پنجمین کنفرانس ملی تحلیل پوششی داده‌ها
13-15 شهریور 1392- مو سسه آموزش عالی آیندگان– تنکابن- ایران

تحلیل پوششی داده ها با مقادیر از دست رفته: نگرش خوشه بندی فازی

محمد تیموری، دانشگاه آزاد اسلامی واحد قزوین، گروه صنایع، تهران، ایران
m66.teimouri@gmail.com
اسماعیل مهدی زاده، دانشگاه آزاد اسلامی واحد قزوین، گروه صنایع، تهران، ایران
emqiau@yahoo.com

چکیده

در تحلیل پوششی داده ها (*DEA*)، داده های ورودی و خروجی همه واحد های تصمیم گیرنده (*DMU*) برای مقایسه لازم هستند. در *DEA* کلاسیک فرض بر این است که، داده های عددی برای ورودی و خروجی ها در دسترس است. با این حال، در بسیاری از برنامه های کاربردی، تعدادی از *DMU* ها دارای ورودی و یا خروجی از دست رفته می باشند. این مقاله یک روش، برای استفاده از *DEA* با ورودی و خروجی از دست رفته با استفاده از مفاهیم خوشه بندی فازی و الگوریتم ژنتیک (GA) معرفی می کند. ما از یک مجموعه داده کامل و واقعی که از مطالعات پیشین گرفته شده است، برای ارزیابی اثر روش پیشنهادی با فرض دارا بودن سطوح مختلف مقادیر از دست رفته استفاده می کنیم.

کلمات کلیدی: واحد تصمیم گیرنده، ورودی و خروجی از دست رفته، خوشه بندی فازی.

1. مقدمه

تحلیل پوششی داده ها یک روش برنامه ریزی مشهور، برای اندازه گیری کارایی نسبی واحد های تصمیم گیرنده می باشد. از زمان ارائه روش تحلیل پوششی داده ها توسط چارنز و همکاران [1]، این روش به عنوان ابزاری موثر برای ارزیابی و الگوبرداری بکار گرفته شده است. سادگی فهم واجرای روش تحلیل پوششی داده ها و در کنار آن، دقت بالا و کاربرد وسیع آن در زمینه های مختلف سیاسی، فرهنگی، اجتماعی و اقتصادی باعث شده است، پژوهشگران زیادی از این روش برای دست یافتن به اهداف خود استفاده کنند. اما عمده این پژوهش ها بر روی مدل های *DEA* که مقادیر ورودی و خروجی آن ها کامل بوده صورت می گرفته است. در *DEA* کلاسیک، فرض بر این است که داده های عددی برای ورودی و خروجی ها در دسترس است. با این حال، در بسیاری از برنامه های کاربردی، گاهی اوقات تعدادی از واحد های تصمیم گیری دارای ورودی و یا خروجی از دست رفته می باشند و این واحدهای تصمیم گیری قادر به گزارش تمام آمارهای مورد نیاز نیستند. در طی دهه گذشته نگرش های مختلفی برای مقابله با این مسائل در نظر گرفته شده است، که از جمله می توان به [2-4] اشاره کرد.

در این مقاله ما از رویکرد خوشه بندی فازی به همراه الگوریتم ژنتیک (GA) برای بازیابی مقادیر از دست رفته (ورودی - خروجی) در واحد های تصمیم گیرنده استفاده می کنیم.

2. الگوریتم خوشه بندی FCM

این الگوریتم یک مجموعه بردار S بعدی $X=\{X_1,X_2,...,X_N\}$ درون C خوشه که $X_{ij}=\{x_{j1},x_{j2},...,x_{js}\}$ معرف j-امین نمونه $j=1,2,...,n$ تقسیم بندی می کند. تابع هدف FCM را می توان به صورت زیر فرموله کرد:

$$\min J_m(\tilde{U},V) = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}\left\|x_j - v_i\right\|^2$$

$$s.t$$

$$\sum_{i=1}^{c}\mu_{ij}=1 \quad \forall j=1,2,...,n \qquad (1)$$

به طوری که μ_{ij} برابر درجه عضویت داده j-ام می باشد. کمترین مقدار J_{FCM} مربوط به بهترین حالت خوشه بندی خواهد بود . V_i نشان دهنده مختصات مرکز i-امین خوشه می باشد $Vi=\{v_{i1}, v_{i2},...,v_{in}\}$. n نیز تعداد ابعاد V_i یا تعداد معیار های تشابه می باشد. m به عنوان پارامتر فازی ساز نامیده می شود، که بازه تغییرات آن به صورت $m \in [1,\infty]$ می باشد. مختصات مرکز خوشه از رابطه زیر به دست می آید:

$$v_{ij}=\frac{\sum_{j=1}^{n}\mu_{ij}^{m}\cdot x_{ji}}{\sum_{j=1}^{n}\mu_{ij}^{m}} \qquad (2)$$

