

ỨNG DỤNG THUẬT TOÁN TỐI ƯU TIẾN HÓA BẦY ĐÀN MỜ TRONG PHÂN TÍCH NHU CẦU KHÁCH HÀNG

Nguyễn Thị Như Na^{a*}

^aKhoa Tự nhiên, Trường Cao đẳng Sư phạm Điện Biên, Lai Châu, Việt Nam

Lịch sử bài báo

Nhận ngày 31 tháng 03 năm 2017 | Chính sửa ngày 21 tháng 04 năm 2017

Chấp nhận đăng ngày 19 tháng 05 năm 2017

Tóm tắt

Bài báo này ứng dụng thuật toán tối ưu tiến hóa bầy đàn mờ cho bài toán phân tích nhu cầu khách hàng. Đây là bài toán có ý nghĩa ứng dụng lớn trong hoạt động sản xuất kinh doanh. Áp dụng thuật toán tối ưu tiến hóa bầy đàn mờ vào bài toán cụ thể là một công ty chuyên cung cấp thiết bị y tế của Mỹ muốn phân tích nhu cầu 500 bệnh viện trong khu vực về các thiết bị và vật tư y tế, hỗ trợ công ty đưa ra chiến lược kinh doanh phù hợp nhất với từng bệnh viện để đạt doanh thu cao.

Từ khóa: Bầy đàn; Di truyền; Phân cụm; Tập mờ; Tiến hóa; Tối ưu.

1. ĐẶT VẤN ĐỀ

Phân tích nhu cầu khách hàng là một việc rất quan trọng trong kinh doanh, phân tích nhu cầu khách hàng, phân khúc thị trường, xác định nhu cầu biến động của khách hàng... Các thông tin thu được sử dụng để hỗ trợ các doanh nghiệp đưa ra chiến lược kinh doanh hiệu quả. Cụ thể, bài toán thực tế phân tích nhu cầu khách hàng là một công ty chuyên cung cấp các thiết bị y tế cho 500 bệnh viện (<http://www.rci.rutgers.edu/~cabrera/sc/cs8/cs8.html>) ở Mỹ muốn tìm giải pháp để tăng doanh số bán hàng. Phân tích nhu cầu tiêu thụ các thiết bị y tế của các bệnh viện ở Mỹ tìm ra bệnh viện có mức tiêu thụ loại thiết bị y tế nào cao nhằm đưa ra chiến lược kinh doanh phù hợp với từng bệnh viện để tăng hiệu quả kinh doanh. Dữ liệu đầu vào bài toán là 500 bệnh viện ở Mỹ được tổ chức trong tập tin *Customer.xls* bao gồm 19 trường và 4000 bản ghi.

Dữ liệu đầu ra bài toán là phân tích 500 bệnh viện thành 3 nhóm với mức tiêu thụ

* Tác giả liên hệ: Email: nhuna.cdsp@gmail.com

thiết bị y tế khác nhau: Bệnh viện có mức tiêu thụ thiết bị y tế thấp; Bệnh viện có mức tiêu thụ thiết bị y tế trung bình; Bệnh viện có mức tiêu thụ thiết bị y tế cao.

19 trường trong tập tin bao gồm các mã như sau:

1. ZIP: Mã bưu điện
2. HID: ID bệnh viện
3. CITY: Tên thành phố
4. STATE: Tên tiểu bang
5. BEDS: Số giường bệnh
6. RBEDS: Số giường chỉnh hình
7. OUT-V: số lượt khám ngoại trú
8. ADM: Chi phí hành chính (1000 \$/năm)
9. SIR: Thu từ nội trú
10. SALESY: Bán trang thiết bị phục hồi chức năng từ ngày 1 tháng 1
11. SALES12: Bán trang thiết bị phục hồi chức năng cuối tháng 12
12. HIP95: Số hoạt động cho hông trong năm 1995
13. KNEE95: Số hoạt động đầu gối trong năm 1995
14. TH: Có hoạt động dạy học không? 0, 1
15. TRAUMA: Có chấn thương không? 0, 1
16. REHAB: Có chỉnh hình không? 0, 1
17. HIP96: Số hoạt động cho hông cho năm 1996
18. KNEE96: Số hoạt động đầu gối cho năm 1996
19. FEMUR96: Số hoạt động cho xương đùi cho năm 1996.

