

عنوان مقاله:

روشی جدید در انتخاب ویژگی برای دسته بندی مستندات متنی

محل انتشار:

اولین همایش منطقه ای بهینه سازی و روش های محاسبه نرم در مهندسی برق و کامپیوتر (سال: 1392)

تعداد صفحات اصل مقاله: 7

نویسندگان:

هاجر فرهمند - دانشجوی کارشناسی ارشد، دانشگاه آزاد اسلامی واحد علوم و تحقیقات بوشهر

علی هارون آبادی - استادیار و عضو هیئت علمی، دانشگاه آزاد اسلامی واحد تهران مرکزی گروه کامپیوتر

سید جواد میرعابدینی - استادیار و عضو هیئت علمی، دانشگاه آزاد اسلامی واحد تهران مرکزی گروه کامپیوتر

خلاصه مقاله:

امروزه اطلاعات زیادی در قالب متون الکترونیکی و پایگاه های داده ای متنی ذخیره شده اند. برای استخراج دانش از این حجم بالای اطلاعات متنی، نیازمند استفاده از روش های خوشه بندی و طبقه بندی مستندات متنی هستیم. در اکثر روشهای طبقه بندی متن معمولا براساس فراوانی کلمات و در نظر گرفتن متن به صورت مجموعه ای از کلمات، طبقه بندی انجام می شود ولی چنین نمایشی باعث ابعاد بالای متون، تعداد بسیار زیاد ویژگی ها و منجر به کاهش کارایی الگوریتم های طبقه بندی می گردد. در این مقاله به منظور کاهش تعداد ویژگی ها و انتخاب ویژگی های اصلی متن، از دانش پس زمینه و تکنیکهای یادگیری ماشین استفاده شده است. در واقع با استفاده از هستان شناس وردنت و دانش پس زمینه ویژگی های اصلی متون انتخاب می شوند و با استفاده از الگوریتم های یادگیری ماشین متون طبقه بندی می شوند. نتایج شبیه سازی روش پیشنهادی، نشان می دهد بهبود موثری در کاهش ابعاد متن و در نتیجه افزایش کارایی طبقه بندی متن ایجاد شده است.

کلمات کلیدی:

انتخاب ویژگی، طبقه بندی متن، هستان شناسی، یادگیری ماشین

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/261650>

