằm đánh giá chính xác hơn các nguồn thông tin dữ liệu muốn khai thác. Thông qua việc thực hiện các thuật toán phân lớp như Decision Tree, Naïve Bayesian, Neural Network và Support Vector Machines trên tập dữ liệu trò chơi Caro, bài báo tập trung đánh giá các kỹ thuật phân lớp đã nêu nhằm có cái nhìn tổng quan hơn về các kỹ thuật này trong khai phá dữ liệu.

**Từ khóa:** Thuật toán phân lớp, khai phá dữ liệu, trò chơi Caro.

**Nhận bài:** 21/08/2023; **Phản biện:** 25/08/2023; **Duyệt đăng:** 29/08/2023.

## 1. Giới thiệu vấn đề

Trong xã hội phát triển như hiện nay lượng thông tin, dữ liệu thu thập lưu trữ ngày càng nhiều, điều đó chứng tỏ khai phá dữ liệu (data mining) là một bước đi đầy tiềm năng giúp các công ty tập trung vào trích rút những thông tin quan trọng nhất trong dữ liệu mà họ đã thu thập được, từ đó đưa ra những quyết định quan trọng mang tính chiến lược. Một trong những kỹ thuật phổ biến trong khai phá dữ liệu là phân lớp (classification), được dùng để dự đoán hay phân loại các trường hợp dữ liệu. Phân lớp dùng để chỉ định các đối tượng vào các lớp hoặc các đối tượng khác nhau. Mục tiêu là đưa ra dự đoán chính xác cho từng trường hợp trong dữ liệu. Phân lớp được ứng dụng nhiều trong phân tích mô hình kinh doanh, tiếp thị, tin dụng, mô hình y sinh học, chuẩn đoán y khoa, phân tích tính hiệu quả trong điều trị bệnh, ...

Trong bài nghiên cứu này tôi sẽ trình bày một số nội dung liên quan đến như: Tìm hiểu tổng quan về khai phá dữ liệu, cách giải quyết bài toán phân lớp và các ứng dụng của bài toán phân lớp; Nghiên cứu một số thuật toán phân lớp trong khai phá dữ liệu, chẳng hạn như phân lớp dựa trên cây quyết định, xác suất có điều kiện, mạng Neural và thuật toán Support Vector Machines; Nghiên cứu và xây dựng tập mẫu từ trò chơi Caro để thử nghiệm các thuật toán phân lớp đã nêu; Thử nghiệm các thuật toán phân lớp trên phần mềm khai phá dữ liệu để từ đó đánh giá, so sánh các kết quả của các thuật toán đã trình bày.

## 2. Một số thuật toán phân lớp

### 2.1. Phân lớp dựa trên Decision tree

Xét  $D$  là tập dữ liệu huấn luyện có thuộc tính phân lớp  $C$ , trong đó

$$C = \{C_1, C_2, \dots, C_m\}$$

$D = D_1 \cup D_2 \cup \dots \cup D_t$  là phân hoạch trên  $D$ , với  $D_i \cap D_j = \emptyset$ .

Mục đích:

- Tìm kiếm độ đo đánh giá mức độ đồng nhất của các đối tượng dựa trên thuộc tính phân lớp.

- Tìm ra phân hoạch của  $D$  có mức độ đồng nhất cực đại.

- Một số độ đo phổ biến Information gain, GINI index, ... đều dựa trên Entropy.

Entropy của tập dữ liệu là lượng thông tin cần để phân loại một phần tử trong tập dữ liệu huấn luyện  $D$ , ký hiệu:  $Info(D)$ .

Entropy của tập dữ liệu ứng với thuộc tính là lượng thông tin cần để phân loại một phần tử trong tập dữ liệu  $D$  dựa trên thuộc tính  $A$ , ký hiệu  $Info_A(D)$ .

Độ lợi thông tin (Information Gain): tối thiểu hoá lượng thông tin cần thiết để phân lớp các mẫu dữ liệu. Độ lợi thông tin ứng với thuộc tính  $A$ , ký hiệu  $Gain(A)$ , là độ sai lệch giữa Entropy ban đầu của tập dữ liệu (trước khi phân hoạch) và Entropy của dữ liệu với thuộc tính  $A$  (sau khi phân hoạch bởi  $A$ ) và được tính như sau:

$$Gain(A) = Info(D) - Info_A(D)$$

### 2.2. Phân lớp dựa trên Naïve Bayesian

Chúng ta có thể ứng dụng thuật toán phân lớp dựa trên xác suất có điều kiện (Naïve Bayes Classification-NBC) để tính tỷ lệ xác suất với rất nhiều các dạng bài toán khác nhau, với dữ liệu càng nhiều thì độ chính xác của thuật toán sẽ càng cao, và khi dữ liệu thay đổi thì kết quả cũng thay đổi theo.

