

انتخاب ویژگی ها از صفحات وب مبتنی بر تئوری مجموعه ناهموار و اتوماتای یادگیر

توزیع شده

بابک نصیری^۱؛ فریبرز محمودی^۲؛ محمدرضا میبیدی^۳

چکیده

صفحات وب حجم انبوهی از اطلاعات را در خود جای داده اند، اما برای وب کاوی بسیاری از آنها زائد و اضافی می باشند. از این رو لازم است بعنوان پیش پردازش، ویژگی های مناسب از بین این حجم اطلاعات استخراج شود. از آنجا که انتخاب ویژگی های مناسب یک مسئله NP-hard بشمار می رود، جستجو برای الگوریتم های تقریبی سریع و کارا همچنان ادامه دارد. در این مقاله یک روش ترکیبی جدید مبتنی بر تئوری مجموعه ناهموار^۱ و اتوماتای یادگیر توزیع شده برای انتخاب ویژگی های مناسب در صفحات وب ارائه شده است. نتایج حاصل از پیاده سازی روش پیشنهادی بر روی چندین مجموعه داده از جمله یک مجموعه مبتنی بر وب، حکایت از کارایی روش پیشنهادی در مقایسه با سایر روشهای شناخته شده دارد.

کلمات کلیدی

انتخاب ویژگی ها، وب کاوی، اتوماتای یادگیر توزیع شده، تئوری مجموعه ناهموار، محتوی کاوی وب.

۱ دانشکده برق، کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی، قزوین، ایران،
nasiri_babak@yahoo.com

۲ دانشکده برق، کامپیوتر و فناوری اطلاعات، دانشگاه آزاد اسلامی، قزوین، ایران،
mahmoudi@qazviniau.ac.ir

۳ دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه امیر کبیر، تهران، ایران،
mmeybodi@aut.ac.ir