

هرس دسته‌بند تجمیعی با استفاده از انتخاب پیشروانه

الگوریتم جستجوی ممنوعه

زهره سادات تقوی^۱، هدیه ساجدی^۲، محمد کلانتری^۳

^۱ کارشناس ارشد، دانشگاه آزاد اسلامی، واحد قزوین، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، قزوین، ایران
zs.taghavi@qiau.ac.ir

^۲ استادیار، دانشگاه تهران، پردیس علوم، دانشکده ریاضی، آمار و علوم کامپیوتر، تهران، ایران
hhsajedi@ut.ac.ir

^۳ استادیار، دانشگاه تربیت دبیر شهید رجایی، دانشکده مهندسی برق و کامپیوتر، تهران، ایران
mkalantari@srutu.edu

چکیده

هرس تجمیع یکی از مؤثرترین عملیات برای افزایش عملکرد پیش‌بینی و کاهش سربار محاسباتی روش‌های تجمیعی در حیطه یادگیری ماشین است. این عملیات که یکی از مسائل *NP-Complete* می‌باشد، توجه محققان زیادی را در این حیطه به خود جلب کرده است. در این مقاله، یک روش هرس دسته‌بند تجمیعی جدید پیشنهاد می‌شود که با استفاده از انتخاب پیشروانه الگوریتم‌های جستجوی ممنوعه، بهترین مدل‌ها را به صورت گام‌به‌گام از تجمیع اولیه انتخاب می‌کند تا در انتها بهترین زیرمجموعه از آن به دست آید. به علاوه در این روش، یک استراتژی جدید و اثرگذار با نام "فراموشی دوره‌ای" که خاص لیست ممنوعه است، پیشنهاد می‌شود. روش پیشنهادی با دو روش پیشرفته مشابه برای هرس یک تجمیع اولیه ناهمگن و روی ۱۰ مجموعه داده استاندارد، مورد مقایسه و ارزیابی قرار می‌گیرد. این آزمایش‌ها براساس معیار دقت تجمیع هرس شده، اندازه آن و میانگین زمان اجرا تحلیل می‌شوند. نتایج آزمایش‌ها نشان می‌دهد که روش پیشنهادی در مقایسه با دو روش مشابه، به طور میانگین منتهی به ۳/۰۶۹٪ بهبود دقت می‌شود. به علاوه، روش پیشنهادی در کاهش اندازه تجمیع اولیه و یافتن مدل‌های دقیق‌تر آن، موفق‌تر است که یکی از دلایل این امر می‌تواند وقت بیشتری باشد که برای جستجوی مدل‌ها صرف می‌کند.

کلمات کلیدی

دسته‌بند تجمیعی، هرس دسته‌بند تجمیعی، انتخاب تجمیع، الگوریتم جستجوی ممنوعه، استراتژی مربوط به لیست ممنوعه.

۱- مقدمه

دارند که نمیتوان تنها از یک دسته‌بند منفرد برای دسته‌بندی آنها استفاده کرد [4]. به عنوان مثال در مسائلی که در آنها مجموعه داده در مکان‌های مختلف به صورت فیزیکی توزیع شده است، نمیتوان از یک دسته‌بند منفرد برای امر دسته‌بندی بهره گرفت؛ یا در کاربردهایی، مجموعه داده آنقدر بزرگ است که با استفاده از یک دسته‌بند پایه پیچیدگی محاسباتی و مشکلات مربوط به حافظه تحمیل می‌گردد. با توجه به مسائل مطرح، گونه دیگری از الگوریتم‌های دسته‌بندی با نام روش "دسته‌بند تجمیعی"^۲ یا "روش ترکیب دسته‌بندها" [10] مطرح شد. این روش امکان ترکیب نتایج دسته‌بندی الگوریتم‌های دسته‌بندی مختلف را فراهم کرده و موجب می‌شود که دسته‌بندها خطای یکدیگر را بپوشانند. این روش با وجود اینکه راه حل بسیاری از

یکی از مراحل اصلی داده‌کاوی بعد از استخراج و آماده‌سازی داده‌ها، تزریق آنها به یک سری الگوریتم‌های یادگیری برای ساخت مدل است تا براساس آن، دانش مورد نظر از داده‌ها استخراج شود. یکی از مطرح‌ترین و پرکاربردترین الگوریتم‌های یادگیری، روش دسته‌بندی^۱ است. نمونه‌هایی از الگوریتم‌های دسته‌بندی عبارتند از درخت تصمیم، شبکه عصبی مصنوعی، ماشین بردار پشتیبان، و روش k -نزدیکترین همسایه [3، 4]. هر یک از این الگوریتم‌ها، در کنار ویژگی‌های خوب و منحصر به فردشان، دارای معایبی هستند که موجب می‌شود نتوان همیشه از آنها برای تمامی مسائل استفاده کرد. به علاوه، برخی از کاربردها وجود