

یادگیری بر خط متریک محلی مبتنی بر روش فعال/غیر فعال

بیدا حمدان^۱، دکتر داود ذبیح زاده^۲، دکتر رضا منصفی^۳

۱- دانشجوی کارشناسی ارشد گروه کامپیوتر دانشکده فنی و مهندسی دانشگاه فردوسی مشهد

۲- عضو هیات علمی گروه کامپیوتر موسسه آموزش عالی اسرار مشهد

۳- دانشیار گروه کامپیوتر دانشکده فنی و مهندسی دانشگاه فردوسی مشهد

چکیده

تعیین شباهت/فاصله داده‌ها در بسیاری از الگوریتم‌های یادگیری ماشین، شناسایی الگو و داده کاوی استفاده می شود. معیار شباهت/فاصله بستگی به داده‌ها دارد، بنابراین ممکن است یک معیار برای یک کاربرد مناسب ولی برای کاربرد دیگر نامناسب باشد، همین مسئله منجر به پیدایش روش‌های یادگیری متریک شده است. در بسیاری از مجموعه داده‌ها که میزان تمایز ویژگی‌ها در نواحی مختلف داده متفاوت است، استفاده از روش‌های سراسری یادگیری متریک، کارایی مطلوبی ندارد از این رو روش‌های محلی ارایه شده اند که چندین نگاشت خطی یا غیرخطی در نواحی مختلف فضای داده یاد می گیرند. از مزایای این روش‌ها می توان به انعطاف پذیری بالا، یادگیری نگاشت‌های غیرخطی، مناسب بودن برای مسائل پیچیده مانند داده‌های ناهمگن اشاره کرد. با این وجود بدلیل تعداد زیاد پارامترهای یادگیری، باید برای مشکلات بیش برآزش و مقیاس پذیری اندازه مجموعه داده‌های آموزشی راه کارهای مناسب در نظر گرفته شود. همچنین مشکل نامتوازن بودن داده‌های کلاس‌های مختلف وجود دارد که باید راه کار مناسب برای این مساله ارایه داد. در این مقاله برای حل این مشکلات، روشی برخط یادگیری متریک محلی ارایه شده است، که با یادگیری متریک سراسری و محلی بصورت توأم، سعی می کند از مشکل عدم توازن داده‌ها در نقاط مختلف جلوگیری کند. روش پیشنهادی مبتنی بر PA (Passive/Aggressive) است و در هر مرحله با دریافت نمونه آموزشی، متریک‌های محلی لازم و متریک سراسری را بهنگام می کند. برای تعیین فاصله بین دوشی از ترکیب متریک محلی و سراسری استفاده می شود. نتایج آزمایش‌ها بر روی مجموعه وسیعی از داده‌های مورد آزمایش تأیید می کند که روش پیشنهادی کارایی نسبت به روش‌های همتا در مرزهای دانش برتری دارد.

واژگان کلیدی: یادگیری متریک، یادگیری شباهت، یادگیری محلی، یادگیری برخط، یادگیری فعال/غیر فعال محلی