Có nhiều cách tiếp cận để giải quyết bài toán phân tích nhu cầu khách hàng như là cách tiếp cận thống trị dựa trên tập rõ (DRSA) được phát triển bởi Greco, Matarazzo, và Slowinski (1998) và Greco, Matarazzo, và Slowinski (2000) rất hữu ích để giảm dữ

liệu trong phân tích định tính. Cách tiếp cận phân tích giỏ hàng (MBA) của Giudici và Passerone (2002) xác định mối liên hệ giữa khách hàng và các sản phẩm khác nhau trong một đơn vị đặc biệt như là bên trong một siêu thị. Các dữ liệu được phân tích trong MBA thường bao gồm tất cả các giao dịch mua hàng được thực hiện trong một khoảng thời gian nhất định để phân tích cấu trúc kết hợp giữa việc bán sản phẩm khác nhau có sẵn dựa vào đó doanh nghiệp lên kế hoạch cho các chính sách tiếp thị tốt hơn. Cách tiếp cận cho bài toán phân tích nhu cầu khách hàng mua bán trên Internet (Song, Kyeong, & Kim, 2001) sử dụng độ đo tương tự và độ đo khác nhau cho các thay đổi của khách hàng, sau đó đánh giá độ đo thay đổi để đưa ra quy tắc thay đổi giúp các nhà quản lý đưa ra chiến lược kinh doanh phù hợp cho từng đối tượng khách hàng dựa trên số liệu kinh doanh cụ thể.

Phân cụm được coi như một công cụ độc lập để xem xét phân bố dữ liệu, làm bước tiền xử lý cho các thuật toán khác. Phân cụm ứng dụng rất nhiều trong các lĩnh vực như phân tích hình ảnh (Pappas, 1992), thông tin địa lý (Aksoy, 2006), khai phá web (Runkler & Bezdek, 2003) ... Phương pháp phân cụm mờ là sự kết hợp của kỹ thuật phân cụm với lý thuyết mờ của Zadeh (1965) đang phát triển và được ứng dụng rộng rãi trong thực tiễn, ví dụ như phân tích rủi ro, dự báo nguy cơ phá sản cho ngân hàng và nhiều bài toán khác. Nhưng những vấn đề được quan tâm nhiều vẫn là nâng cao chất lượng phân cụm (Chen & Ludwig, 2014), tính toán thông qua một số độ đo chất lượng cụ thể... Và một số nghiên cứu ứng dụng thuật toán tối ưu tiến hóa phân cụm mờ như là nghiên cứu phân cụm mờ bằng PSO (Runkler & Katz, 2006) giải quyết bài toán phân cụm mờ bằng cực tiểu hóa mô hình *Fuzzy C – Means* (FCM), ứng dụng trên bộ dữ liệu bệnh ung thư phổi. Nghiên cứu về vấn đề phát hiện, định vị trực quan các tín hiệu không cố định của Biswal, Dash, và Panigrahi (2009). Các nghiên cứu dựa trên tối ưu tiến hóa mờ để giải bài toán phân tích nhu cầu khách hàng như trên là rất ít.

Bài báo này sẽ áp dụng thuật toán tối ưu tiến hóa mờ do Pang, Wang, Zhou, và Dong (2004) đề xuất cho bài toán phân cụm mờ để xác định nghiệm tối ưu toàn cục cho bài toán phân tích nhu cầu khách hàng. Với dữ liệu thực tế một công ty chuyên cung cấp các thiết bị y tế Mỹ muốn phân tích nhu cầu khách hàng là 500 bệnh viện trong khu vực để có kế hoạch và chiến lược kinh doanh phù hợp nhất cho từng đối tượng khách hàng, thích ứng với nhu cầu khách hàng mà đạt được doanh thu cao. Thuật toán tối ưu hóa bầy

đàn mờ (FPSO) (Pang và ctg., 2004) dựa trên thuật toán tối ưu bầy đàn cho bài toán người bán hàng. Thuật toán tối ưu bầy đàn (PSO) được Eberhart và Kenneday (1995) giới thiệu thuộc về lớp các bài toán tiến hóa, dựa trên khái niệm trí tuệ bầy đàn để giải bài toán tối ưu tiến hóa. PSO được áp dụng rộng rãi để cải tiến hiệu suất các thuật toán khác. Các ứng dụng như là bài toán lập kế hoạch (Weijun, Zhiming, Wei, & Genke, 2004), người bán hàng (Wang, Huang, Zhou, & Pang, 2003).

