

بهبود الگوریتم بهینه‌سازی کلونی مورچه با الگوریتم K نزدیک‌ترین همسایه برای انتخاب

ویژگی در دسته‌بندی اسناد متنی

دلشاد زنده روی^۱، فرهاد سلیمانیان قره‌چپق^{۲*}

۱- گروه مهندسی کامپیوتر، واحد مراغه، دانشگاه آزاد اسلامی، مراغه، ایران

۲- گروه مهندسی کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران

چکیده

امروزه اطلاعات در قالب‌های متفاوت بصورت مفید و غیر مفید روز به روز با سرعت بیشتر در حال تولید می‌باشند. دسته‌بندی این اسناد با وجود داده‌ها و اطلاعات فراوان و ویژگی‌های مختلف بسیار مشکل و وقت گیر می‌باشد. این ویژگی‌ها می‌توانند شامل ویژگی‌های نامرتب و نویز باشند که صحت و دقت دسته‌بندی اسناد را پایین می‌آورند. پس باید به وسیله ابزاری موثر بتوانیم اطلاعات و داده‌ها را آنالیز کرده و ویژگی‌های موثری از آنها استخراج کرده تا بتوانیم اطلاعات را به شکل مطلوب و مفیدی دسته‌بندی نماییم. برای انتخاب ویژگی‌ها از الگوریتم فرا ابتکاری بهینه‌سازی کلونی مورچه استفاده شده است که این الگوریتم دارای دقت و صحت و دوباره‌خوانی خوبی در انتخاب ویژگی و دسته‌بندی متون است ولی دارای عملکرد صد درصدی نمی‌باشد. در این مقاله برای دسته‌بندی اسناد متنی از K نزدیک‌ترین همسایه استفاده شده است. یعنی ابتدا با الگوریتم بهینه‌سازی کلونی مورچه بهترین ویژگی‌ها تک تک انتخاب می‌شوند و سپس با دسته‌بندی K نزدیک‌ترین همسایه دسته‌بندی اسناد انجام می‌گیرد. از مجموعه داده‌های Reuters- 21578 و WEBKB و CADI2 برای ارزیابی روش پیشنهادی استفاده شده است. با توجه به نتیجه حاصل شده میزان دقت دسته‌بندی، بطور قابل ملاحظه‌ای افزایش یافته و کارایی دسته‌ها نیز به میزان قابل توجهی بهبود یافته است.

واژگان کلیدی: دسته‌بندی متن، انتخاب ویژگی، الگوریتم بهینه‌سازی کلونی مورچه، الگوریتم K

نزدیک‌ترین همسایه