Sử dụng chuỗi thời gian mờ dự báo số lượng doanh nghiệp nhỏ và vừa

NGUYỄN TÂN ÂN, Trường Đại học Sư phạm Hà Nội
NGUYỄN VĂN TRÚC, Bộ Kế hoạch và Đầu tư

Chiến lược phát triển của tất cả các quốc gia trên thế giới đều quan tâm hàng đầu đến việc phát triển doanh nghiệp (DN). Nếu dự báo được tình hình và quy luật phát triển của nó thì chúng ta sẽ có kế hoạch và chính sách phù hợp để hỗ trợ và thúc đẩy sự phát triển của DN. Sau đây chúng tôi xin giới thiệu một mô hình dự báo dựa trên chuỗi thời gian mờ.

Dự báo là một phát biểu về tương lai, là cơ sở để đưa ra những quyết định. Đã có nhiều phương pháp dự báo được sử dụng trong các mô hình dự báo khác nhau. Tuy nhiên, do đây là bài toán phức tạp, nên các phương pháp vẫn còn bộc lộ nhiều nhược điểm: độ phức tạp tính toán lớn, độ chính xác không cao, không tính được ảnh hưởng của các yếu tố khách quan và kết quả dự báo không phải bao giờ cũng đạt yêu cầu nên con người vẫn phải chịu nhiều bất ngờ trước những biến động của cuộc sống. Đã có rất nhiều dự báo sai, hoặc không dự báo được dẫn đến những hậu quả tai hại. Vì thế nghiên cứu về dự báo luôn là vấn đề có tính thời sự.

Để dự báo, thông thường phải thu thập thông tin về đối tượng dự báo, về những đối tượng liên quan... Trên cơ sở kết quả xử lý những thông tin đó, đưa ra những phán đoán của mình về tương lai. Tuy vậy, không phải bao giờ cũng thu thập được đầy đủ những thông tin cần thiết, ngay cả những thông tin đã thu thập được cũng không phải bao giờ cũng chính xác. Thông tin không đầy đủ, không chính xác làm cho việc xử lý bài toán dự báo càng thêm rắc rối. Tuy nhiên, đây là thực tế không thể né tránh, nên các phương pháp xử lý thông tin không đầy đủ, không chính xác được áp dụng khi giải quyết bài toán dự báo để đưa ra lời dự báo cũng là những vấn đề cần quan tâm.

Khái niệm tập mờ được đưa ra từ năm 1965 và ngày càng được ứng dụng trong nhiều lĩnh vực khác nhau, nhất là trong điều khiển và trí tuệ nhân tạo. Dựa vào kết quả thống kê và các thao tác hợp lý trên chuỗi thời gian mờ, ta có thể làm giảm đi nhiều độ phức tạp tính toán, rút ra được những quy luật của một quá trình, từ đó đưa ra phán đoán đủ chính xác về tương lai. Trong nghiên cứu này, chúng tôi áp dụng phương pháp tính toán trên chuỗi thời gian mờ để dự báo sự biến thiên của số lượng các DN nhỏ và vừa hàng năm ở Việt Nam.

Mô hình dự báo mờ

Chuỗi thời gian mờ [5]

Định nghĩa Tập mờ [3,4]:

Cho tập vũ trụ $U = \{u_1, u_2, ..., u_n\}$. Một tập mờ A trên U được xác định bởi:

$$A = \{ u_1, \mu_A(u_1)), (u_2, \mu_A(u_2), ..., (u_n, \mu_A(u_n))\}$$

Trong đó, $\mu : U \to [0,1]$ gọi là hàm thuộc, giá trị $\mu_A(u_i)$ $i = 1,2,...,n$ là độ thuộc của u_i vào tập mờ A.

