

ỨNG DỤNG PHƯƠNG PHÁP KHAI PHÁ DỮ LIỆU TRONG QUẢN LÝ NĂNG LƯỢNG ĐIỆN

■ Lư Nhật Vinh ¹
Cao Quang Lộc ²

TÓM TẮT

Trong bối cảnh cơ sở dữ liệu sử dụng năng lượng điện ở nước ta hiện nay chỉ đáp ứng cho mục đích lưu trữ và thống kê ở mức thông thường thì việc quản lý, thống kê, khai thác dữ liệu tiêu thụ điện một cách hiệu quả hơn là yêu cầu rất cần thiết. Một trong những phương pháp đó là ứng dụng logic mờ vào cơ sở dữ liệu sử dụng năng lượng điện và dùng các phương pháp khai phá dữ liệu để khai thác các thông tin tiềm ẩn bên trong dữ liệu.

ABSTRACT

Application of data exploitation method in electricity management

In the present situation, database of electricity consumption in Vietnam is used only for the usual purpose of statistics storage, so that there is a need for more effective management, compilation and exploitation of this database. One of the methods is the application of fuzzy logic in data mining and exploitation for hidden information in this database of electricity consumption.

1. Giới thiệu

Cơ sở dữ liệu sử dụng điện với các con số tiêu thụ điện trên hóa đơn thu tiền khách hàng mỗi tháng chỉ thể hiện số lượng tiêu thụ điện nhưng không đánh giá được mức độ tiêu thụ điện.

Để biểu diễn và đánh giá mức độ tiêu thụ điện, chúng ta xây dựng 6 mức tiêu thụ điện theo từng ngành nghề và biến ngôn ngữ logic mờ tiêu thụ. Trên cơ sở biến ngôn ngữ tiêu thụ chúng ta xây dựng mức sử dụng.

Các mức tiêu thụ điện theo ngành nghề có thể được xác định qua kinh nghiệm của người sử dụng. Tuy nhiên, chúng ta có thể sử dụng các phương pháp gom cụm dữ liệu (clustering) hỗ trợ xác định các mức tiêu thụ một cách khách quan theo tình hình sử dụng điện hiện tại.

Trên cơ sở các mức sử dụng, chúng ta dùng luật kết hợp (association rules) để khai thác các luật sử dụng điện.

2. Mức tiêu thụ

Mức tiêu thụ là các chỉ số tiêu thụ điện được áp dụng cho từng ngành nghề theo từng thời điểm (ngày áp dụng) để làm cơ sở đánh giá và tính toán mức sử dụng điện.

Có sáu mức tiêu thụ từ mức 1 đến mức 6:

- Mức 1: Mức rất ít

¹TS, Khoa Kỹ thuật – Công nghệ, Đại học Văn Hiến

²ThS, Công ty GEMADEPT

- Mức 2: Mức ít
- Mức 3: Mức vừa
- Mức 4: Mức nhiều
- Mức 5: Mức rất nhiều
- Mức 6: Mức rất rất nhiều

Gọi $m_1, m_2, m_3, m_4, m_5, m_6$ lần lượt là các mức tiêu thụ mức 1, mức 2, mức 3, mức 4, mức 5, mức 6 của ngành nghề n trong thời điểm t ; gọi M là số tiêu thụ điện lớn nhất của ngành nghề n trong thời điểm t , ta có:

