



دانشگاه صنعتی امیرکبیر  
(پلی تکنیک تهران)  
دانشکده مدیریت علم و فناوری

گزارش کار پروژه پایانی

خوشه بندی فازی تصویر: یک روش جدید هوش محاسباتی

نگارش

رضا اکبری مقدم

استاد

دکتر مهدی قطعی

بهمن ماه ۹۹

## فهرست

چکیده.....	۴
مقدمه.....	۴
الگوریتم FC-PFS.....	۷
پیاده سازی الگوریتم بر روی مجموعه داده .....	۸
نتیجه گیری.....	۱۲

۵ .....	شکل ۱
۵ .....	شکل ۲
۵ .....	شکل ۳
۶ .....	شکل ۴
۶ .....	شکل ۵
۶ .....	شکل ۶
۷ .....	شکل ۷
۸ .....	شکل ۸
۸ .....	شکل ۹
۸ .....	شکل ۱۰
۹ .....	شکل ۱۱
۹ .....	شکل ۱۲
۹ .....	شکل ۱۳
۱۰ .....	شکل ۱۴
۱۰ .....	شکل ۱۵
۱۰ .....	شکل ۱۶
۱۰ .....	شکل ۱۷
۱۱ .....	شکل ۱۸
۱۱ .....	شکل ۱۹
۱۱ .....	شکل ۲۰
۱۱ .....	شکل ۲۱
۱۱ .....	شکل ۲۲

## چکیده

خوشه بندی فازی مخصوصاً خوشه بندی فازی C-means (FCM) یک ابزار مفید برای پردازش الگوهای پیشنهادی از پایگاه داده میباشد. با این وجود کیفیت خوشه بندی این الگوریتم بالا نمیباشد تا زمانی که براساس مجموعه های فازی سنتی، که برای مقابله با محدودیت ها پیشنهاد شده است. این مقاله بهبود دیگری از FCM را در مجموعه های فازی تصویری در نظر میگیرد که یک تعمیمی از مجموعه فازی سنتی و مجموعه فازی شهودی و ارائه یک الگوریتم جدید به نام FCPFS میباشد.

## مقدمه

خوشه بندی فازی به عنوان ابزاری مفید در فرآیندهای شناسایی الگو و کشف دانش در نظر گرفته میشود. FCM یک روش شناخته شده در خوشه بندی فازی است که به عنوان یک ابزار قوی در استخراج قوانین و داده کاوی در نظر گرفته میشود. خواسته های رو به رشد برای بهره برداری از سیستمهای هوشمند و کاملاً خودمختار ، FCM را با چالش بزرگی روبرو کرده است که برای انواع مختلف برنامه هایی مانند تجزیه و تحلیل داده ها ، تشخیص الگو ، تصویر تقسیم بندی ، آنالیز موقعیت یابی گروهی ، تصاویر ماهواره ای و تجزیه و تحلیل مالی مانند آنچه که امروزه دیده می شود استفاده شود. با این وجود ، کیفیت خوشه بندی FCM از آن زمان پیشرفتی نکرده است. انگیزه این مقاله این است که یک روش خوشه بندی فازی جدید طراحی کند که بتواند کیفیت خوشه بندی بهتری نسبت به FCM بدست آورد. تا کنون روش های مختلفی برای بهبود این الگوریتم ارائه شده است. این روشها بر عدم اطمینان مرتبط با دستگاه fuzzifier متمرکز بودند که میزان ابهام FCM را کنترل می کند. حتی اگر کیفیت خوشه بندی آنها از کیفیت بهتری برخوردار باشد ، زمان محاسبات FCM بسیار زیاد است به طوری که محققان ترجیح می دهند FCM را روی مجموعه های فازی شهودی (IFS) گسترش دهند. اخیراً ، ( Cuong ) مجموعه های فازی تصویری را ارائه داده است (PFS) ، که تعمیم مجموعه های فازی سنتی است و مجموعه های فازی شهودی. مدل های مبتنی بر PFS می توانند در موقعیت هایی که نیاز به نظرات انسان است اعمال شوند و شامل پاسخ های بیشتر از انواع: بله ، ممتنع ، نه است. به طوری که آنها می توانند نتایج دقیق تری را برای الگوریتم های خوشه بندی مستقر در PFS ارائه دهند. در این بخش به قوانین اساسی FC-PFS را میپردازد:

- طبقه بندی الگوریتمهای خوشه بندی فازی موجود در ادبیات که به ما کمک می کند جریان در حال رشد را درک کنیم و دلایلی که باید از PFS برای خوشه بندی استفاده شود در قسمت ۲,۱ معرفی شده

است. برخی از عملیات اساسی فازی تصویر، معیارهای فاصله تصویر و روابط فازی تصویر در این بخش نیز ذکر شده است.