به طوری که j متغیری است، برای نشان دادن فضای معیارها، $j=1,2,...,n$.

گام های روش FCM به صورت زیر می باشد:

گام1: مقدار c را مشخص کرده $(2 \le c < n)$ و یک مقدار برای m انتخاب می کنیم. ماتریس افراز اولیه، $U^{(0)}$ را حدس می زنیم. هرگام از این الگوریتم با یک r مشخص می گردد، $r=0, 1, 2, ...$.

گام 2: مرکز خوشه ها $\{V_i^{(r)}\}$را در هر تکرار محاسبه می کنیم.

گام 3: ماتریس افراز را برای تکرار rام، $\tilde{U}^{(r)}$ به صورت زیر به هنگام می کنیم:

$$\mu_{ij}^{(r+1)}=\left[\sum_{k=1}^{c}\left(\frac{d_{ij}^{(r)}}{d_{kj}^{(r)}}\right)^{2/(m-1)}\right]^{-1} \qquad (m\neq 1) \qquad (3)$$

گام 4: اگر $\|\tilde{U}^{(r+1)} - \tilde{U}^{(r)}\| \le \varepsilon$ آن گاه محاسبات به پایان می رسد در غیر این صورت به گام دوم برمی گردیم.

3. روش پیشنهادی

در این روش قبل از حل مجموعه داده ناکامل توسط مدل DEA (در این مقاله از مدل CCR وروردی محور استفاده شده است)، ورودی و خروجی های از دست رفته، توسط روش خوشه بندی فازی و با استفاده از الگوریتم ژنتیک مورد بازیابی قرار گرفته و پس از آن این مجموعه داده که کامل شده است، برای حل به مدل CCR فرستاده می شود. برای یک مجموعه داده (DMU ها) ناکامل s بعدی (ورودی و خروجی) $\ddot{X} = \{\ddot{x}_1 \ddot{x}_2 \ \dots \ \ddot{x}_n\}$ با g ورودی - خروجی از دست رفته روش پیشنهادی می تواند به صورت گام های زیر توصیف شود:

گام 1- برای هر DMU ناکامل $\ddot{x}_b$ $(1 \le b \le n)$ که j امین ورودی یا خروجی آن $\ddot{x}_{jb}$، از دست رفته است. q همسایه (DMU) نزدیکتر برای آن را با استفاده از فاصله جزیی [5] پیدا می کنیم، سپس بازه $[\ x_{jb}^{-}\ x_{jb}^{+}\ ]$ را برای $\ddot{x}_{jb}$ بر اساس همسایه ها تعیین می کنیم.

گام 2- مقادیر m، c و ε را برای خوشه بندی انتخاب کرده و همچنین مقادیر پارامترهای الگوریتم ژنتیک شامل اندازه جمعیت M، حداکثر تعداد نسل G، احتمال تقاطع Pc و احتمال جهش P_m را تنظیم می کنیم.

گام 3- در تکرار t $(t=12 \ \dots \ T)$، مجموعه داده ناقص $\ddot{X}$ با استفاده از الگوریتم ژنتیک کامل می شود و سپس مجموعه داده کامل X با استفاده از الگوریتم خوشه بندی فازی FCM خوشه بندی می شود.

گام 4- مقدار برازندگی برای هر جواب با استفاده از رابطه (1) محاسبه می شود و بهترین جواب ها ذخیره می شوند.

گام 5- اگر تعداد تکرارها الگوریتم پیشنهادی برابر با $t=T$ شد، متوقف می شویم و مجموعه داده کامل X را برای انجام تحلیل پوششی داده ها بر می گزینیم. در غیر این صورت به گام 3 می رویم.