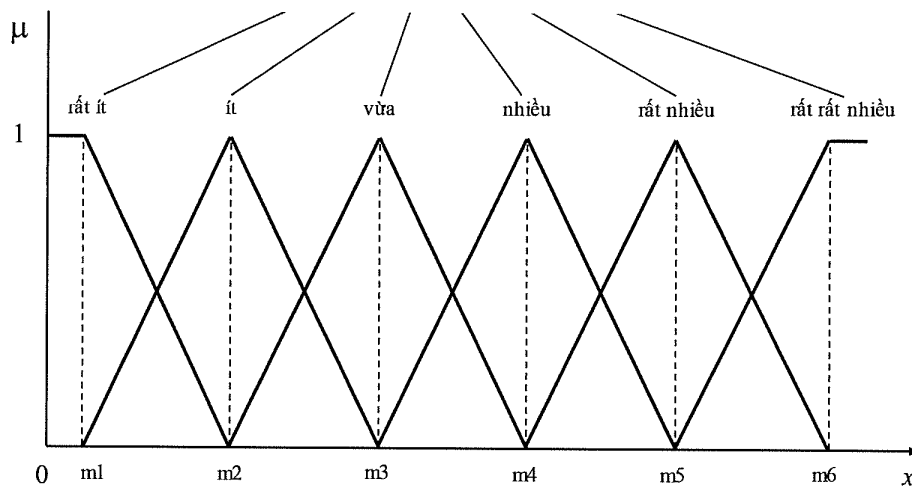
$$0 \leq m_1 \leq m_2 \leq m_3 \leq m_4 \leq m_5 \leq m_6 \leq M$$

3. Biến ngôn ngữ tiêu thụ

Biến ngôn ngữ tiêu thụ là số logic mờ mô tả mức độ tiêu thụ điện. Các giá trị của nó gồm có: *rất ít, ít, vừa, nhiều, rất nhiều, rất rất nhiều.*

Biến ngôn ngữ tiêu thụ được mô tả trên tập $U = [0, 1, \dots, n]$ hoặc biến điều khiển x (biến cơ sở) thay thế tiêu thụ theo số tiêu thụ điện bằng các tam giác và các phần của hình thang để chỉ rõ các thuật ngữ rất ít, ít, vừa, nhiều, rất nhiều, rất rất nhiều.

Các thuật ngữ của biến ngôn ngữ được biểu diễn như sau:



Các hàm ... của biến ngôn ngữ tiêu thụ ...

$$\mu_{\text{rất ít}}(x) = \begin{cases} 1 & \text{với } 0 \leq x \leq m_1 \\ \frac{m_2 - x}{m_2 - m_1} & \text{với } m_1 \leq x \leq m_2 \end{cases} \quad (2)$$

$$\mu_{\text{ít}}(x) = \begin{cases} \frac{x - m_1}{m_2 - m_1} & \text{với } m_1 \leq x \leq m_2 \\ \frac{m_3 - x}{m_3 - m_2} & \text{với } m_2 \leq x \leq m_3 \end{cases} \quad (3)$$

$$\mu_{\text{vừa}}(x) = \begin{cases} \frac{x - m_2}{m_3 - m_2} & \text{với } m_2 \leq x \leq m_3 \\ \frac{m_4 - x}{m_4 - m_3} & \text{với } m_3 \leq x \leq m_4 \end{cases} \quad (4)$$

$$\mu_{\text{nhiều}}(x) = \begin{cases} \frac{x - m_3}{m_4 - m_3} & \text{với } m_3 \leq x \leq m_4 \end{cases} \quad (5)$$

$$\mu_{\text{rất nhiều}}(x) = \begin{cases} \frac{m_5 - x}{m_5 - m_4} & \text{với } m_4 \leq x \leq m_5 \end{cases} \quad (6)$$

$$\mu_{\text{rất rất nhiều}}(x) = \begin{cases} \frac{m_6 - x}{m_6 - m_5} & \text{với } m_5 \leq x \leq m_6 \\ \frac{x - m_5}{m_6 - m_5} & \text{với } m_5 \leq x \leq m_6 \\ 1 & \text{với } x \geq m_6 \end{cases} \quad (7)$$

4. Mức sử dụng

Mức sử dụng là một giá trị biểu diễn mức độ sử dụng điện.

Cho tập $X = [0, 1, \dots, n]$, tập $Y = [1.0, 1.1, \dots, 6.0]$, mức sử dụng là một ánh xạ f từ tập X vào tập Y .

$$f: X \rightarrow Y$$

$$y = f(x)$$

Mức sử dụng được tính toán dựa trên giá trị của các hàm thành viên của biến ngôn ngữ *tiêu thụ*.