Bài báo này ứng dụng thuật toán tối ưu tiến hóa phân cụm mờ cho công ty ở Mỹ chuyên cung cấp thiết bị y tế cho 500 bệnh viện khu vực, muốn phân tích nhu cầu về vật tư và thiết bị y tế của các bệnh viện trên. Kết quả thực nghiệm thu được 3 cụm khách hàng với các mức tiêu thụ vật tư và thiết bị y tế khác nhau, dựa vào đó công ty có cơ sở khoa học đưa ra giải pháp chiến lược kinh doanh phù hợp nhất cho từng nhóm đối tượng và từng đối tượng khách hàng làm tăng hiệu quả hoạt động kinh doanh. Ngoài ra, ứng dụng của thuật toán này có thể mở rộng không chỉ cho công ty chuyên cung cấp các thiết bị y tế mà còn có thể ứng dụng cho các doanh nghiệp kinh doanh sản xuất các mặt hàng khác.

2. THUẬT TOÁN TỐI ƯU TIẾN HÓA CHO PHÂN CỤM MỜ

2.1. Phương pháp phân cụm mờ

Bài toán phân cụm N vector $X = \{x_1, x_2, \dots, x_N\}$ thành c cụm dựa trên tính toán tối thiểu hóa hàm mục tiêu để đo chất lượng của cụm và tìm tâm cụm sao cho hàm độ đo không tương tự là nhỏ nhất. Một phân cụm mờ vector $X = \{x_1, x_2, \dots, x_N\}$ được biểu diễn bởi ma trận $U = [U_{ki}]_{N \times c}$ sao cho một điểm dữ liệu có thể thuộc về nhiều nhóm và được xác định bằng giá trị hàm thuộc u . Ma trận giá trị hàm thuộc có dạng như sau:

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1c} \\ \vdots & \ddots & \vdots \\ u_{N1} & \cdots & u_{Nc} \end{bmatrix} \quad (1)$$

Thuật toán phân cụm mờ đã được xuất phát từ việc cực tiểu giá trị hàm mục tiêu:

$$J_m = \sum_{k=1}^N \sum_{j=1}^c u_{kj}^m d(x_k, z_j) \quad (2)$$

Trong đó: $d(x_k, z_j)$ là một độ đo không tương tự.

Giải bài toán $J_m(u, z) \rightarrow \min$ với ràng buộc sau:

$$\begin{cases} 0 \leq u_{kj} \leq 1 \\ \sum_{j=1}^c u_{kj} = 1 \\ 0 \leq \sum_{k=1}^N u_{kj} \leq N \end{cases} \quad (3)$$

2.2. Áp dụng thuật toán tối ưu bầy đàn cho phân cụm mờ (FPSO)

Pang và ctg. (2004) đã đề xuất thuật toán tối ưu bầy đàn mờ. Trong thuật toán này vị trí và vận tốc của các cá thể xác định lại tương ứng với các biến mờ. Phương pháp này mô tả cho bài toán phân cụm mờ.

Với X là vị trí các cá thể, thể hiện mối quan hệ mờ của các đối tượng dữ liệu. Tâm cụm $Z = \{z_1, z_2, \dots, z_c\}$, X được biểu diễn như sau:

$$X = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{N1} & \cdots & \mu_{Nc} \end{bmatrix} \quad (4)$$

Trong đó: μ_{ij} là một hàm thuộc của cá thể i thuộc cụm j :

$$\begin{aligned} \mu_{ij} &\in [0, 1], \forall i = 1, 2, \dots, N; \forall j = 1, 2, \dots, c \\ \sum_{j=1}^c \mu_{ij} &= 1, \forall i = 1, 2, \dots, N \end{aligned} \quad (5)$$

Ma trận vị trí của mỗi cá thể giống ma trận mờ μ trong thuật toán FCM. Ngoài ra, vận tốc của mỗi cá thể là ma trận N dòng và c cột, các cá thể của ma trận trong phạm vi $[-1, 1]$.

$$V = \begin{bmatrix} v_{11} & \cdots & v_{1c} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{Nc} \end{bmatrix} \quad (6)$$

Ta có công thức (7) cập nhật vị trí và vận tốc của các cá thể:

$$\begin{aligned} V(t+1) &= w.V(t) + (c_1 r_1)(pbest(t) - X(t)) + (c_2 r_2)(gbest(t) - X(t)) \\ X(t+1) &= X(t) \oplus V(t+1) \end{aligned} \quad (7)$$

Sau khi cập nhật ma trận vị trí của cá thể, có thể không thỏa mãn điều kiện như trong (8). Vì vậy phải chuẩn hóa ma trận vị trí.

$$\begin{aligned} \mu_{ij} &\in [0,1], \forall i = 1, 2, \dots, N; \forall j = 1, 2, \dots, c \\ \sum_{j=1}^c \mu_{ij} &= 1, \forall i = 1, 2, \dots, N \end{aligned} \quad (8)$$