Có thể biểu diễn tập mờ A_i như sau:

$$A_i = \frac{a_{i1}}{u_1} + \frac{a_{i2}}{u_2} + ... + \frac{a_{ik}}{u_s} + ... + \frac{a_{im}}{u_m}$$

Trong đó, u_s là độ thuộc của a_{ik} vào tập mờ A_i

- Định nghĩa Chuỗi thời gian mờ [5]:

Coi tập con của R được gọi là tập vũ trụ, trên đó đã định nghĩa các tập mờ $\mu_i(t)$ $i = 1,2,...$ Khi đó, mỗi một bộ $F_i(t)$ gồm các μ_i được gọi là chuỗi thời gian mờ.

Mô hình dự báo mờ và phương pháp tính toán

Giả sử đã biết số liệu của những năm trước: $A_1, A_2, ..., A_n$. Đây chính là những tập mờ được đặc trưng bởi các hàm thuộc μ_i. Cần tính A_{n+1} là số liệu của năm cần dự báo.

Bài toán trở thành, biết:

$A_1 \to A_2$
$A_2 \to A_3$
$A_2 \to A_3$
...
$A_{n-1} \to A_n$

Tìm $A_{n+1} = ?$

Thông thường người ta làm như sau [2]:

Từ các quan hệ $A_i \to A_{i+1}$ đã biết, ta tìm các quan hệ $R_i = A_1^T \bullet A_1$, với là phép hợp thành nào đó. Tức là ta tìm được:

$R_1 = A_1^T \bullet A_1$
$R_2 = A_2^T \bullet A_2$
...
$R_n = A_n^T \bullet A_n$

Từ các R_i ta tìm được R và cuối cùng $A_{n+1} = R \bullet A_n$

Việc tính toán như vậy thường phải thực hiện một khối lượng tính toán lớn, phức tạp. Trong trường hợp phải tìm các quan hệ mờ thì khối lượng tính toán còn lớn hơn nữa.

Để đạt được mục tiêu tìm A_{n+1}, trong [5], Shyi – Uing Chen đã đưa ra phương pháp tính toán dựa trên lý thuyết tập mờ, theo đó khối lượng tính toán giảm đi rất nhiều, nhưng vẫn cho kết quả đủ tốt. Phương pháp đó như sau:

Bước 1: Dựa trên dãy thời gian $A_1, A_2, ..., A_n$, ta xác định tập vũ trụ U là $[D_{\min} - D_1, D_{\max} + D_2]$ với D_1, D_2 là các số bù. Tiếp đó phân vùng tập vũ trụ U thành k vùng đều nhau u_i:

$$u_1 = [u_1, u_{12}]; u_2 = [u_{12}, u_2]; ...;$$
$$u_k = [u_{k1}, u_{k2}]$$

Trong đó $u_{11} = D_{\min} - D_1$, $u_{k2} = D_{\max} + D_2$.

Bước 2:

Ta định nghĩa các tập mờ

$$A_1 = \frac{a_1}{u_1} + \frac{a_{12}}{u_2} + ... + \frac{a_{1m}}{u_m}$$

$$A_2 = \frac{a_{21}}{u_1} + \frac{a_2}{u_2} + ... + \frac{a_{2m}}{u_m}$$

...

$$A_i = \frac{a_{i1}}{u_1} + \frac{a_{i2}}{u_2} + ... + \frac{a_{is}}{u_s} + ... + \frac{a_{im}}{u_m}$$

...

$$A_n = \frac{a_{n1}}{u_1} + \frac{a_{n2}}{u_2} + ... + \frac{a_m}{u_m}$$

Trong đó u_s là các phân vùng và a_{is} là độ thuộc của u_s vào tập mờ A_i, $i = 1,2,...,n$, $s = 1,2,...,m$

Bước 3:

Từ dãy thời gian ta có:

$A_1 \to A_1$

$A_2 \to A_3$

...

$A_{n-1} \to A_n$

Trong các quan hệ trên có thể có nhiều hơn một quan hệ có cùng vế trái. Chia các quan hệ trên vào các nhóm có cùng vế trái.