Gọi x là số tiêu thụ điện, $f(x)$ là mức sử dụng, dùng các công thức (2) - (7) để tính các hàm thành viên theo x .

Gọi μ_1 là $\mu_{\text{rất ít}}(x)$, μ_2 là $\mu_{\text{ít}}(x)$, μ_3 là $\mu_{\text{vừa}}(x)$,
 μ_4 là $\mu_{\text{nhiều}}(x)$, μ_5 là $\mu_{\text{rất nhiều}}(x)$, μ_6 là $\mu_{\text{rất rất nhiều}}(x)$

ta có:

$$f(x) = \begin{cases} 1 + \mu_2 & \text{với } (\mu_1 > 0 \text{ và } \mu_2 > 0) \text{ hoặc } \mu_1 = 1 \\ 2 + \mu_3 & \text{với } (\mu_2 > 0 \text{ và } \mu_3 > 0) \text{ hoặc } \mu_2 = 1 \\ 3 + \mu_4 & \text{với } (\mu_3 > 0 \text{ và } \mu_4 > 0) \text{ hoặc } \mu_3 = 1 \\ 4 + \mu_5 & \text{với } (\mu_4 > 0 \text{ và } \mu_5 > 0) \text{ hoặc } \mu_4 = 1 \\ 5 + \mu_6 & \text{với } (\mu_5 > 0 \text{ và } \mu_6 > 0) \text{ hoặc } \mu_5 = 1 \text{ hoặc } \mu_6 = 1 \end{cases} \quad (8)$$

5. Gom cụm xác định các mức tiêu thụ

Mức tiêu thụ điện theo ngành nghề có thể được xác định qua kinh nghiệm của người sử dụng. Tuy nhiên, chúng ta có thể sử dụng các phương pháp gom cụm dữ liệu hỗ trợ xác định các mức tiêu thụ một cách khách quan theo tình hình sử dụng điện hiện tại.

5.1. Độ đo khoảng cách

Số tiêu thụ điện của hóa đơn là một số nguyên $x \in [0, 1, \dots, n]$. Ta sử dụng độ đo khoảng cách Euclide để tính khoảng cách giữa các hóa đơn.

Cho i, j là hai hóa đơn trong CSDL sử dụng điện, gọi t_i, t_j lần lượt là số tiêu thụ điện của hóa đơn i, j , khoảng cách $d(i, j)$ giữa hóa đơn i và hóa đơn j được xác định như sau:

$$d(i, j) = \sqrt{|t_i - t_j|^2} = |t_i - t_j| \quad (9)$$

5.2. Thuật toán gom cụm

Mục đích của việc gom cụm là xác định các mức tiêu thụ theo từng ngành nghề áp dụng cho một khoảng thời gian nào đó, tức là xác định giá trị các trọng tâm của cụm theo từng ngành nghề trong khoảng thời gian tương ứng.

Người ta có thể thử nghiệm nhiều thuật toán gom cụm khác nhau trên dữ liệu hóa đơn để chọn ra thuật toán hiệu quả và phù hợp với đặc tính của loại dữ liệu này. Ở đây, chúng ta sử dụng *thuật toán k-means* để thực hiện.

Thuật toán k-means xác định các mức tiêu thụ

Đầu vào của thuật toán:

Số lượng cụm k cần phân hoạch và không gian dữ liệu có n hóa đơn.

Đầu ra của thuật toán:

k cụm hóa đơn và giá trị của k trọng tâm tương ứng.

Thuật toán gồm bốn bước:

1. Chọn ngẫu nhiên k hóa đơn làm trọng tâm ban đầu của k cụm.
2. Gán hoặc gán lại từng hóa đơn vào cụm có trọng tâm gần hóa đơn đang xét nhất. Độ đo khoảng cách được tính theo công thức (9). Nếu không có phép gán lại nào thì dừng thuật toán.
3. Tính lại trọng tâm cho từng cụm: Giá trị của trọng tâm là trung bình cộng giá trị tiêu thụ điện của tất cả hóa đơn trong cụm.
4. Quay lại bước 2.