Để chuẩn hóa ta chuyển các số âm trong ma trận thành 0. Nếu các cá thể trên một hàng của ma trận là 0, chúng cần chuẩn hóa lại bằng cách lấy ngẫu nhiên trong $[0,1]$, ta có ma trận được chuẩn hóa được biểu diễn như trong (9).

$$X_{normal} = \begin{bmatrix} \mu_{11}/\sum_{j=1}^c \mu_{1j} & \cdots & \mu_{1c}/\sum_{j=1}^c \mu_{1j} \\ \vdots & \ddots & \vdots \\ \mu_{N1}/\sum_{j=1}^c \mu_{Nj} & \cdots & \mu_{Nc}/\sum_{j=1}^c \mu_{Nj} \end{bmatrix} \quad (9)$$

Thuật toán FPSO cũng tương tự như các thuật toán cải tiến khác, đều cần một hàm đánh giá kết quả tổng quát gọi là hàm độ đo thích nghi. Công thức (10) là hàm độ đo thích nghi sử dụng để đánh giá các kết quả.

$$f(X) = \frac{K}{J_m} \quad (10)$$

Trong đó K là hằng số; J_m là hàm mục tiêu $J_m(\mu, Z) = \sum_{i=1}^N \sum_{j=1}^c \mu_{ij}^m \|x_i - z_j\|^2$

Từ các công thức trên, ta có thuật toán FPSO như sau:

Bước 1. Khởi tạo tham số P, c_1, c_2, w và tham số đếm vòng lặp tối đa $Maxiter$.

Bước 2. Tạo P cá thể trong quần thể ($X, pbest, gbest, V$ ma trận $N \times c$ cá thể).

Bước 3. Khởi tạo $X, V, pbest$ cho mỗi cá thể và $gbest$ cho quần thể.

Bước 4. Tính các tâm cụm cho mỗi cá thể bằng công thức (11):

$$z_j = \frac{\sum_{i=1}^N \mu_{ij}^m x_i}{\sum_{i=1}^N \mu_{ij}^m} \quad (11)$$

Bước 5. Tính giá trị hàm độ đo thích nghi cho mỗi cá thể bằng công thức (10).

Bước 6. Tính $pbest$ cho mỗi cá thể.

Bước 7. Tính $gbest$ cho quần thể.

Bước 8. Cập nhật ma trận vận tốc cho mỗi cá thể bằng công thức (9).

Bước 9. Cập nhật ma trận vị trí cho mỗi cá thể bằng công thức (7).

Bước 10. Nếu thỏa gặp điều kiện kết thúc, quay lại Bước 4.

Điều kiện kết thúc là số lần lặp tối đa hoặc giá trị $gbest$ trong vòng lặp không cải thiện nữa.

2.3. Ưu điểm và nhược điểm của bài toán

Thuật toán FPSO là sự kết hợp ưu điểm của thuật toán FCM, dễ dàng giải quyết các bài toán tối ưu hàm mục tiêu khác nhau của thuật toán PSO để phân cụm trong môi trường mờ. Hiệu suất tốt hơn so với một số thuật toán phân cụm mờ như FCM. Sử dụng FPSO dễ dàng giải quyết các bài toán phân cụm mờ, tối ưu hóa tổ hợp khó giải quyết trong phạm vi lớn trong môi trường mờ (Mehdizadeh & Moghaddam, 2008). Tuy nhiên, nhược điểm là thuật toán FPSO hội tụ chậm hơn so với thuật toán phân cụm mờ FCM.

3. KẾT QUẢ THỰC NGHIỆM

Thuật toán chạy với tham số cụm $c=3$ với nhãn các cụm như sau: Bệnh viện có mức tiêu thiết bị y tế thấp, trung bình, cao. Tham số mờ $m=2$, tiêu chuẩn hội tụ (*epsilon*) $Eps = 0.01$, và số lần lặp tối đa $MaxTest = 1000$. Quy tắc: Khách hàng sẽ thuộc về cụm mà có độ thuộc lớn nhất. Nếu có hai độ thuộc lớn nhất bằng nhau trở lên thì chọn một trong số các cụm đó để đưa vào.

Cụm 1. Gồm những khách hàng là những bệnh viện có mã bệnh viện ID và độ thuộc: ID(37011) với độ thuộc (0.3723); ID(55511) và độ thuộc (0.3728); ID(24016) và độ thuộc (0.3698); ID(2016) và độ thuộc (0.3763); ID(53016) và độ thuộc (0.3764)... cụ thể ở Bảng 1.

