



خوشه بندی داده ها، مروری بر روش های موجود و مقایسه عملکرد آن ها

محمود شاکری

دانشگاه جامع علمی کاربردی واحد جهاد دانشگاهی مشهد
mah_shakeri@yahoo.com

محمد عبدالمهدی

دانشگاه جامع علمی کاربردی واحد جهاد دانشگاهی مشهد
mabdilahi512@yahoo.com

چکیده

یکی از مهمترین اعمال در داده کاوی خوشه بندی داده های موجود در یک دیتاست می باشد. این تکنیک به دنبال کشف ساختارهایی بوده که منجر به گروه بندی نمونه های موجود در یک پایگاه داده بوده، به گونه ای که نمونه های مشابه درون دسته هایی که بیشترین شباهت را با هم داشته قرار گرفته، در حالی که دارای تفاوتی قابل قبول با نمونه های سایر گروه ها داشته باشند. الگوریتم های خوشه بندی را می توان به چند دسته کلی الگوریتم های خوشه بندی تفکیکی، سلسله مراتبی، مبتنی بر چگالی و مبتنی بر گرید تقسیم نمود. این تحقیق ابتدا به مرور روشهای خوشه بندی مطرح پرداخته و چند الگوریتم از هر روش را معرفی کرده است. در ادامه همچنین به معرفی روشهای دیگری از جمله روشهای خوشه بندی مبتنی بر الگوریتم های فرا ابتکاری و خوشه بندی های آنلاین پرداخته شده و سپس چالشهای موجود در خوشه بندی از قبیل انتخاب تعداد بهینه خوشه ها، کاهش ابعاد، مدیریت داده های پرت و روشهای مقابله با آنها معرفی شده اند. در نهایت روشها و الگوریتم های موجود از نظر برخی از پارامترها مقایسه شده و به بررسی مزایا و معایب هر الگوریتم پرداخته شده است.

واژگان کلیدی: داده کاوی، خوشه بندی، روشهای خوشه بندی، مقایسه خوشه ها، معیار بهینگی