نهمین سمپوزیوم پیشرفت های علوم و تکنولوژی
مشهد- ۱۳ آذر- ۱۳۹۳
کمیسیون دوم: همایش ملی علوم و مهندسی کامپیوتر
با محوریت
امنیت ملی و توسعه پایدار
9th SASTech Symposium on Advances in Science & Technology (2014)
برگزارکننده: موسسه آموزش عالی خاوران
www.khi.ac.ir

بهبود خوشه‌بندی K-Means با الگوریتم مگس میوه

جواد حمیدزاده[۱]، علی زمانی خلیل‌آباد[۲]، علی آرچین[۳]

[۱]استادیار ، دانشکده مهندسی کامپیوتر و فناوری اطلاعات ، دانشگاه صنعتی سجاد ،مشهد، ایران
ایمیل: j_hamidzadeh@sadjad.ac.ir
[۲]دانشجوی کارشناسی ارشد ، مهندسی کامپیوتر نرم‌افزار ، دانشگاه آزاد اسلامی، فردوس، ایران
ایمیل : ali.zamani78@gmail.com
[۳]دانشجوی کارشناسی ارشد ، مهندسی کامپیوتر نرم‌افزار ، دانشگاه آزاد اسلامی، فردوس، ایران
ایمیل : aliarchin54@gmail.com

آدرس پست الکترونیکی نویسنده رابط : ali.zamani78@gmail.com

نام ارائه دهنده : علی زمانی خلیل اباد

کدمقاله:

خلاصه

خوشه‌بندی داده‌ها به کلاس‌ها و یا دسته‌های متناسب، یکی از مباحث مهم و مطرح در تشخیص الگوست. آنچه در خوشه‌بندی حائز اهمیت است انجام این کار به‌گونه‌ایست‌که، داده‌هایی که درست طبقه‌بندی نشده‌اند به حداقل برسند و یا به‌عبارت‌دیگر در هرکلاس داده‌هایی قرار بگیرند که حداکثر نزدیکی و مشابهت را با هم داشته باشند. هدف این مقاله اینست که باکمک الگوریتم بهینه‌سازی مگس میوه، مدل پیشنهادی جدیدی که آن را FOA-Clustering نام نهاده‌ایم ، جهت بهبود روش K-Means معرفی کنیم. در پایان، روش مزبور بر روی مجموعه‌ای از داده‌ها،آزمایش شده است. نتایج، نشان دهنده برتری روش پیشنهادی ما نسبت به سایرروش‌های مرز دانش است .

کلمات کلیدی:تشخیص الگو، خوشه‌بندی، K-Means، الگوریتم بهینه‌سازی مگس میوه، FOA

۱. مقدمه

یکی از روش‌های حیاتی کنترل و مدیریت داده‌ها، خوشه‌بندی داده‌های با خواص مشابه، درون مجموعه‌ای از دسته‌ها است. خوشه‌بندی فرآیندی است که در آن مجموعه‌ای از اشیاء داده به گروه‌های مجزایی از کلاس، خوشه، تقسیم می‌شوند، به‌طوری‌که اشیای یک کلاس تا حد امکان به یکدیگر شبیه بوده و با اشیاء دیگر کلاس‌ها، متفاوت می‌باشند. خوشه‌بندی درزمینه‌های بسیاری ازجمله در شناسایی الگو، یادگیری ماشین، داده‌کاوی، بازیابی اطلاعات و داده‌ورزی زیستی کاربرد دارد. این روش ابزاری برای طبقه‌بندی غیر نظارت‌شده داده‌ها است، در حقیقت خوشه‌بندی یک تکنیک جداسازی بدون ناظر است که معمولاً داده‌ها را به‌صورت بردارهایی در یک فضای چندبُعدی دریافت کرده و آن‌ها را در دسته‌ها یا خوشه‌هایی قرار می‌دهد. با این ویژگی که داده‌های یک خوشه از یک تشابهی سود می‌برند درحالی‌که داده‌های خوشه‌های مختلف، بالعکس، تشابهی باهم ندارند. بنابراین باید به دنبال معیاری برای مشابهت دسته‌ای از داده‌ها باشیم به‌گونه‌ای که این داده‌ها در خوشه یکسانی قرار گیرند. هدف اصلی خوشه‌بندی K-Means این است که مجموع عدم تشابه بین تمام اشیاء یک خوشه از مراکز خوشه‌های متناظرشان کمترین باشد. مهم‌ترین مشکل الگوریتم K-Means [۱] این است که نتایج

$$D = \|X - Z\| \qquad (۱)$$