- مدل فازی تصویری پیشنهادی برای خوشه بندی و آن راه حل ها در بخش ۲,۲ ارائه شده است.
- در بخش ۲,۳ الگوریتم پیشنهادی FC-PFS شرح داده شده است.

در ادامه مقاله ما به بررسی سایر الگوریتم ها و ادبیات این موضوع نمیپردازیم و مستقیم به سراغ نحوه عملکرد و نتایج این الگوریتم را بررسی میکنیم. بزداک مسئله خوشه بندی فازی را که درجه عضویت یک نقطه داده  $X_k$  به خوشه  $J$  ام را به وسیله  $u_{kj}$  بررسی میکند. تفاوت های این روش با خوشه بندی سخت در اینجا واضح است که یک نقطه داده ممکن است به خوشه های دیگر بر اساس درجه عضویتش تعلق داشته باشد. اصلاح FCM با معیارهای اندازه گیری جدید کافی نیست و استفاده از برخی پسوندهای FS مانند T2FS و IFS انتخاب خوبی میباشد. هوانگ و ری پیشنهاد استقرار FCM در T2FS را ارائه میدهد تا محدودیت های عدم اطمینان و تمرکز بر fuzzifier در کنترل میزان ابهام در FCM را برطرف کند.

$$\tilde{A} = \{(x, u, \mu_{\tilde{A}}(x, u)) | \forall x \in A, \forall u \subseteq J, x \in [0, 1]\}.$$

شکل ۱

در شکل ۱ رابطه T2FS را مشاهده میکنید که  $J$  یک زیرمجموعه از  $X$ ،  $u$  درجه عضویت فازی میباشد. زمانی که  $u$  برابر با یک باشد  $A$  فاصله T2FS نامیده میشود. همچنین زمانی که  $u$  برابر صفر باشد  $A$ ، FS را برمیگرداند. IFS، که عناصری را تشکیل می دهد که با ارزش عضویت و عدم عضویت مشخص می شود، معنی مفیدی برای توصیف آن است زمانی که با داده های مبهم و نامشخص سر و کار داشته باشید.

$$\hat{A} = \{(x, \mu_{\hat{A}}(x), \gamma_{\hat{A}}(x)) | x \in X\},$$

شکل ۲

در معادله شکل ۲  $u$  درجه عضویت هر عضو  $x$  و  $\gamma$  درجه عدم عضویت را نشان میدهد. شاخص فازی شهودی یک عنصر (همچنین شناخته شده است به عنوان درجه تردید) عدم تعین را نشان می دهد که نشان داده شده به عنوان :

$$\pi_{\hat{A}}(x) = 1 - \mu_{\hat{A}}(x) - \gamma_{\hat{A}}(x), \quad \forall x \in X.$$

شکل ۳

زمانی که برابر صفر است IFS، FS را برمیگرداند. درجه تردید را می توان از طریق تابع عضویت توسط اپراتور تولید ارزیابی کرد ،

$$\pi_{\hat{A}}(x) = 1 - \mu_{\hat{A}}(x) - (1 - \mu_{\hat{A}}(x)^{\alpha})^{1/\alpha},$$

شکل ۴

که در آن  $0 < \alpha$  ضرب نمایش است.

اخیرا PFS، cuong را ارائه داده است که یک تعمیم از FS و IFS است. تعریف PFS بیان شده است

در زیر:

$$\dot{A} = \{ \langle x, \mu_{\dot{A}}(x), \eta_{\dot{A}}(x), \gamma_{\dot{A}}(x) \rangle | x \in X \},$$

شکل ۵

که  $u$  درجه مثبت هر عضو  $x$  و  $n$  درجه خنثی و  $\gamma$  درجه منفی را نشان میدهد.