6. Khai thác dữ liệu sử dụng điện bằng luật kết hợp

6.1. Mức sử dụng chẵn

Mức sử dụng chẵn là giá trị làm tròn từ *mức sử dụng*. Mức sử dụng chẵn phục vụ cho việc khai thác dữ liệu bằng luật kết hợp.

Gọi f là mức sử dụng, m là mức sử dụng chuẩn, ta có:

$$m = \begin{cases} 1 & \text{với } 1 \leq f \leq 1.4 \\ 2 & \text{với } 1.5 \leq f \leq 2.4 \\ 3 & \text{với } 2.5 \leq f \leq 3.4 \\ 4 & \text{với } 3.5 \leq f \leq 4.4 \\ 5 & \text{với } 4.5 \leq f \leq 5.4 \\ 6 & \text{với } 5.5 \leq f \leq 6 \end{cases} \quad (10)$$

5

6.2. Luật sử dụng điện cần khai thác

Cho Ngữ cảnh khai thác dữ liệu (NCKTDL) sử dụng điện (O, I, R) và ngưỡng $minsupp \in [0,1]$, ngưỡng $minconf \in [0,1]$. Cho S là tập phổ biến theo ngưỡng $minsupp$ và luật kết hợp hợp lệ $r: X \rightarrow Y$ được tạo ra từ S .

r là luật kết hợp cần khai thác nếu r có 3 dạng sau:

1. $Nnld_i, Dvhcd_j \rightarrow MucSuDung_h$: Khách hàng thuộc ngành nghề i và sống ở đơn vị hành chính j sử dụng điện ở mức h
2. $Nnld_i \rightarrow MucSuDung_h$: Khách hàng thuộc ngành nghề i sử dụng điện ở mức h
3. $Dvhcd_j \rightarrow MucSuDung_h$: Khách hàng sống ở đơn vị hành chính j sử dụng điện ở mức h

Vì vậy, sau khi khai thác được danh sách luật kết hợp, chúng ta chỉ giữ lại những luật có ba dạng cần khai thác.

6.3. Áp dụng thuật toán khai thác luật kết hợp

Chúng ta dùng các thuật toán của lý thuyết luật kết hợp để xây dựng một công cụ khai thác luật kết hợp tổng quát cho mọi NCKTDL (O, I, R) với ngưỡng $minsupp \in [0,1]$ và ngưỡng $minconf \in [0,1]$. Đầu vào của công cụ khai thác luật kết hợp tổng quát là ma trận nhị phân biểu diễn NCKTDL (O, I, R) , ngưỡng $minsupp$ và $minconf$.

Để khai thác dữ liệu sử dụng điện ta thực hiện các bước sau:

1. Lấy tập dữ liệu cần khai thác theo NCKTDL sử dụng điện.
2. Tạo ma trận nhị phân biểu diễn NCKTDL.
3. Dùng công cụ khai thác luật kết hợp tổng quát để tìm các luật kết hợp.
4. Chuyển đổi các luật kết hợp dạng nhị phân thành các luật kết hợp sử dụng điện.
5. Giữ lại các luật kết hợp cần khai thác.

7. Cài đặt thử nghiệm

Chúng tôi đã tiến hành cài đặt thử nghiệm trên môi trường Microsoft SQL Server 2008 và Microsoft Visual C# 2008. Dữ liệu thử nghiệm được dùng từ dữ liệu sử dụng điện năm 2008 của Công ty Điện Lực TP.HCM.

Sau khi định nghĩa 6 mức tiêu thụ điện cho tất cả các ngành nghề và cập nhật các mức sử dụng mờ vào CSDL, hệ thống có khả năng tìm kiếm, thống kê, đánh giá dữ liệu theo mức độ sử dụng điện.