Cụm 2. Gồm những khách hàng là những bệnh viện có mã bệnh viện ID và độ thuộc: ID(1104) với độ thuộc (0.4234); ID(1602) với độ thuộc (0.4517); ID(1605) với độ thuộc (0.2738); ID(1801) với độ thuộc (0.4082); ID(1830) với độ thuộc (0.4662); ID(1841) với độ thuộc (0.4303)... cụ thể ở Bảng 2.

Cụm 3. Gồm những khách hàng là những bệnh viện có mã bệnh viện ID và độ thuộc: ID(43014) với độ thuộc (0.4406); ID(21514) độ thuộc (0.4469); ID(46014) với độ thuộc (0.4402); ID(62014) với độ thuộc (0.4859); ID(78014) với độ thuộc (0.4938); ID(20901) với độ thuộc (0.48844); ID(153014) với độ thuộc (0.4874); ID(174014) với độ thuộc (0.4083)... cụ thể ở Bảng 3.

Xây dựng hàm thuộc cho các biến ngôn ngữ. Trước hết xác định biến ngôn ngữ:

$$RBEDS = \{RBEDS \text{ thấp}; RBEDS \text{ cao}\}$$

$$OUT_V = \{OUT_V \text{ thấp}; OUT_V \text{ cao}\}$$

$$ADM = \{ADM \text{ thấp}; ADM \text{ cao}\}$$

$$SIR = \{SIR \text{ thấp}; SIR \text{ cao}\}$$

$$SALESY = \{SALESY \text{ thấp}; SALESY \text{ cao}\}$$

$$SALES12 = \{SALES12 \text{ thấp}; SALES12 \text{ cao}\}$$

$$HIP95 = \{HIP95 \text{ thấp}; HIP95 \text{ cao}\}$$

$$KNEE95 = \{KNEE95 \text{ thấp}; KNEE95 \text{ cao}\}$$

$$HIP96 = \{HIP96 \text{ thấp}; HIP96 \text{ cao}\}$$

$$KNEE96 = \{KNEE96 \text{ thấp}; KNEE96 \text{ cao}\}$$

$$FEMUR96 = \{FEMUR96 \text{ thấp}; FEMUR96 \text{ cao}\}$$

Không gian tham chiếu của các thuộc tính được thể hiện ở Bảng 4.

Xác định hàm thuộc cho các biến ngôn ngữ: Với biến X, và không gian tham chiếu $U = [0, u]$. Khi đó: X cao $(x) = x/u; 0 \leq x \leq u$, X thấp $(x) = 1 - x/u$. Ví dụ: Với thuộc tính BEDS: $0 \leq BEDS \leq 1476$; với $BEDS = 1001$ thì BEDS cao $(1001) = 1001/ 1476 = 0.678184$; BEDS thấp $= 1 - 0.678184$. Tương tự với các thuộc tính còn lại.

Xác định luật mờ: (Theo MISO - với n đầu vào và 1 đầu ra)

- *Luật 1.* Nếu BEDS cao và RBEDS cao và OUT_V cao và REHAB cao thì tiêu thụ mức cao.
- *Luật 2.* Nếu RBED và OUT_V và SIR cao và HIP95 cao thì tiêu thụ mức cao.

Bảng 1. Bệnh viện thuộc Cụm 1

ID	BEDS	RBEDS	OUT_V	ADM	SIR	SALESY	SALES	HIP	KNEE	TH	TRAUMA	REHAB	HIP	KNEE	FEMUR
							12	95	95				96	96	96
37011	250	12	107297	9203	5503	1	1	66	46	1	0	1	91	47	62
5551	236	25	263189	7891	6748	3	3	66	56	1	0	1	65	70	52
24016	155	10	53398	7530	6426	1	1	38	24	1	0	1	54	39	57
2016	475	13	63656	17424	10940	0	0	154	112	1	0	1	147	124	137
53016	320	18	48556	11935	11185	8	16	126	47	1	0	1	145	80	127

Bảng 2. Bệnh viện thuộc Cụm 2

ID	BEDS	RBEDS	OUT -V	ADM	SIR	SALESY	SALES 12	HIP 95	KNEE 95	TH	TRAUMA	REHAB	HIP 96	KNEE 96	FEMUR 96
1104	324	0	95702	10406	13648	28	29	122	94	0	0	0	152	95	116
1602	80	80	0	907	0	318	491	0	0	0	0	1	0	0	0
1605	211	0	9275	7571	15766	277	513	209	143	1	0	0	207	178	177
1801	82	82	0	440	0	50	82	0	0	0	0	1	0	0	0
1830	151	0	25550	6199	5257	12	16	40	15	0	0	0	39	19	87
1841	290	0	92768	8662	5388	56	96	55	26	0	0	0	49	40	80

