کاهش ابعاد از طریق متغیرهای مهم

خلاصه مطالب

بسیاریاز مجموعه داده ا با مقیاس وسیع نیاز دارند که از لحاظ ابعادی کاهش پیدا کنند تا پژوهش و تحلیل بیشتر بتواند رخ دهد. متغیرهای مهم زیر مجموعه ای از متغیرهای اصلی هستند که تا اندازه ای ساخت و اطلاعاتی را که در متغیرهای اصلی وجود دارد را حفظ می کنند. کاهش ابعاد با استفاده از متغیرهای اصلی بررسی می شود و یک الگوریتم جدید را برای تعیین یک چنین متغیر مهمی پیشنهاد می دهد. این روش تست شده است و با ۱۱ متغیر دیگر در روش های انتخابی پژوهش های مشابه، مقایسه شده و مشخص شده است که به شدت میثر می باشد. همچنین الحاقات این روش نیز توسعه یافته اند که شامل روشی برای تعیین متغیرهای مهم خطی مقیاس های مکرر و تکنیکی برای ایجاد فوایدی به منظور تغییر فرآیند گزینش می باشد. همچنین این روش از طریق داده های واقعی توضیح داده می شود که شامل یک داده انگلیسی بزرگتر است که مربوط به خروجی حاصل شده پس از پیوند کلی زانو می باشد.

۱-مقدمه

تمرکز این مقاله روی کاهش ابعد داده های چند متغیری می باشد که به وسیله نیاز برای حفظ تنها یک زیر مجموعه از متغیرهای اصلی و به منظور بررسی بیشتر محدود شده اند ما به پیروی از (Mccabe 1984) یک چنین مجموعه ای را تحت معیارهای مشخص شده متغیرهای مهم مجموعه داده ها می نامیم. مزیت ویژه کاربرد PVs برای کاهش ابعاد این است که همه متغیرهای باقی مانده می توانند کنار گذاشته شوند. تحلیل اجزای مهم (PCA) به شدت در تهیه یک ارائه کاهش ابعاد مؤثر است اما به طور نمونه، هنوز هم مستلزم همه متغیرهای اصلی می باشد. Mccabe (1984)دراین متن راجع به نقایص شیوه های PCA را تحت نظرتعدادی از معیارهای بهینه سازی،مهیا می کند. برای مثال انتخاب زیر مجموعه های متغیر اصلی برای برنامه های کاربردی در مدل سازی پس رفت، یک تاریخچه طولانی و متداوم دارد. (krzanowki&marrot 1994) دیدگاهی که در اینجا توصیف می شود به هیچ برنامه کاربردی مشابهی ارتباط نداشت اما بوسیله نیاز برای کاهش ابعادی قبل از اجرای مدل سازی نموداری خطی برانگیخته شد. چنین الزامی به صورت فزاینده از خصوصیات بسیاری از حوزه های کاربردی، به ویژه آنهایی که در مجموعه دادگان بسیار بزرگ با هزاران متغیر،دخیل هستند، می باشد.