

عنوان مقاله:

افزایش دقت در تجمیع داده های بلادرنگ بزرگ با استفاده از کاهش ویژگی غیرموثر

محل انتشار:

فصلنامه جهان مهارت، دوره 1، شماره 1 (سال: 1399)

تعداد صفحات اصل مقاله: 8

نویسنده:

صبا چهل امیران - دپارتمان مهندسی برق و کامپیوتر، دانشکده زینب کبری، دانشگاه فنی و حرفه ای استان همدان، همدان، ایران

خلاصه مقاله:

اصطلاح داده های بزرگ برای داده های محاسباتی و یا اطلاعاتی که می تواند تجزیه و تحلیل شود یا برای استفاده از تکنیک ها و ابزارهای سنتی یادگیری ماشین به کار رود، مفید است. تعریف کلی از داده های بزرگ نشان می دهد که داده های محاسباتی بیش از حد سریع، گسترده و یا بیش از حد برای پردازش سخت هستند. مسئله انتخاب ویژگی، یکی از مسائلی است که در مبحث یادگیریماشین و همچنین شناسایی الگوی آماری مطرح است. این مسئله در بسیاری از کاربردها (مانند طبقه بندی) اهمیت زیادی دارد، زیرا در این کاربردها تعداد زیادی از ویژگی ها وجود دارد که بسیاری از آنها یا بلااستفاده هستند و یا اینکه بار اطلاعاتی چندانی ندارند. حذف نکردن این ویژگی ها مشکلی از لحاظ اطلاعاتی ایجاد نمی کند، ولی بار محاسباتی را برای کاربرد مورد نظر بالا می برد؛ و علاوه براین باعث می شود که اطلاعات غیرمفید زیادی را به همراه داده های مفید ذخیره گردد. در این پژوهش با استفاده از استخراج ویژگی و الگوریتم ژنتیک روشی ارائه شده است تا تحلیل نظرات کاربران درباره فیلم های سینمایی را انجام گیرد، در این روش با انتخاب ویژگیها نظر در قالب یک برداری عددی ذخیره می شود و سپس توسط الگوریتم ژنتیک نوع جمله از نظر مثبت و منفی بودن مشخص شده و تحلیل آماری انجام می شود.

کلمات کلیدی:

انتخاب ویژگی، الگوریتم ژنتیک، نظردکاوی، شبکه های اجتماعی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1351976>