درجه ممتنع هر عنصر محاسبه میشود به وسیله  $\xi_{\dot{A}}(x) = 1 - (\mu_{\dot{A}}(x) + \eta_{\dot{A}}(x) + \gamma_{\dot{A}}(x))$ ,  $\forall x \in X$  که اگر  $\xi_{\dot{A}}(x)$  برابر صفر باشد همان IFS سنتی را برمیگرداند.

سرانجام استفاده از تکنیک های شبیه به عملکرد تولید یاگر و با استفاده از معادله شکل ۴ که به جای  $\mu_{\hat{A}}(x)$  قرار دادیم تا مقدار درجه امتناع یک عنصر را بیابیم:

$$\xi_{kj} = 1 - (u_{kj} + \eta_{kj}) - (1 - (u_{kj} + \eta_{kj})^{\alpha})^{\frac{1}{\alpha}},$$

شکل ۶

## الگوریتم FC-PFS

---

### Fuzzy Clustering Method on Picture Fuzzy Sets

---

- I:** Data  $X$  whose number of elements ( $N$ ) in  $d$  dimensions; Number of clusters ( $C$ ); threshold  $\varepsilon$ ; fuzzifier  $m$ ; exponent  $\alpha$  and the maximal number of iteration  $\max Steps > 0$
- O:** Matrices  $u$ ,  $\eta$ ,  $\xi$  and centers  $V$ ;
- 

### FC-PFS Algorithm:

---

- 1:  $t = 0$
  - 2:  $u_{kj}^{(t)} \leftarrow random$ ;  $\eta_{kj}^{(t)} \leftarrow random$ ;  $\xi_{kj}^{(t)} \leftarrow random$  ( $k = \overline{1, N}$ ,  $j = \overline{1, C}$ ) satisfying constraint (36)
  - 3: Repeat
  - 4:  $t = t + 1$
  - 5: Calculate  $V_j^{(t)}$  ( $j = \overline{1, C}$ ) by equation (42)
  - 6: Calculate  $u_{kj}^{(t)}$  ( $k = \overline{1, N}$ ;  $j = \overline{1, C}$ ) by equation (40)
  - 7: Calculate  $\eta_{kj}^{(t)}$  ( $k = \overline{1, N}$ ;  $j = \overline{1, C}$ ) by equation (41)
  - 8: Calculate  $\xi_{kj}^{(t)}$  ( $k = \overline{1, N}$ ;  $j = \overline{1, C}$ ) by equation (39)
  - 9: Until  $\|u^{(t)} - u^{(t-1)}\| + \|\eta^{(t)} - \eta^{(t-1)}\| + \|\xi^{(t)} - \xi^{(t-1)}\| \leq \varepsilon$  or  $t > \max Steps$
- 

شکل ۷

مجموعه داده های سایت UCI مانند iris,wine,wdbc,glass,heart,cmc در این گزارش استفاده شده است. که شکل ۸ گزارشی از این مجموعه داده ها میباشد.

Dataset	No. of elements	No. of attributes	No. of classes	Elements in each classes
IRIS	150	4	3	(50, 50, 50)
WINE	178	13	3	(59, 71, 48)
WDBC	569	30	2	(212, 357)
GLASS	214	9	6	(70, 76, 17, 13, 9, 29)
IONOSPHERE	351	34	2	(126, 225)
HABERMAN	306	3	2	(225, 81)
HEART	270	13	2	(150, 120)
CMC	1473	9	3	(415, 227, 831)

شکل ۸

همچنین در این گزارش از معیارهای (MA) mean accuracy، (DB) Davies-Bouldin، (RI) Rand Index استفاده شده است.