- Mỗi mẫu dữ liệu được biểu diễn bằng  $X = (x_1, x_2, \dots, x_n)$  tương ứng với các thuộc tính  $A_1, A_2, \dots, A_n$ .

- Xét các lớp  $C_1, C_2, \dots, C_m$ . Cho trước mẫu chưa biết  $X$ , thuật toán gán  $X$  vào  $C_j$  nếu  $P(C_j|X) > P(C_l|X)$  với  $1 \leq j \leq m, j \neq l$ . Chúng ta có cực đại  $P(C_j|X)$ . Lớp  $C_j$  có  $P(C_j|X)$  là cực đại được gọi là giả thuyết hậu nghiệm cực đại.

- Dựa theo công thức Bayes, ta có công thức sau:

$$P(C_j|X) = \frac{P(X|C_j)P(C_j)}{P(X)}$$

Để phân lớp mẫu chưa biết  $X$ , ta tính  $P(X|C_j)P(C_j)$

cho từng  $C_j$ . Sau đó mẫu  $X$  được gán vào  $C_j$  nếu  $P(C_j|X) > P(C_l|X)$  với mọi  $1 \leq l \leq m, j \neq l$ .

### 2.3. Phân lớp dựa trên thuật toán Support Vector Machines

Để phân lớp kỹ thuật SVM nguyên thủy sẽ chia không gian dữ liệu thành 2 phần và quá trình này lặp lại nhiều lần. Khi đó hàm quyết định phân dữ liệu vào lớp thứ  $i$  của tập  $n$ .

Bài toán phân nhiều lớp sử dụng phương pháp SVM hoàn toàn có thể thực hiện giống như bài toán hai lớp. Bằng cách sử dụng chiến lược "một-đối-một".

Giả sử bài toán cần phân loại có  $k$  lớp ( $k > 2$ ), chiến lược "một-đối-một" sẽ tiến hành  $k(k-1)/2$  lần phân lớp nhị phân sử dụng phương pháp SVM. Mỗi lớp sẽ tiến hành phân tách với  $k-1$  lớp còn lại để xác định  $k-1$  hàm phân tách dựa vào bài toán phân hai lớp bằng phương pháp SVM.

#### 2.4. Phân lớp dựa trên mạng Neural

Mạng neural nhân tạo có các thành phần cơ bản sau:

- Đầu vào: là tập các tín hiệu của neural, các tín hiệu này thường được đưa vào dưới dạng một vector  $N$  chiều ( $x_1, x_2, \dots, x_N$ ).

- Liên kết: mỗi liên kết được thể hiện bởi một trọng số gọi là trọng số liên kết, ký hiệu  $W_{ki}$  (trọng số liên kết giữa tín hiệu vào thứ  $i$  với neural  $k$ ). Thông thường, các trọng số này được khởi tạo một cách ngẫu nhiên ở thời điểm khởi tạo mạng và được cập nhật liên tục trong suốt quá trình học.

- Hàm tổng (Summing function): thường được dùng để tính tổng của tích các đầu vào với trọng số liên kết của nó.

- Ngưỡng (bias): thường được đưa vào như một thành phần của hàm tuyến.

- Hàm truyền (transfer function): hàm này được dùng để giới hạn phạm vi đầu ra của mỗi neural. Nó nhận đầu vào là kết quả của hàm tổng và ngưỡng đã cho trước. Phạm vi đầu ra của mỗi neural được giới hạn trong đoạn  $[0, 1]$  hoặc  $[-1, 1]$ . Các hàm truyền có thể là hàm tuyến tính hoặc phi tuyến. Việc lựa chọn hàm tùy thuộc vào từng lớp bài toán và kinh nghiệm.

- Đầu ra (output): là tín hiệu đầu ra của một neural, với mỗi neural sẽ có tối đa một đầu ra.

### 3. Đánh giá các thuật toán phân lớp trên tập dữ liệu trò chơi caro

Caro là một trò chơi phổ biến giữa hai người chơi, kết quả có thể một trong hai người thắng hoặc hòa. Ở đây ta nghiên cứu về khả năng phân lớp của một số thuật toán trên tập dữ liệu trò chơi Caro.

