

عنوان مقاله:

دسته بندی متون زبان فارسی با استفاده از نگاشت کلمه به بردار

محل انتشار:

کنفرانس ملی سیستم های هوشمند و محاسبات سریع (سال: 1399)

تعداد صفحات اصل مقاله: 16

نویسندگان:

محمدرضا شمس - استادیار گروه مهندسی کامپیوتر، مرکز آموزش عالی شهرضا، شهرضا، ایران

مهدی مرادی - دانشجوی کارشناسی گروه مهندسی کامپیوتر، مرکز آموزش عالی شهرضا، شهرضا، ایران

خلاصه مقاله:

دسته بندی متن یکی از کلیدی ترین و مهم ترین فعالیت های حوزه متن کاوی است. هدف از دسته بندی متن، قرار دادن هر متن در دسته های از پیش تعیین شده بر اساس یادگیری اولیه سیستم است. تحقیقات گسترده ای که در حیطه دسته بندی در زبان های مختلف و به خصوص در زبان انگلیسی انجام شده است. امام فعالیت های انجام شده در زبان فارسی برای این حوزه بسیار محدود و دارای دو چالش اساسی است. اولین چالش این روش ها استفاده از شیوه های سنتی و قدیمی دسته بندی برای زبان فارسی و چالش دوم، ارزیابی روش ارائه شده بر روی مجموعه محدود و کم تعدادی از متون است. از این رو در این مقاله روشی برای دسته بندی متون فارسی بر پایه نگاشت کلمه به بردار به عنوان یکی از حوزه های جدید و کاربردی یادگیری ماشین پیشنهاد می شود و تلاش می شود روش پیشنهادی از جنبه های مختلف و بر روی یک مجموعه 90 هزارتایی از اخبار فارسی ارزیابی و بررسی گردد تا خروجی روش قابل اطمینان و سازگار با شرایط مختلف باشد. در روش پیشنهادی ابتدا پیش پردازش های لازم برای زبان فارسی بر روی مجموعه داده اعمال می گردد. سپس به کمک نگاشت کلمه به بردار، فضای مسئله به یک فضای n بعدی از اعداد نگاشت می شود. در این نگاشت هر کلمه به برداری از اعداد تبدیل می شود به نحوی که کلمات مشابه از نظر برداری به هم نزدیک باشند. سپس این بردارها مبنای آموزش یک دسته بند قرار گرفته و عمل پیشگویی موضوع هر متن طبق این آموزش انجام می شود. نتایج ارزیابی ها نشان دهنده موفقیت روش پیشنهادی در دسته بندی متن های زبان فارسی است.

کلمات کلیدی:

متن کاوی، دسته بندی متن، استخراج ویژگی، شبکه عصبی مصنوعی، نگاشت کلمه به بردار، ماشین بردار، پشتیبان

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1152617>

