

عنوان مقاله:

الگوریتم ++K-means با پیشنهادی با روش نگاشت - کاهش تکراری نموی جهت بهبود مقیاس پذیری و افزایش کارایی داده کاوی داده های بزرگ

محل انتشار:

هشتمین کنفرانس بین المللی فناوری اطلاعات، کامپیوتر و مخابرات (سال: 1398)

تعداد صفحات اصل مقاله: 18

نویسندگان:

مراد درخشان - استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، تهران، ایران

جاسم قاسمی نژاد - دانشجوی کارشناسی ارشد رشته نرم افزار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه پیام نور، تهران، ایران

خلاصه مقاله:

با توسعه ی سریع شبکه ها، رسانه های ذخیره سازی و افزایش گنجایش مجموعه های داده های بزرگ در تمام حوزه های علم و فناوری از جمله علوم فیزیکی، زیستی و پزشکی به سرعت در حال رشد و گسترش است. برای پشتیبانی از داده کاوی داده های بزرگ به سکوی پردازشی با عملکرد بالا مورد نیاز است تا بتواند قدرت داده های بزرگ را رها سازد. یک سیستم باید به دقت طراحی شود تا بتوان داده های بدون ساخت یافته با روابط پیچیده شان را به الگوهای مفیدی تبدیل کرد. به کارگیری یک زیرساخت موازی، پشتیبانی از زبان برنامه نویسی موازی، مدل های نرم افزاری برای تجزیه و تحلیل داده کاوی توزیع شده از اهداف بسیار مهم برای پردازش داده های بزرگ است. در حال حاضر، پردازش داده های بزرگ وابسته به مدل های برنامه نویسی موازی همچون نگاشت کاهش است. زمانی که نگاشت کاهش با داده های بزرگ سروکار دارد، با موانع و چالش هایی مواجه است. نگاشت کاهش برای الگوریتم های تکرارشونده ی داده کاوی همچون الگوریتم خوشه بندی K-means، مناسب نیست. هدف پژوهشی پایان نامه، بهبود مقیاس پذیری و کارایی الگوریتم ++K-means برای کاهش داده های بزرگ است. ++K-means مراکز را به ترتیب انتخاب میکند تا به یک راه حل بهینه ی قابل اثباتی برسد اما به علت مقیاس پذیری ضعیفی که دارد با افزایش داده ها، بازده اش کم میشود. در الگوریتم ++K-means تمام رکوردهای ساختار به تمام رکوردهای وضعیت وابسته اند. نگاشت کاهش تکراری نموی از روابط بین وضعیت و ساختار پشتیبانی میکند. در حقیقت، با به کارگیری نگاشت کاهش تکراری نموی، هزینه های شروع کار و به دنبال آن تعدادی از کارهای نگاشت کاهش به شدت کاهش مییابد. الگوریتم ++K-means پیشنهادی با روش نگاشت کاهش تکراری نموی مقیاس پذیری و کارایی بالایی دارد و از این پس میتوانیم به راحتی نتایج داده کاوی را همواره به روز نگه داریم.

کلمات کلیدی:

داده کاوی، داده های عظیم، نگاشت کاهش، نگاشت کاهش تکراری نموی، الگوریتم ++K-means

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1010132>