#### 3.1. Phân lớp dữ liệu dựa trên WEKA

WEKA là một phần mềm được viết bằng ngôn ngữ lập trình Java. Đây là một công cụ rất mạnh trong xây dựng các mô hình khai phá dữ liệu, cho phép triển khai hầu hết các kỹ thuật khai phá dữ liệu như phân lớp, phân cụm, khai phá luật kết hợp,... Các chức năng chính của WEKA Explorer thể hiện trong các thẻ của màn hình chính, bao gồm:

- Preprocess (tiền xử lý dữ liệu): chức năng này cho phép mở, lưu, điều chỉnh tập tin dữ liệu và các thuật toán áp dụng cho bước tiền xử lý dữ liệu.

- Classify (phân lớp): Chức năng này cung cấp các mô hình phân lớp dữ liệu, người dùng có thể chọn các thuật toán phân lớp như cây quyết định, mạng Neural, xác suất có điều kiện,... để phân lớp.

- Cluster (phân cụm): Chức năng này cung cấp các mô hình phân cụm như thuật toán k-means,...

- Associate (luật kết hợp): Chức năng này cho phép khai thác tập phổ biến và luật kết hợp nhờ vào các thuật toán Apriori, FPGrowth,...

- Select attributes (lựa chọn thuộc tính): Cho phép lựa chọn các thuộc tính thích hợp nhất trong tập dữ liệu theo tiêu chí nào đó.

- Visualize: Biểu diễn dữ liệu dưới dạng biểu đồ.

#### 3.2. Các phương pháp đánh giá

Về cơ bản, quá trình phân lớp dữ liệu bao gồm hai bước như sau:

- Bước 1 (Học - learning): bước này sẽ xây dựng mô hình mô tả tập các lớp dữ liệu hay khái niệm xác định trước. Đầu vào của quá trình này là tập dữ liệu có cấu trúc được mô tả bằng các thuộc tính, chẳng hạn các thuộc tính của trò chơi Caro đã đề cập. Trong tập dữ liệu này, mỗi phần tử dữ liệu đã được gán thuộc về một lớp xác định trước. Đầu ra của bước này là tập các quy tắc phân lớp (mô hình) dưới dạng if-then, cây quyết định, công thức logic, mạng neural.

- Bước 2 (Phân lớp - Classification): Dùng mô hình đã xây dựng ở Bước trước để phân lớp dữ liệu mới.

Trong Bước 1, ta cần phải đánh giá được độ chính xác mang tính chất dự đoán cho từng mô hình phân lớp (các quy tắc phân lớp). Độ chính xác của mô hình trên tập dữ liệu kiểm tra là tỉ lệ phần trăm các mẫu trong tập dữ liệu kiểm tra được mô hình phân lớp đúng (so với thực tế). Có nhiều phương pháp chia tập dữ liệu để đánh giá. Người ta thường sử dụng tập dữ liệu kiểm tra độc lập với tập dữ liệu huấn luyện. Nếu độ chính xác của mô hình có thể chấp nhận được, thì mô hình được sử dụng để phân lớp cho những dữ liệu khác, hoặc những dữ liệu mà giá trị thuộc tính phân lớp chưa biết hay còn thiếu.

Bài báo sử dụng một số phương pháp đánh giá được hỗ trợ trong WEKA như sau:

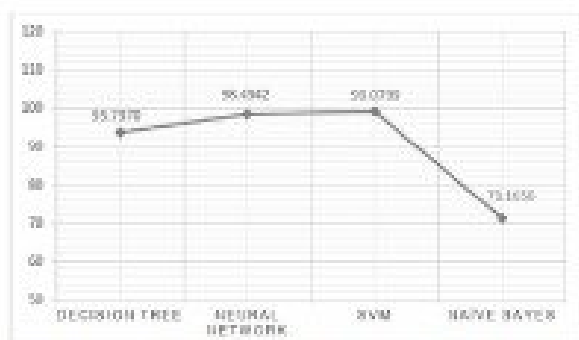
- Phương pháp đánh giá dựa vào tập dữ liệu huấn luyện.

- Phương pháp chia phần trăm.

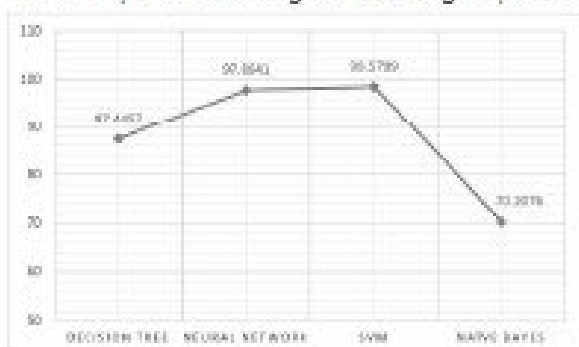
- Phương pháp xác nhận chéo.