گام 6- روش CCR ورودی محور را برای مجموعه داده کامل X انجام می دهیم و سپس رتبه بندی واحد های تصمیم گیری و همچنین واحدهای کارا و ناکارا را مشخص می کنیم.

4. مثال عددی

برای اعتباردهی روش پیشنهادی، ما از یک مجموعه داده کامل و واقعی که از مطالعات پیشین گرفته شده است [4]، شامل 22 تا DMU، که هر DMU شامل چهار ورودی و سه خروجی است ، استفاده می کنیم. همچنین برای ارزیابی اثر الگوریتم، فرض دارا بودن سطح های مختلف مقادیر از دست رفته (5٪، 10٪ و 20٪) را در نظر می گیریم. همچنین در این مقاله، ما از معیار متوسط درصد خطای مطلق (MAPE) برای ارزیابی دقت پیش بینی الگوریتم پیشنهادی استفاده می کنیم [4]. بهترین بازیابی مقادیر زمانی بدست می آید که MAPE کمترین مقدار را بگیرد. مقادیر مربوط به کارایی های بدست آمده توسط DMU های مختلف با توجه به در صد های مختلف از بین رفتگی و همچنین مقدار MAPE مربوط به آن ها در جدول 1 نشان داده شده است.

جدول 1: ورودی و خروجی هر 3 زیر واحد

DMU	کارایی با مقادیر کامل	کارایی با 5% مقادیر از دست رفته	کارایی با 10% مقادیر از دست رفته	کارایی با 20% مقادیر از دست رفته
1	1.000	1.000	1.000	0.706
2	0.558	0.550	0.558	0.985
3	1.000	0.897	0.649	0.509
4	0.860	0.868	0.870	0.857
5	0.611	0.550	0.460	0.870
6	1.000	1.000	1.000	0.896
7	1.000	0.896	0.758	0.704
8	0.655	0.648	0.650	0.613
9	1.000	0.972	0.924	0.938
10	0.629	0.640	0.633	0.629
11	0.933	1	0.992	1
12	0.617	0.609	0.617	0.655
13	1.000	0.971	0.769	0.564
14	0.342	0.341	0.308	0.226
15	0.598	0.601	0.598	0.590
16	0.835	0.830	0.899	0.938
17	0.426	0.426	0.403	0.689
18	1.000	1.000	0.930	0.945
19	0.932	0.932	0.981	0.992
20	1.000	1.000	0.675	0.432
21	1.000	0.666	0.476	0.867
22	0.610	0.612	0. 561	0.480
MAPE		**3.871**	**11.445**	**23.635**

5. نتیجه‌گیری

این مقاله یک روش، برای استفاده از DEA با ورودی و خروجی از دست رفته با استفاده از مفاهیم خوشه بندی فازی و الگوریتم ژنتیک معرفی می کند. نتایج محاسباتی بدست آمده نشان می دهد که الگوریتم پیشنهادی از دقت بالایی در بازیابی مقادیر از دست رفته برخوردار است. ما از یک مجموعه داده کامل و واقعی که از مطالعات پیشین گرفته شده است، برای ارزیابی اثر روش پیشنهادی با فرض دارا بودن سطح های مختلف مقادیر از دست رفته استفاده کردیم.

مراجع

[1] Charnes, A., W. W. Cooper, E. Rhodes, "Measuring the efficiency of decision making units" ,*European Journal of Operational Research*, 2 (1978) 429–444.
[2] Kuosmanen, T., "Modeling blank data entries in data envelopment analysis" ,*Econometrics working paper archive at WUSTL*. No. 0210001 (2002).
[3] Smirlis, Y. G., E. K. Maragos, D. K. Despotis, "Data envelopment analysis with missing values: An interval DEA approach" ,*Applied Mathematics and Computation*, 177(1) (2006) 1–10.
[4] Ben-Arieh, D., D. -K. Gullipalli, "Data Envelopment Analysis of clinics with sparse data: Fuzzy clustering approach" ,*Computers & Industrial Engineering*, 63 (2012) 13–21.
[5] Hathaway, R. J., J. C. Bezdek, "Fuzzy c-means clustering of incomplete data" ,*IEEE Transactions on Systems, Man, and Cybernetics—Part b: Cybernetics*, 31(5) (2001) 735–744.