Thông kê dữ liệu

Dữ liệu cần thống kê

Từ ngày: 01/01/2008 Đến ngày: 31/12/2008

Ngành nghề*: 4401 | Sinh hoạt tiêu dùng dân cư Thành thị

- 4201 - Các Đại sứ quán, tổ chức LHQ, CQBD nước ngoài [65]
- 4300 - II. Văn phòng doanh nghiệp [66]
- 4400 - III. Sinh hoạt tiêu dùng dân cư [67]
- 4401 - 1. Sinh hoạt tiêu dùng dân cư Thành thị [68]
- 4402 - 2. Sinh hoạt tiêu dùng dân cư Nông thôn, miền núi... [69]
- 4600 - IV. Tổ chức quốc tế [70]
- 5000 - E. HOẠT ĐỘNG KHÁC [71]
- 5100 - I. Hoạt động văn hoá trường học... [72]
- 5101 - 1. Nhà hát, rạp chiếu bóng, rạp xiếc [73]
- 5102 - 2. Câu lạc bộ văn hoá, vui chơi... [74]
- 5103 - 3. Trường học [75]
- 5104 - 4. Bệnh viện, trạm xá [76]

Chọn tất cả | Bỏ tất cả | Số ngành nghề được chọn: 1 / 87

Đơn vị hành chính*

- 701 - TP.HCM [1]
- 70101 - Quận 1 TP.HCM [2]
- 7010101 - Phường Tân Định Quận 1 [3]
- 7010103 - Phường Đa Kao Quận 1 [4]
- 7010105 - Phường Bến Nghé Quận 1 [5]
- 7010107 - Phường Bến Thành Quận 1 [6]
- 7010109 - Phường Nguyễn Thái Bình Quận 1 [7]
- 7010111 - Phường Phạm Ngũ Lão Quận 1 [8]
- 7010113 - Phường Cầu Ông Lãnh Quận 1 [9]
- 7010115 - Phường Cô Giang Quận 1 [10]
- 7010117 - Phường Nguyễn Cư Trinh Quận 1 [11]
- 7010119 - Phường Cầu Kho Quận 1 [12]
- 70103 - Quận 2 TP.HCM [313]

Chọn tất cả | Bỏ tất cả | Số đơn vị hành chính được chọn: 1 / 319

Tiêu chí thống kê

Mức sử dụng từ 3.5 đến 4.4 [Nhiều] [Nhiều]

Lấy 50 Ngành nghề Đơn vị hành chính Khách hàng [Sử dụng nhiều nhất]

Thống kê dữ liệu

Kết quả thống kê

Mã KH	Tên khách hàng	Số quyền	Số hóa đơn	Ngày phát hành	Số tiêu thụ
PE01000002513	LA THI TAM	01A4460C1	0130246	05/03/2008	7,307
PE01000002704	TRIEU QUANG HONG	01A6600-1	0129030	05/03/2008	3,175
PE01000002746	PHAN THI THU SUONG	01A7020C1	0130307	05/03/2008	8,362
PE01000002750	NGUYEN T MINH TRANG	01A6840-1	0129220	05/03/2008	5,866
PE01000002876	PHU DONG	01A4900C1	0130261	05/03/2008	8,008
PE01000002941	NGUYEN BICH THU	01A4720-1	0111949	05/03/2008	2,781

Dòng: 50 / 160 [Tạo ra bởi: ADMIN 15/05/2009 10:34:11 PM] [Sửa đổi bởi: ADMIN 22/09/2009 2:35:02 PM]

Khách hàng: PE01000003628
MSD SLHD STT
4 2 6,077
Tổng số hóa đơn và tiêu thụ: 2 và 6,077

Khách hàng: PE01000018291
MSD SLHD STT
4 3 10,382
Tổng số hóa đơn và tiêu thụ: 3 và 10,382

Khách hàng: PE01000018298
MSD SLHD STT
4 1 4,749
Tổng số Hóa đơn và Tiêu thụ: 1 và 4,749

Thời gian thực hiện: 0 giờ 0 phút 1 giây

Chức năng khai thác dữ liệu ngoài khả năng tìm ra các *luật sử dụng điện*, nó còn cho phép trả lời các câu hỏi liên quan đến tình hình sử dụng điện theo nhu cầu của người sử dụng.

