

**K روش ترکیبی جدید و موثر در خوشه‌بندی فازی بر اساس قاعده نزدیک‌ترین همسایگی و FCM**

سعید رستگار

s.rastegar36@gmail.com

حسین میار نعیمی

h\_miare@nit.ac.ir

چکیده:

آنالیز خوشه‌بندی برای کلاسه بندی مجموعه ای از داده ها به زیرگروه های مشابه بکار می رود. اساس آن بر مبنای روشهای مبتنی بر یادگیری بدون مربی است و یکی از تکنیکهای مهم در زمینه شناسایی آماری الگوها است. در تحقیقات و مقالات مختلفی که تا کنون در این زمینه صورت گرفته، نشان داده شده است که الگوریتمهای خوشه‌بندی که از معیارهای فازی استفاده می‌کنند، مزیت آشکاری نسبت به نمونه‌های قطعی (*Crisp*) دارند. در خوشه بندی فازی (نرم) هر نمونه به چندین کلاستر با درجات عضویت مختلف اختصاص می‌یابد. اکثر الگوریتم‌های خوشه‌بندی فازی بر مبنای روش *fuzzy c-mean (FCM)* می‌باشند. با وجود مزیت *FCM* در شناسایی دسته‌ها نسبت به روش سخت نظیر یعنی *hard C-Mean (HCM)*، نیاز به دانستن مراکز اولیه و تعداد خوشه‌ها از معایب روش فازی است چراکه معمولاً این مراکز اولیه به صورت دلخواه مقاردهی می‌شوند و وجود این مقادیر اولیه متفاوت ممکن است باعث رسیدن تابع هدف فازی به نقاط مینیمم محلی متفاوت شوند. بنابراین با این امید که بعضی از اجراها به نقاط مینیمم گلوبال برسند، خوشه‌بندی با مقادیر اولیه متفاوتی تکرار می‌شود. لذا در این مقاله الگوریتم جدید و موثری برای خوشه بندی فازی بدون احتیاج به پیش فرض مراکز و تعداد خوشه‌ها پیشنهاد شده‌است که می‌تواند با دقت بالایی خوشه بندی را به سمت خوشه بندی بدون مربی (*unsupervised learning*) سوق دهد. با استفاده از معیارهای سنجش خوشه‌بندی فازی، یافتن مقدار بهینه خوشه‌ها توسط الگوریتم ترکیبی حاصل می‌شود. این الگوریتم بر اساس استراتژی خوشه بندی دو لایه پایه گذاری شده است، در لایه اول یک جستجوی بدون مربی به دنبال مراکز اولیه خوشه‌ها با استفاده از قاعده نزدیکترین همسایگی (*K-NN*) مورد استفاده قرار می‌گیرد. سپس برای بدست آوردن خوشه بندی بهینه، دومین لایه برخلاف روش متداول که تکرارهای متوالی تا همگرایی دارد، شامل فقط یک تکرار از *FCM* است. الگوریتم جدید پیشنهاد شده و روش *FCM* متداول را روی 6 مجموعه داده استاندارد، که برای ارزیابی و مقایسه این قبیل الگوریتمها به کار می‌روند، پیاده‌سازی کرده‌ایم. نتایج بدست آمده نشان می‌دهد که الگوریتم جدید نسبت به *FCM*، دقت بیشتری هم در یافتن تعداد واقعی کلاسترها و هم در اختصاص درست نمونه‌ها به هر کلاستر دارد..

کلمات کلیدی

الگوریتم ترکیبی؛ خوشه بندی نرم؛ یادگیری بدون مربی؛ قاعده نزدیک ترین همسایگی