

عنوان مقاله:

موازنه بین سرعت و دقت دسته بندی صفحات وب در انگشت نگاری صفحات وب

محل انتشار:

نهمین کنفرانس فناوری اطلاعات و دانش (IKT 2017) (سال: 1396)

تعداد صفحات اصل مقاله: 8

نویسندگان:

مریم طائی - دانشگاه اصفهان

علی بهلولی - دانشگاه اصفهان

مرجان کائدی - دانشگاه اصفهان

خلاصه مقاله:

انگشت نگاری وبسایت یک حالت خاص از تحلیل داده های وب است که طی آن می توان بدون رمزگشایی محتوای ترافیک، مرور مجموعه ای از صفحات وب از پیش تعیین شده توسط کاربر را تشخیص داد. این کار با بهره گیری از الگوریتم های یادگیری ماشین و با استفاده از مجموعه ویژگی های استخراج شده از داده های ترافیک کاربر انجام می گیرد. روش های بسیاری برای دسته بندی این نوع ترافیک با کمک گرفتن از مجموعه های متفاوت از ویژگی های ترافیک و الگوریتم های دسته بندی مختلف پیشنهاد شده است. در این روش ها به منظور دستیابی به دقت مطلوب و انعطاف پذیری دسته بندی در شرایط مختلف، مجموعه داده های آموزشی و ویژگی هایی با ابعاد بالا مورد استفاده قرار می گیرند. این مساله منجر به کاهش سرعت آموزش و آزمون دسته بند می گردد. این پژوهش ابتدا روشی برای استفاده از دو دسته بند k نزدیکترین همسایه و شبکه عصبی برای مساله انگشت نگاری وبسایت ارایه می دهد. سپس به بررسی اثر استفاده از روش تحلیل مولفه های اساسی به عنوان یک روش رایج برای کاهش ابعاد داده در عملکرد دسته بندهای فوق می پردازد. به منظور دستیابی به بهترین نتایج، میزان داده لازم برای تشکیل مولفه های اساسی و تعداد مولفه های مناسب برای دستیابی به بهترین دقت ممکن بررسی می شود. نتایج آزمایش ها حاکی از آن است که استفاده از روش تحلیل مولفه های اساسی باعث افزایش نرخ مثبت کاذب و کاهش نرخ مثبت واقعی نسبت به روش اولیه میگردد اما بهبود چشمگیر سرعت آموزش و آزمون دسته بندهای فوق را به دنبال دارد. با توجه به اهمیت سرعت دسته بندی در کاربرد انگشت نگاری، روش پیشنهادی باعث ایجاد موازنه بین سرعت و دقت دسته بند می گردد.

کلمات کلیدی:

انگشت نگاری وبسایت، دسته بندی ترافیک، تحلیل مولفه های اساسی، شبکه های عصبی، k نزدیکترین همسایه

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/727224>