Khai thác dữ liệu

Dữ liệu cần khai thác

Từ ngày: 01/01/2008 Đến ngày: 31/12/2008

Ngành nghề*

- 2000 - B. CÔNG NGHIỆP, XÂY DỰNG [9]
- 2102 - 2. Khai thác dầu thô, khí tự nhiên [12]
- 2105 - 5. Khai thác cát, đá, sỏi [15]
- 2201 - 1. Chế biến thực phẩm [17]
- 2202 - 2. Sản xuất đồ uống [18]
- 2203 - 3. SX thuốc lá, thuốc láo [19]
- 2204 - 4. SX sợi, dệt [20]
- 2205 - 5. SX trang phục, nhuộm [21]
- 2206 - 6. Thuốc, sơ chế da, giấy dáp [22]
- 2207 - 7. Chế biến gỗ và các SP từ gỗ, tre [23]
- 2208 - 8. SX giấy và các SP từ giấy [24]
- 2209 - 9. Xuất bản, sao in [25]

Chọn tất cả | Bỏ tất cả | Số ngành nghề được chọn: 0 / 56

Đơn vị hành chính*

- 701 - TP.HCM [1]
- 70101 - Quận 1 TP.HCM [2]
- 7010101 - Phường Tân Định Quận 1 [3]
- 7010103 - Phường Đa Kao Quận 1 [4]
- 7010105 - Phường Bến Nghé Quận 1 [5]
- 7010107 - Phường Bến Thành Quận 1 [6]
- 7010109 - Phường Nguyễn Thái Bình Quận 1 [7]
- 7010111 - Phường Phạm Ngũ Lão Quận 1 [8]
- 7010113 - Phường Cầu Ông Lãnh Quận 1 [9]
- 7010115 - Phường Cô Giang Quận 1 [10]
- 7010117 - Phường Nguyễn Cư Trinh Quận 1 [11]
- 7010119 - Phường Cầu Kho Quận 1 [12]
- 70103 - Quận 2 TP.HCM [313]

Chọn tất cả | Bỏ tất cả | Số đơn vị hành chính được chọn: 0 / 319

Tiêu chí khai thác

Ngành nghề Đơn vị hành chính Cả hai

Độ ủng hộ tối thiểu 3 (%)

Độ tin cậy tối thiểu 70 (%)

Mức sử dụng từ 2.5 đến 4.4 [Vừa] [Nhiều]

Số tiêu thụ từ 0 đến 540,140

Khai thác dữ liệu

Kết quả khai thác

Luật 5: Khách hàng sống ở <Phường 05 Quận 3> sử dụng điện <Vừa>

Luật 6: Khách hàng sống ở <Phường Tân Định Quận 1> sử dụng điện <Vừa>

Luật 7: Khách hàng sống ở <Phường 06 Quận 3> sử dụng điện <Vừa>

Luật 8: Khách hàng sống ở <Phường 07 Quận 3> sử dụng điện <Vừa>

Luật 9: Khách hàng sống ở <Phường 08 Quận 3> sử dụng điện <Vừa>

Luật 10: Khách hàng sống ở <Phường Đa Kao Quận 1> sử dụng điện <Vừa>

Luật 11: Khách hàng sống ở <Phường Bến Nghé Quận 1> sử dụng điện <Vừa>

Luật 12: Khách hàng sống ở <Phường Bến Thành Quận 1> sử dụng điện <Vừa>

Luật 13: Khách hàng sống ở <Phường Nguyễn Thái Bình Quận 1> sử dụng điện <Vừa>

Luật 14: Khách hàng sống ở <Phường Phạm Ngũ Lão Quận 1> sử dụng điện <Vừa>

Luật 15: Khách hàng thuộc ngành <I. Bán buôn, bán lẻ> sống ở <TP.HCM> sử dụng điện <Vừa>