### پیاده سازی الگوریتم بر روی مجموعه داده

در این قسمت به بررسی الگوریتم بر روی مجموعه داده Haberman میپردازیم که دارای ۴ ویژگی سن، سال عمل بیمار، تعداد گره های مثبت شناسایی شده، وضعیت بقا. در این قسمت  $N=306$  و  $r=4$  و  $C=3$  میباشد. مقادیر اولیه مثبت حالت های خنثی و امتناع ماتریس های  $3 \times 306$  میباشد به صورت زیر تنظیم شده است:

```
([0.32342 , 0.14869 , 0.368887]
 , [0.332441 , 0.133411 , 0.32217]
 ...
 , [0.3470441 , 0.16581 , 0.34774])
```

شکل ۹- مقدار  $u(0)$ 

```
([0.45558 , 0.39744 , 0.559987]
 , [0.44841 , 0.366411 , 0.501217]
 ...
 , [0.412441 , 0.301181 , 0.514001])
```

شکل ۱۰ مقدار  $\eta(0)$



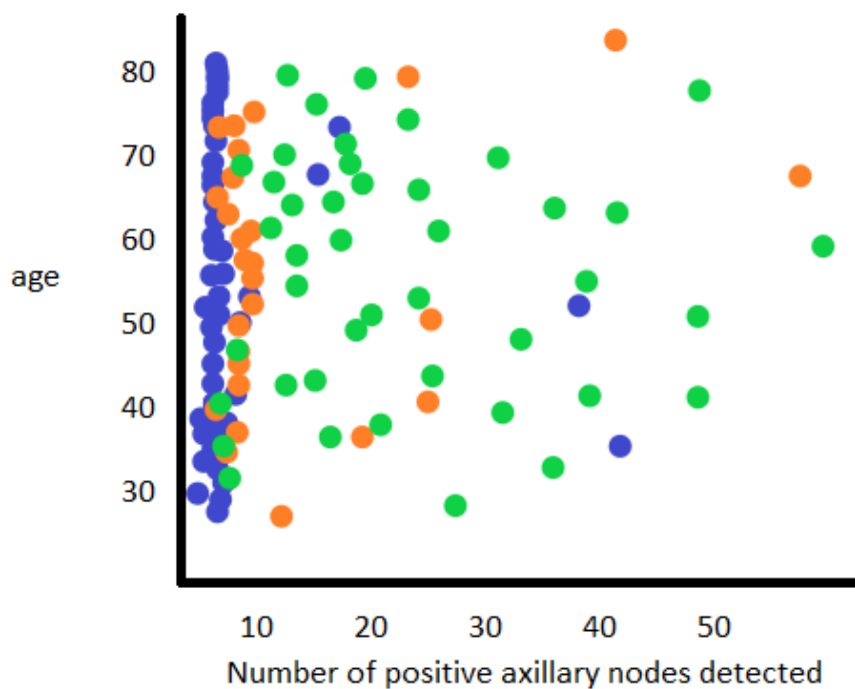
```
([0.223558 , 0.10224 , 0.411987]
 , [0.21841 , 0.12001 , 0.416901]
 , ...
 , [0.215541 , 0.11451 , 0.47811])
```

شکل ۱۱- $\xi(0)$

پراکندگی این نقاط براساس این تنظیمات براساس شکل ۱۳ میباشد. که براساس مرحله پنجم الگوریتم FC-PFS مراکز خوشه به صورت شکل زیر میباشد:

```
([5.4095 , 3.1004 , 2.30487 , 1.658412]
 , [5.9149 , 3.4102 , 2.4901 , 1.7861]
 , [5.47201 , 3.45714 , 2.3474 , 1.58744])
```

شکل ۱۲



شکل ۱۳

حال به بررسی مقادیر مثبت خنثی و امتناع جدید میپردازیم:

```
([0.37899 , 0.14588 , 0.34110]
 , [0.33949 , 0.112202 , 0.312001]
 , ...
 , [0.39881 , 0.11104 , 0.34558])
```

شکل ۱۴-  $u(1)$

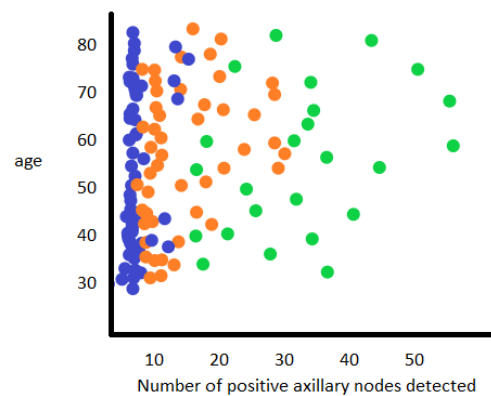
```
([0.4995 , 0.3204 , 0.58887]
 , [0.42149 , 0.34402 , 0.5001]
 , ...
 , [0.40101 , 0.31414 , 0.57414])
```