#### 3.3. Kết quả thực nghiệm

Kết quả thực nghiệm sẽ so sánh độ chính xác khi phân lớp dựa vào các thuật toán: Decision Tree, Neural Network, Support Vector Machine và Naïve Bayes. Sau khi thực hiện các phương pháp đánh giá dựa vào tập huấn luyện, chia phần trăm và xác nhận chéo với mỗi thuật toán đã đề cập trên phần mềm WEKA.



Hình 1. Độ chính xác lớn nhất của từng thuật toán.



Bảng 1 mô tả độ chính xác sau khi sử dụng phần mềm WEKA dùng các phương pháp đánh giá dựa vào tập huấn luyện, chia phần trăm và xác nhận chéo của từng thuật toán.

Bảng 1. Độ chính xác của các thuật toán phân lớp

| Thuật toán phân lớp (các thiết lập)  | Phương pháp thực hiện            | Độ chính xác (%) |
|--|----------------------------------|------------------|
| Decision Tree J4.8 (Tree → J48)  | Training Set                     | 93,7370          |
|  | 10-fold Cross-Validation         | 84,5511          |
|  | Percentage split (Holdout) (66%) | 84,0491          |
| Neural Network (Functions → MultilayerPerceptron, hiddenLayers=2, trainingTime=50) | Training Set                     | 98,4342          |
|  | 10-fold Cross-Validation         | 98,2255          |
|  | Percentage split (Holdout) (66%) | 96,9325          |
| Support Vector Machine (Functions → SMO)   | Training Set                     | 99,3299          |
|  | 10-fold Cross-Validation         | 99,3299          |
|  | Percentage split (Holdout) (66%) | 99,0799          |
| Naive Bayes (Bayes → NaiveBayes)   | Training Set                     | 69,8330          |
|  | 10-fold Cross-Validation         | 69,6242          |
|  | Percentage split (Holdout) (66%) | 71,1656          |

#### 4. Kết luận

Từ kết quả nghiên cứu lý thuyết và thực nghiệm, chúng tôi đã thu nhận được những kiến thức hữu ích thuộc lĩnh vực khai phá dữ liệu ứng dụng trong trò chơi. Các kết quả thử nghiệm thể hiện ở Hình 1, 2 và bảng 1 cho thấy rằng các thuật toán phân loại có thể đánh giá thành công các trò chơi đã kết thúc và người mới có thể tiến hành đánh giá một cách chính xác. Một số thuật toán cho độ chính xác của việc phân lớp xấp xỉ 99%. Tuy nhiên có một số thuật toán cho độ chính xác thấp hơn, khoảng 70%. Điều này nói lên rằng, với độ chính xác lên đến 99%, chúng ta có thể ứng dụng khai phá dữ liệu trong trò chơi Caro. Các câu hỏi nghiên cứu đã được trả lời, cụ thể hơn, các thuật toán phân lớp trong học máy có thể phân loại thành công trò chơi Caro và theo kết quả thực nghiệm Thuật toán phân lớp SVM cho kết quả tốt nhất trong việc phân loại trò chơi Caro trong các thuật toán đã đề cập. □

#### Tài liệu tham khảo

- [1]. Nguyễn Hà Nam, Nguyễn Tri Thành, Hà Quang Thụy, *Giáo trình khai phá dữ liệu*, NXB Đại học Quốc gia Hà Nội, 2013.
- [2]. Đỗ Phúc, *Giáo trình khai phá dữ liệu*, NXB Đại học quốc gia TP. Hồ Chí Minh, 2017.
- [3]. Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining: Concepts and Techniques*, 3rd ed, Morgan-Kaufmann Publishers, 2012.

### RESEARCHING AND EVALUATING SOME CLASSIFICATION ALGORITHMS ON THE CARO GAME DATASET

Ho Thi Huyen Thuong

Vinh University

Email: thuong.dhvinh@gmail.com

**Abstract:** The article explores some classification algorithms on the Caro game dataset in order to more accurately evaluate the data sources to be exploited. Through the implementation of classification algorithms such as Decision Trees, Naive Bayesian, Neural Network and Support Vector Machines on the Caro game dataset, the article focuses on evaluating the aforementioned classification techniques to get an overview about these techniques in data mining.

**Keywords:** Classification algorithm, data mining, Caro game.