

عنوان مقاله:

بهبود دقت طبقه‌بندی اسناد با استفاده از ترکیب الگوریتمهای یادگیری ماشین

محل انتشار:

چهارمین کنفرانس بین المللی مهندسی برق و کامپیوتر (سال: 1395)

تعداد صفحات اصل مقاله: 20

نویسندگان:

حدیث شعیری - دانشجویی کارشناسی ارشد رشته نرم افزار، دانشگاه آزاد اسلامی واحد بروجرد

ابراهیم شیری - استادیار دانشکده ریاضی و علوم کامپیوتر دانشگاه صنعتی امیرکبیر

خلاصه مقاله:

باتوجه به داده های عظیم موجود در پایگاه های داده اینترنت و مراکز اطلاعاتی طبقه بندی متن یک نقش مهم در سازمان دهی بازایی و جستجوی اطلاعات موجود در این پایگاه های داده می باشد در واقع می توان گفت طبقه بندی متن اختصاص دادن یک متن به یک طبقه از پیش تعیین شده می باشد برای اختصاص یافتن یک متن به یک طبقه خاص از ویژگیهای متن موجود استفاده میشود و باتوجه به ویژگیهای فراوانی که برای متون وجود دارد یک چالش اساسی انتخاب زیرمجموعه ای از ویژگیهای مناسب برای یک متن می باشد از این رو در این مقاله از فیلترهای مختلفی جهت انتخاب ویژگی از جمله حذف کلمات متوقف کننده n-gram و ... استفاده شده است در این پژوهش از مجموعه داده استاندارد خبرگزاری رویترز Reuters-21578 با 8 دسته مختلف و در مجموع بیش از 7000 سند متنی استفاده شده است از الگوریتم های ترکیبی جهت طبقه بندی استفاده کرده ایم که نتایج حاصل از آزمایشها انجام شده بر روی مجموعه داده نشان میدهد که الگوریتم پیشنهادی بهترین کارایی جهت طبقه بندی متن را داشته و باعث بهبود دقت و بازخوانی جهت طبقه بندی متن شده است باتوجه به نتایج شبیه سازی میزان دقت طبقه بندی اسناد برابر 96.66 درصد شده است

کلمات کلیدی:

طبقه بندی متن ، یادگیری ماشین، الگوریتم های ترکیبی ، فرکانس عبارت عکس فرکانس عبارت، n-gram

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/609136>