Bảng 3. Bệnh viện thuộc Cụm 3

ID	BEDS	RBEDS	OUT -V	ADM	SIR	SALESY	SALES 12	HIP 95	KNEE 95	TH	TRAUMA	REHAB	HIP 96	KNEE 96	FEMUR 96
43014	1005	0	504125	34214	27830	0	2	504	271	1	1	0	519	289	303
21514	710	0	421456	37745	25796	0	0	461	458	1	1	0	437	372	169
46014	219	0	47036	7829	9163	6	18	663	463	1	0	0	618	410	51
62014	391	0	0	14594	8852	7	13	82	49	1	1	0	80	47	77
78014	290	0	72061	11767	7426	0	0	78	51	1	0	0	101	53	103
209014	275	0	118791	8788	6473	1	15	51	30	1	0	0	52	32	68
153014	321	0	66061	17475	9375	0	0	104	93	1	1	0	102	60	109

Bảng 4. Không gian tham chiếu của các thuộc tính

	BEDS	RBEDS	OUT -V	ADM	SIR	SALESY	SALES 12	HIP 95	KNEE 95	TH	TRAUMA	REHAB	HIP 96	KNEE 96	FEMUR 96
Min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Max	1476	850	1987000	66440	70300	1209	2770	1421	868	1	1	1	137 3	1081	489

- *Luật 3.* Nếu BEDS cao và RBEDS thấp và HIP95 cao và KNEE95 cao thì tiêu thụ mức trung bình.
- *Luật 4.* Nếu BEDS thấp và RBEDS cao và HIP95 thấp và KNEE95 thấp thì tiêu thụ mức thấp.
- *Luật 5.* Nếu BEDS cao và RBEDS thấp và HIP96 cao và KNEE96 cao thì tiêu thụ mức trung bình.
- *Luật 6.* Nếu BEDS thấp và RBEDS cao và HIP96 thấp và KNEE96 thấp thì tiêu thụ mức thấp.
- *Luật 7.* Nếu BEDS thấp và RBEDS cao và OUT_Vcao và REHAB thấp thì tiêu thụ mức thấp.
- *Luật 8.* Nếu BEDS thấp và RBEDS thấp và TH cao và REHAB cao thì tiêu thụ ở mức trung bình.
- *Luật 9.* Nếu BEDS cao và RBEDS thấp và TH cao và REHAB cao thì tiêu thụ ở mức cao.
- *Luật 10.* Nếu BEDS cao và RBEDS thấp và HIP96 cao và KNEE96 cao thì tiêu thụ ở mức cao.
- *Luật 11.* Nếu BEDS thấp và RBEDS thấp và SALELY cao và SALES12 cao thì tiêu thụ ở mức thấp.
- *Luật 12.* Nếu BEDS cao và RBEDS cao và SALELY cao và SALES12 cao thì tiêu thụ ở mức cao.

Qua tính toán với số liệu cụ thể ở các bảng trên cho thấy:

Cụm 1 là nhóm các bệnh viện có các chỉ số RBEDS (số giường chỉnh hình) cao và chỉ số REHAB (có chỉnh hình) thường xuyên, có chỉ số TH (có hoạt động dạy học) thường xuyên. Các chỉ số khác của các thuộc tính ở mức trung bình. Cụm 1 là cụm các bệnh viện có mức tiêu thụ thiết bị y tế mức trung bình.

Cụm 2 là nhóm các bệnh viện có chỉ số SALELY (bán trang thiết bị phục hồi chức năng ngày 1 tháng 1) và chỉ số SALES12 (bán trang thiết bị phục hồi chức năng cuối tháng 12) cao hơn. Các chỉ số khác của các thuộc tính ở mức thấp. Cụm 2 là cụm các bệnh viện có mức tiêu thụ thiết bị y tế mức thấp.

Cụm 3 là nhóm các bệnh viện có chỉ số ADM (chi phí hành chính), SIR (thu từ nội trú), HIP95 (số hoạt động cho hông năm 1995), KNEE95 (số hoạt động cho đầu gối năm 1995), TH (có hoạt động dạy học), HIP96 (số hoạt động cho hông năm 1996), KNEE96 (số hoạt động cho đầu gối năm 1996), FEMUR96 (số hoạt động cho xương đùi năm 1996) cao hơn. Cụm 3 là cụm các bệnh viện có mức tiêu thụ thiết bị y tế mức cao.

