

عنوان مقاله:

داده کاوی بزرگ مقرون به صرفه در زمینه ابر: یک مطالعه موردی با K-means

محل انتشار:

شانزدهمین کنفرانس ملی علوم و مهندسی کامپیوتر و فناوری اطلاعات (سال: 1401)

تعداد صفحات اصل مقاله: 16

نویسندگان:

اسماعیل جهانگشته - استاد دانشگاه آزاد واحد ایرانشهر

پرویز گزمه - استاد دانشگاه آزاد واحد ایرانشهر

حسینعلی رفیعی - دانشجوی مهندسی نرم افزار کامپیوتر، دانشگاه آزاد واحد ایرانشهر

خلاصه مقاله:

کاوش داده ی بزرگ، اغلب نیازمند منابع محاسباتی فوق العاده می باشد. این امر به یک مانع عمده در رابطه با استفاده ی وسیع از تجزیه و تحلیل داده های بزرگ تبدیل شده است. محاسبات ابری به محققانی که در زمینه ی داده فعالیت می کنند، اجازه ی دسترسی به منابع محاسباتی، بر اساس تقاضای ساخت راه حل های تحلیلی داده ای بزرگ در ابر را می دهد. هر چند، هزینه ی پولی کاوش داده ی بزرگ در ابر، هنوز هم می تواند بر خلاف انتظار مانع، بالا باشد. برای مثال، اجرای مثال های ۱۰۰ m4- xlarge Amazon EC2 به مدت یک ماه هزینه ای در حدود \$۱۷,۴۹۵,۰۰۰ را به دنبال دارد. در این زمینه، مسئله ی حیاتی به منظور تجزیه و تحلیل هزینه ی بهره وری (هزینه ی مقرون به صرفه) داده کاوی بزرگ در ابر، چگونگی دستیابی به یک نتیجه ی رضایت بخش کافی با حداقل هزینه ی محاسباتی ممکن است. در سناریو های داده کاوی بزرگ حقیقی، دقت ۱۰۰٪ غیر ضروری است. در عوض، اغلب، دستیابی به یک دقت کافی، برای مثال، ۹۹٪ یا هزینه ی کمتر مانند، ۱۰٪ نسبت به هزینه ی دستیابی با دقت، ۱۰۰٪ ترجیح داده می شود. در این مقاله، ما به کشف و نمایش داده کاوی بزرگ مقرون به صرفه به همراه یک مطالعه ی موردی و با استفاده از K-means اقدام می کنیم. با استفاده از مطالعه ی موردی، در می یابیم که دست یابی به دقت ۹۹٪ تنها نیاز به هزینه ی محاسباتی ۰.۳۲٪-۴۶.۱۷٪ مربوط به دقت ۱۰۰٪ دارد. این یافته، سنگ بنای لازم را برای داده کاوی مقرون به صرفه در انواع دامنه ها قرار می دهد.

کلمات کلیدی:

محاسبات ابری؛ داده کاوی؛ مقرون به صرفه؛ داده بزرگ؛ K-means

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1513577>