Luật 16: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường Nguyễn Cư Trinh Quận 1> sử dụng điện <Vừa>

Luật 17: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường 05 Quận 3> sử dụng điện <Vừa>

Luật 18: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường Tân Định Quận 1> sử dụng điện <Vừa>

Luật 19: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường 06 Quận 3> sử dụng điện <Vừa>

Luật 20: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường 08 Quận 3> sử dụng điện <Vừa>

Luật 21: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường Đa Kao Quận 1> sử dụng điện <Vừa>

Luật 22: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường Bến Nghé Quận 1> sử dụng điện <Vừa>

Luật 23: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường Bến Thành Quận 1> sử dụng điện <Vừa>

Luật 24: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường Nguyễn Thái Bình Quận 1> sử dụng điện <Vừa>

Luật 25: Khách hàng thuộc ngành <I. Sinh hoạt tiêu dùng dân cư Thành thị> sống ở <Phường Phạm Ngũ Lão Quận 1> sử dụng điện <Vừa>

Thời gian thực hiện: 0 giờ 1 phút 25 giây

Kết quả thực thi chương trình từ số liệu được dùng từ dữ liệu sử dụng điện năm 2008 của Công ty Điện lực TP.HCM với 1.097.944 hóa đơn, 103.741 khách hàng, 87 ngành nghề và 319 đơn vị hành chính cho thấy chương trình có thể hỗ trợ khai thác luật tiềm ẩn theo nhu cầu của người sử dụng trong ngành điện, từ đó cho phép theo dõi tình hình sử dụng điện, cung cấp thông tin hỗ trợ ra quyết định trong việc quản lý, định hướng phát triển và sử dụng hiệu quả năng lượng điện.

8. Kết luận

Việc cải tiến và đưa tri thức logic mờ vào CSDL sử dụng điện là một yêu cầu rất cần thiết. Trên cơ sở các mức sử dụng mờ, chúng ta có thể thống kê, đánh giá tình hình sử dụng điện.

Khai thác các luật sử dụng điện cung cấp các thông tin tiềm ẩn bên trong số liệu sử dụng điện, hỗ trợ các thông tin chiến lược cho các cấp lãnh đạo định hướng các chính sách khai thác và phát triển năng lượng điện.

Ngoài logic mờ và luật kết hợp, chúng ta có thể tìm hiểu và ứng dụng những phương pháp khai phá dữ liệu khác để khai thác các tri thức tiềm ẩn khác ngoài luật sử dụng điện như: cây quyết định (decision tree), mạng nơ-ron (neural network), ...

TÀI LIỆU THAM KHẢO

1. Professor Ho Tu Bao (2000), *Introduction to Knowledge Discovery and Data Mining*, Institute of Information Technology National Center for Natural Science and Technology, pp.58-59, 78-79.
2. George Bojadziev, Maria Bojadziev (2007), *Fuzzy Logic for Business, Finance, and Management (2nd Edition)*, Simon Fraser University, Canada, British Columbia Institute of Technology, Canada, pp.37-39, 42, 45-46.
3. Professor Charles Elkan (2004), *Clustering with k-means: faster, smarter, cheaper*, University of California, San Diego.
4. GS.TSHK Hoàng Văn Kiếm, TS. Đỗ Văn Nhơn (2007), *Chuyên đề Phương pháp toán trong tin học*, ĐHQG TP.HCM, Chương trình Thạc sĩ CNTT qua mạng.
5. PGS.TS. Đỗ Phúc (2007), *Bài giảng chuyên đề Khai phá dữ liệu và Nhà kho dữ liệu*, ĐHQG TP.HCM, Chương trình Thạc sĩ CNTT qua mạng.