Vậy để tăng doanh thu cho công ty kinh doanh các thiết bị y tế, công ty sẽ tập trung vào 3 nhóm khách hàng với những chiến lược khác nhau cho từng nhóm. Với nhóm 1, công ty tập trung vào cung cấp các thiết bị liên quan đến hoạt động chỉnh hình và các thiết bị phục vụ cho quá trình học tập và đào tạo chuyên môn. Với nhóm 2, công ty tập trung cung cấp các trang thiết bị liên quan đến phục hồi chức năng. Với nhóm 3, có chi phí hành chính, thu từ nội trú khá cao, các chỉ số khác tương đối cao so với hai nhóm trên, có thể nói nhóm 3 là nhóm khách hàng tiềm năng có khả năng tiêu thụ nhiều thiết bị y tế phục vụ cho các hoạt động chữa bệnh như: Hoạt động cho hông, hoạt động cho đầu gối, hoạt động cho xương đùi, và các trang thiết bị liên quan đến hoạt động học tập và đào tạo chuyên môn. Cung cấp và phân phối các thiết bị y tế phù hợp cho từng nhóm ở trên, công ty kinh doanh sẽ nhanh chóng thúc đẩy quá trình bán hàng và từ đó doanh thu sẽ tăng cao hơn.

4. KẾT LUẬN

Bài toán phân tích nhu cầu khách hàng là bài toán có tính ứng dụng cao, có ý nghĩa quan trọng trong hoạt động sản xuất kinh doanh. Bài toán phân tích nhu cầu khách hàng đã được giải quyết trên thuật toán tối ưu tiến hóa phân cụm mờ.

Thuật toán tối ưu tiến hóa mờ được phát triển từ thuật toán tối ưu tiến hóa trên môi trường mờ đã được ứng dụng trong bài toán phân tích nhu cầu khách hàng cụ thể là của một công ty chuyên cung cấp các thiết bị y tế cho 500 bệnh viện khu vực. Công ty

muốn phân tích nhu cầu tiêu thụ khác nhau về thiết bị và vật tư y tế của các bệnh viện. 500 bệnh viện được phân thành 3 cụm với các mức tiêu thụ vật tư và thiết bị y tế khác nhau: Bệnh viện có mức tiêu thụ thiết bị y tế thấp; Bệnh viện có mức tiêu thụ thiết bị y tế trung bình; Bệnh viện có mức tiêu thụ thiết bị y tế cao. Từ đó, công ty đã có cơ sở để đưa ra chiến lược kinh doanh phù hợp nhất đáp ứng nhu cầu từng nhóm bệnh viện và từng bệnh viện cụ thể để hoạt động kinh doanh đạt hiệu quả và lợi nhuận cao. Hướng phát triển tiếp theo, chúng tôi sẽ áp dụng thuật toán trên trong nhiều lĩnh vực, như là hỗ trợ ra quyết định, thông tin địa lý, xử lý ảnh, v.v.

TÀI LIỆU THAM KHẢO

- Angeline, P. J. (1995). *Adaptive and self-adaptive evolutionary computations*. Paper presented at The Computational Intelligence: A Dynamic Systems Perspective, USA.
- Aksoy, E. (2006). *Clustering with GIS: An attempt to classify Turkish district data*. Paper presented at The XXIII FIG Congress, Germany.
- Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Berlin, Germany: Springer Science & Business Media.
- Biswal, B., Dash, P. K., & Panigrahi, B. K. (2009). Power quality disturbance classification using fuzzy C-means algorithm and adaptive particle swarm optimization. *IEEE Transactions on Industrial Electronics*, 56(1), 212-220.
- Bùi, C. C., & Nguyễn, D. P. (2006). *Hệ mờ mạng nơron và ứng dụng*. Hà Nội, Việt Nam: NXB Khoa học và Kỹ thuật.
- Brimberg, J., Hansen, P., Mladenović, N., & Taillard, E. D. (2000). Improvements and comparison of heuristics for solving the uncapacitated multisource Weber problem. *Operations Research*, 48(3), 444-460.
- Chen, M., & Ludwig, S. A. (2014). Particle swarm optimization based fuzzy clustering approach to identify optimal number of clusters. *Journal of Artificial Intelligence and Soft Computing Research*, 4(1), 43-56.
- Du, M. O., Hansen, P., Jaumard, B., & Mladenovic, N. (1999). An interior point algorithm for minimum sum-of-squares clustering. *SIAM Journal on Scientific Computing*, 21(4), 1485-1505.
- Eberhart, R., & Kennedy, J. (1995). *A new optimizer using particle swarm theory*. Paper presented at The Sixth International Symposium on IEEE, USA.
- Fischetti, M., Lancia, G., & Serafini, P. (2002). Exact algorithms for minimum routing cost trees. *Networks*, 39(3), 161-173.
- Giudici, P., & Passerone, G. (2002). Data mining of association structures to model consumer behavior. *Computational Statistics & Data Analysis*, 38(4), 533-541.

