

عنوان مقاله:

خوشه بندی با استفاده از انتخاب زیرمجموعه ویژگی خودکار و تعیین تعداد خوشه های تطبیقی

محل انتشار:

پنجمین کنفرانس ملی محاسبات توزیعی و پردازش داده های بزرگ (سال: 1398)

تعداد صفحات اصل مقاله: 10

نویسندگان:

حبیب ایزدخواه - استادیار دانشکده ریاضی، دانشگاه تبریز، دانشکده علوم ریاضی، گروه علوم کامپیوتر

صابر شیری پورچرلو - دانشجوی کارشناسی ارشد،، موسسه آموزش عالی دانشوران تبریز، دانشکده کامپیوتر و فناوری اطلاعات

خلاصه مقاله:

حجم و سرعت و تنوع داده های تولیدی توسط آزمایشها علمی و تجربی، به صورت چشمگیری در حال افزایش است. استخراج دانش نهفته در این نوع از داده های که به داده های عظیم مشهور است، با استفاده از روشهای سنتی داده کاوی از دقت کافی برخوردار نیست و نیاز به روشهای جدید احساس میشود. خوشه بندی های بزرگ یکی از محبوب ترین روشهایی است که باهدف تقسیم یک گروه از نمونه داده های بدون برچسب در زیرگروه (خوشه)، به طوریکه نمونه های داده متعلق به یک خوشه شبیه به یکدیگر و به نمونه داده های متعلق به خوشه های دیگر غیرمشابه باشند، ارائه شده است. در این تحقیق برای غلبه بر مشکلات خوشه بندی های بزرگ روش جدیدی پیشنهاد شده است که از انتخاب زیرمجموعه ای از ویژگیها بر اساس انحراف معیار داده های عددی و پراکندگی در داده های ترتیبی بهره میبرد. صفت های عددی که دارای انحراف معیار کمی هستند از توزیع نامتوازی برخوردارند و تاثیر چندانی در تعیین گروه داده ها ندارند. همچنین ویژگیهای ترتیبی با پراکندگی کم داده ها باعث میشود که بیشتر نمونه ها در گروه یکسانی قرار گیرند که دقت خوشه بندی را کاهش میدهد. نتایج آزمایشها نشان داده است روش پیشنهادی دقتی در حدود 95% برای خوشه بندی مجموعه داده های بزرگ داشته و با روشهای پیشین معرفی شده در نشریات قابل مقایسه هست.

کلمات کلیدی:

داده کاوی، خوشه بندی، انتخاب زیرمجموعه ویژگی خودکار، خوشه بندی تطبیقی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/961874>

