

عنوان مقاله:

کاربرد روش های آماری در تحلیل و پردازش داده های بزرگ

محل انتشار:

چهارمین کنفرانس ملی محاسبات توزیعی و پردازش داده های بزرگ (سال: 1397)

تعداد صفحات اصل مقاله: 10

نویسنده:

آسیه ابطی - استادیار، گروه ریاضی و آمار، واحد شیراز، دانشگاه آزاد اسلامی، شیراز، ایران

خلاصه مقاله:

با گسترش علم در زمینه های مختلف، هر روز با انبوهی از اطلاعات مواجه خواهید شد. ذخیره سازی و نتیجه گیری از این اطلاعات به دست آمده خود سبب پیدایش علوم جدیدی شده است. داده بزرگ و تحلیل داده بزرگ یکی از این موارد است که در دهه ی اخیر مورد توجه محققان زیادی قرار گرفته است. داده های بزرگ به عنوان مجموعه داده هایی شناخته شده است که توانایی برنامه های کاربردی و تکنولوژی های معمول را در مدیریت و تجزیه و تحلیل دچار چالش کرده است. وجود هرگونه خطا و اشتباهی در تحلیل داده ها می تواند به نتیجه گیری های نادرستی منجر گردد. استفاده از روش های آماری زمینه های علمی، برای تجزیه و تحلیل داده های عینی را فراهم می سازد و منجر به ارایه ی نتایج با معنا و معتبر از اطلاعات جمع آوری شده می گردد. حجم، مقیاس و ساختار داده های حال حاضر چالش های آماری ایجاد کرده است. برای درک قوی از اطلاعات موجود در داده هایی با حجم زیاد استفاده از روش های کلاسیک آماری به تنهایی کارا نیست. ایجاد ارتباط بین روش های آماری و علوم رایانه ای، سبب می شود که روش های کارا و سریعی برای استخراج اطلاعات نهفته ی مجموعه ای بزرگ از داده ها ارایه گردد و منجر به طیف گسترده ای از بینش ها و منافع شود. با کمک آمار و سایر رشته های وابسته به آن مانند داده کاوی، یادگیری ماشین، شبکه های عصبی، تحلیل شبکه های اجتماعی، پردازش سیگنال، تشخیص الگو، روش های بهینه سازی و تجسم رویکردها می توان اطلاعات نهفته ی مجموعه ای بسیار بزرگ از داده ها را به دست آورد. ترکیب مدل سازی لگاریتم خطی به عنوان یک روش آماری با رده بندی درخت رگرسیونی که از تکنیک های یادگیری ماشین یکی از روش هایی است که در تحلیل داده های رسته ای نقش مهمی دارند و در این تحقیق مورد بررسی قرار داده شده است.

کلمات کلیدی:

داده بزرگ، مدل سازی لگاریتم خطی، نظریه یادگیری، درخت رده بندی رگرسیونی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/772503>