- Greco, S., Matarazzo, B., & Slowinski, R. (1998). *A new rough set approach to evaluation of bankruptcy risk*. Paper presented at The Operational Tools in the Management of Financial Risks, USA.
- Greco, S., Matarazzo, B., & Slowinski, R. (2000). Extension of the rough set approach to multicriteria decision support. *INFOR: Information Systems and Operational Research*, 38(3), 161-195.
- Hall, L. O., Ozyurt, I. B., & Bezdek, J. C. (1999). Clustering with a genetically optimized approach. *IEEE Transactions on Evolutionary Computation*, 3(2), 103-112.
- Hansen, P., & Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical Programming*, 79(1-3), 191-215.
- Hansen, P., & Mladenović, N. (1999). *An introduction to variable neighborhood search*. Paper presented at Meta-heuristics, USA.
- Mehdizadeh, E., & Moghaddam, R. T. (2008). Fuzzy particle swarm optimization algorithm for a supplier clustering problem. *Journal of Industrial Engineering*, 1(1), 17-24.
- Mladenović, N., & Hansen, P. (1997). Variable neighborhood search. *Computers & Operations Research*, 24(11), 1097-1100.
- Pappas, T. N. (1992). An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Signal Processing*, 40(4), 901-914.
- Pang, W., Wang, K. P., Zhou, C. G., & Dong, L. J. (2004). *Fuzzy discrete particle swarm optimization for solving traveling salesman problem*. Paper presented at Computer and Information Technology Conference, USA.
- Richter, F. (2016). *Identität, Ethnizität und Nationalismus in Kurdistan*. Berlin, Germany: LIT Verlag Münster.
- Runkler, T. A., & Bezdek, J. C. (2003). Web mining with relational clustering. *International Journal of Approximate Reasoning*, 32(2-3), 217-236.
- Runkler, T. A., & Katz, C. (2006). *Fuzzy clustering by particle swarm optimization*. Paper presented at Fuzzy Systems, 2006 IEEE International Conference on IEEE, USA.
- Salerno, J. (1997). *Using the particle swarm optimization technique to train a recurrent neural model*. Paper presented at Tools with Artificial Intelligence, The Ninth IEEE International Conference, USA.
- Song, H. S., Kyeong, K. J., & Kim, S. H. (2001). Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications*, 21(3), 157-168.
- Stefanowski, J. (2009). *Data mining-clustering* [PowerPoint slides]. Retrieved from <http://www.cs.put.poznan.pl/jstefanowski/sed/DM-7clusteringnew.pdf>.
- Wang, K. P., Huang, L., Zhou, C. G., & Pang, W. (2003). *Particle swarm optimization for traveling salesman problem*. Paper presented at Machine Learning and Cybernetics, 2003 International Conference on IEEE, USA.

Weijun, X., Zhiming, W., Wei, Z., & Genke, Y. (2004). *A new hybrid optimization algorithm for the job-shop scheduling problem*. Paper presented at American Control Conference IEEE, USA.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.

AN APPLICATION OF FUZZY PARTICLE SWARM OPTIMIZATION FOR CUSTOMER ANALYSIS

Nguyen Thi Nhu Na^{a*}

^a*The Faculty of Natural Sciences, Dienbien College of Education, Laichau, Vietnam*

^{*}*Corresponding author: Email: nhuna.cdsp@gmail.com*

Article history

Received: March 31st, 2017 | Received in revised form: April 21st, 2017

Accepted: May 19th, 2017

Abstract

This article presents the application of the Fuzzy Particle Swarm Optimization algorithm for analyzing customer needs. Applying Fuzzy Particle Swarm Optimization algorithm to the problem of a US medical device supplier wanting to analyze the needs of 500 hospitals, in the region in terms of, medical equipment and supplies, the needs analysis assisted. The supplier offering the most suitable business strategies for each hospital to achieve higher revenues.

Keywords: Clustering; Fuzzy set; Genetic; Optimal; Swarm.