Bước 4:

Tính toán giá trị ra:

Ta có các trường hợp sau:

Trường hợp 1: Với các nhóm chỉ có một quan hệ đơn lẻ, tức là trong nhóm chỉ có một quan hệ $A_i \to A_j$ (vế trái suy ra một vế phải) ta tính như sau: Tìm u_s có độ thuộc vào A_j lớn nhất, lấy điểm giữa của u_s làm giá trị ra.

Trường hợp 2: Vế trái có nhiều vế phải ta tính như sau:

Xét tất cả các vế phải tìm những u_s nào tương ứng với các vế phải đó. Lấy giá trị điểm giữa các u_s này cuối cùng lấy trung bình cộng của các giá trị đó làm giá trị ra.

Trường hợp 3: Với những A_j mà A_j không thuộc nhóm nào, như vậy ta xét như là $A_j \to A_j$ và xem u_s nào tương ứng với A_j, lấy điểm giữa làm giá trị ra.

Ví dụ:

Từ số liệu thống kê ta có dãy thời gian. Từ dãy thời gian này đưa ra con số dự báo cho năm tiếp theo:

Năm 2000 - 14457 DN
Năm 2001 - 19800 DN
Năm 2002 - 21535 DN
Năm 2003 - 27698 DN
Năm 2004 - 37230 DN
Năm 2005 - 39959 DN
Năm 2006 - 40347 DN

Phương pháp tính toán như sau [5]:

Bước 1: Dựa vào số liệu thực tế $D_{\min} = 14457$ $D_{\max} = 40347$ ta xác định được tập vũ trụ U là $[D_{\min} - D_1, D_{\max} + D_2]$ với $D_1 = 457, D_2 = 653$ là các số bù. Tập vũ trụ sẽ là U= [14000,41000] . Từ đó có các phân vùng vũ trụ như sau:

u_1=[14000, 17000]; u_2=[17000, 20000];
u_3=[20000, 23000]; u_4=[23000, 26000];
u_5=[26000, 29000]; u_6=[29000, 32000];
u_7=[32000, 35000]; u_8=[35000, 38000];
u_9=[38000, 41000]

Sử dụng 9 tập mờ A_1; A_2; A_3; A_4; A_5; A_6; A_7; A_8; A_9:

$$A_1 = \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9}$$

$$A_2 = \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9}$$

$$A_3 = \frac{0}{u_1} + \frac{0.5}{u_2} + \frac{1}{u_3} + \frac{0.5}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9}$$

$$A_4 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0.5}{u_3} + \frac{1}{u_4} + \frac{0.5}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9}$$

$$A_5 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0.5}{u_4} + \frac{1}{u_5} + \frac{0.5}{u_6} + \frac{0}{u_7} + \frac{0}{u_8} + \frac{0}{u_9}$$

$$A_6 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0.5}{u_5} + \frac{1}{u_6} + \frac{0.5}{u_7} + \frac{0}{u_8} + \frac{0}{u_9}$$

$$A_7 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0.5}{u_6} + \frac{1}{u_7} + \frac{0.5}{u_8} + \frac{0}{u_9}$$

$$A_8 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0.5}{u_7} + \frac{1}{u_8} + \frac{0.5}{u_9}$$

$$A_9 = \frac{0}{u_1} + \frac{0}{u_2} + \frac{0}{u_3} + \frac{0}{u_4} + \frac{0}{u_5} + \frac{0}{u_6} + \frac{0}{u_7} + \frac{0.5}{u_8} + \frac{1}{u_9}$$

...

Số lượng doanh nghiệp được thành lập theo từng năm tương ứng với các tập mờ trên như Bảng 1

BẢNG 1: SỐ LIỆU DN ĐƯỢC THÀNH LẬP THEO TỪNG NĂM TƯƠNG ỨNG VỚI CÁC TẬP MỜ

Năm	*Số DN thành lập thực tế*	*Số DN thành lập được mờ hóa*
2000	14457	A_1
2001	19800	A_2
2002	21535	A_3
2003	27698	A_5
2004	37230	A_8
2005	39959	A_9
2006	40347	A_9
2007	?	

