

عنوان مقاله:

الگوریتم خوشه بندی دادههای بزرگ در داده کاوی

محل انتشار:

اولین همایش ملی علوم کاربردی و مهندسی (سال: 1395)

تعداد صفحات اصل مقاله: 22

نویسنده:

فتاح بیات - گروه ریاضی واحد ملایر دانشگاه آزاد اسلامی ملایر ایران مربی ریاضی کاربردی دانشگاه آزاد اسلامی ملایر ایران

خلاصه مقاله:

کاربردهای زیادی وجود دارند که در آنها خوشه‌بندی مجموعه الگوهای بزرگ ضروری است. اکثر تقریبها و الگوریتمهای تعریف شده توانایی دستکاری این مجموعه دادههای بزرگ را ندارند. در این مقاله به بررسی مزایا و معایب هر یک از الگوریتمهای خوشه‌بندی برای دادههای حجیم نظیر COBWEB، CURE، CLARANS، BIRCH، پرداخته شده است. سپس یک روش پیشنهادی جدید خوشه بندی برای دادههای بسیار بزرگ ارائه شده است و هدف این بوده است که مناسب برای حالتی باشد که امکان بارگذاری همزمان همه دادهها در حافظه وجود ندارد و دادهها بصورت بلاکهای متوالی بارگذاری میشوند. این الگوریتم از سه فاز اصلی تشکیل شده است، در فاز نخست فضای کل دادهها به ناحیه‌های یکسان تقسیم‌بندی می‌شود، در فاز دوم ناحیه در برگیرنده هر داده مشخص شده و دادههای تعلق یافته به هر ناحیه با کمیت‌های حداقلی به صورت فشرده ذخیره میشوند و هر ناحیه متناظر با یک خوشه اولیه میباشد. همچنین در فاز سوم، بعضی از خوشه‌های اولیه بصورت سلسله‌مراتبی با هم ترکیب شده و خوشه‌های نهایی را تشکیل میدهند. برای ارزیابی کارایی روش پیشنهادی، مجموعه داده‌های مصنوعی تولید شده و الگوریتم به این داده‌ها اعمال شده است و نتیجه خوشه‌بندی خوبی حاصل شده است.

کلمات کلیدی:

داده‌کاوی، خوشه‌بندی، مجموعه دادههای بزرگ

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/616610>