شکل ۱۵-  $\eta(1)$

```
([0.2015 , 0.11104 , 0.42001]
 , [0.2799 , 0.11002 , 0.419929]
 , ...
 , [0.25552 , 0.1005031 , 0.497474])
```

شکل ۱۶-  $\xi(1)$

پراکندگی اولین مرحله تکرار در شکل ۱۷ نمایش داده شده است.



شکل ۱۷

به همین ترتیب محاسبات ادامه پیدا میکنند تا زمانی که شرط توقف اعمال شود.

```
([0.3095 , 0.1004 , 0.30487]
 , [0.37949 , 0.10102 , 0.39901]
 , ...
 , [0.31201 , 0.18714 , 0.314414])
```

شکل ۱۸ -  $u^*$

```
([0.46015 , 0.39104 , 0.5901]
 , [0.49949 , 0.36002 , 0.501929]
 , ...
 , [0.40121 , 0.3155031 , 0.51474])
```

شکل ۱۹ -  $\eta^*$

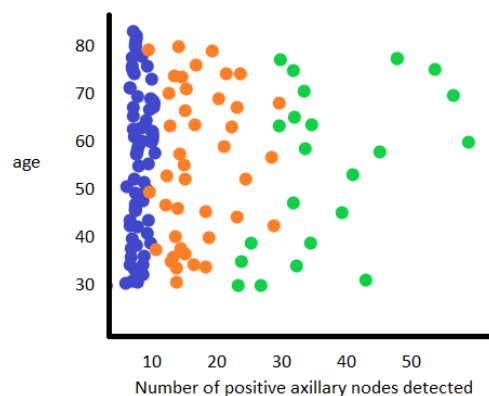
```
([0.20101 , 0.19804 , 0.42702]
 , [0.27821 , 0.11092 , 0.410529]
 , ...
 , [0.2182 , 0.181031 , 0.47974])
```

شکل ۲۰ -  $\xi^*$

همچنین مراکز نهایی خوشه بندی در ماتریس زیر نمایش داده شده است :

```
([5.88875 , 3.7804 , 2.95117 , 1.65612]
 , [5.3349 , 3.6252 , 2.669901 , 1.7661]
 , [5.33251 , 3.425614 , 2.3664 , 1.58844])
```

شکل ۲۱



شکل ۲۲

در جدول زیر نیز به بررسی و ارزیابی خوشه بندی بر روی این مجموعه داده پرداخته ایم.

معیار الگوریتم	MA	Rand Index	DB
FC-PFS	۸۰	۴۹.۶	۲.۷
FCM	۶۰	۴۸.۹	۳.۴
IFCM	۵۵	۴۹.۱	۲.۶
KFCM	۷۵	۵۰.۱	۵.۱
KIFCM	۸۰	۴۹.۷۵	۴.۷

از اطلاعات این جدول میتوان دریافت که FC-PFS کیفیت خوشه بندی بهتری نسبت به دیگر الگوریتم ها دارد.

## نتیجه گیری

در این مقاله با هدف ارتقا کیفیت خوشه بندی FCM مراحل و تنظیماتی اعمال شد که با ترکیب تعمیم های خوشه بندی فازی کیفیت بهتری نسبت به الگوریتم های مرتبط دیگر ارائه شد. نتایج تجربی انجام شده در مجموعه داده های معیار مخزن یادگیری ماشین UCI این مورد را دوباره تأیید کرده اند واقعیت حتی در مورد مقادیر نمایشگر تغییر کرده و اثربخشی الگوریتم پیشنهادی را نشان داد. به علاوه آثار این مضمون با هدف اصلاح این الگوریتم در محیط های توزیع شده و اعمال آن در برخی از برنامه های پیش بینی شده است مانند پیش بینی سهام و پیش بینی وضع هوا.

## منابع

- ۱- <https://link.springer.com/article/10.1007/s00500-015-1712-7>