Bước 2:

Từ Bảng 1 ta có:

$A_1 \to A_2$; $A_2 \to A_3$; $A_3 \to A_5$; $A_5 \to A_8$; $A_8 \to A_9$; $A_9 \to A_9$

Bước 3:

Xếp các quan hệ trên vào các nhóm, theo dữ liệu thực tế, ta có :

Nhóm 1	$A_1 \to A_2$
Nhóm 2	$A_2 \to A_3$
Nhóm 3	$A_3 \to A_5$
Nhóm 4	$A_5 \to A_8$
Nhóm 5	$A_8 \to A_9$
Nhóm 6	$A_9 \to A_9$

[2001]: từ $A_1 \to A_2$ điểm giữa của u_2 là 15500, số DN thành lập năm 2001 được dự báo là 15500

[2002]: Từ $A_2 \to A_3$, điểm giữa của u_3 là 18500, số DN thành lập năm 2002 được dự báo là 18500

[2003]: $A_3 \to A_5$, điểm giữa của u_5 là 27500, số DN thành lập năm 2003 được dự báo là 27500

[2004]: $A_5 \to A_8$, điểm giữa của u_8 là 36500, số DN thành lập năm 2004 được dự báo là 36500.

[2005]: $A_8 \to A_9$, điểm giữa của u_9 là 39500, số DN thành lập năm 2005 được dự báo là 39500.

[2006]: $A_9 \to A_9$, điểm giữa của u_9 là 39500, số DN thành lập năm 2006 được dự báo là 39500.

Kết quả dự báo so với thực tế như Bảng 2:

Năm 2006 sai số là 40347 - 39500 = 847.

BẢNG 2: KẾT QUẢ DỰ BÁO SO VỚI THỰC TẾ

Năm	*Số DN thành lập thực tế (từ số liệu đã có)*	*Số DN thành lập theo dự đoán*
2000	14457	A_1
2001	19800	15500
2002	21535	18500
2003	27698	27500
2004	37230	36500
2005	39959	39500
2006	40347	39500
2007	58195 (bất thường)	42500

Ta thấy $\frac{847}{40347} = 0,02$ là sai lệch lý tưởng.

Tất nhiên đây là những số liệu lấy từ thực tế để minh họa cho phương pháp tính toán. Và việc lấy này thật may mắn.

Năm 2007 sai số là 58195 - 42500 = 15695.

Trường hợp này tỷ lệ là: $\frac{15695}{58195} = 0,27$.

Sai số khoảng 3/10. Trường hợp này phải xét thêm những yếu tố ảnh hưởng bất thường đến sự phát triển về số lượng các DN.

Kết luận

Theo các tính toán trình bày ở trên, thì việc chia khoảng là việc rất quan trọng, nếu chia khoảng nhỏ thì khối lượng tính toán sẽ lớn, nhưng kết quả sẽ tốt hơn. Vấn đề này cần tiếp tục nghiên cứu. Trong trường hợp dãy thời gian biến đổi quá thất thường, kết quả dự báo sẽ kém chính xác, khi đó cần bổ sung thêm thông tin để dự báo chính xác hơn.■

TÀI LIỆU THAM KHẢO :

[1]. Nguyễn Quang Dong, *Phân tích chuỗi thời gian trong tài chính*, NXB Khoa học và Kỹ thuật. Hà Nội 2010

[2]. H. Bintley, *Time series analysis with REVEAL*, Fuzzy sets and systems 23 (1987), 97-118

[3]. L.A. Zadeh, *Fuzzy sets* // Inform and Control, 8, 1965, 338-353

[4]. L.A. Zadeh, *Fuzzy sets and systems* // Proc, Symp, Systems Theory, Brooklyn Polytech, Inst. Brooklyn, New York, pp, 29-39, 1966.

[5]. Shyi-Ming Chen, *Forecasting enrollments based on fuzzy time series*. Fuzzy sets and systems 81 (1996), 311-319