

عنوان مقاله:

مقاوم سازی روش های یادگیری متریک در مقابل داده های پرت

محل انتشار:

کنگره مشترک سیستم های فازی و هوشمند ایران (نوزدهمین کنفرانس سیستم های فازی و هفدهمین کنفرانس سیستم های هوشمند)
(سال: 1399)

تعداد صفحات اصل مقاله: 6

نویسندگان:

داود ذبیح زاده خواجهی - استادیار گروه فناوری اطلاعات دانشکده مهندسی، دانشگاه فناوری نوین سبزوار، سبزوار ایران

رضا منصفی - استاد گروه مهندسی کامپیوتر دانشکده مهندسی، دانشگاه فردوسی مشهد، ایران

خلاصه مقاله:

در بسیاری از الگوریتم های یادگیری ماشین، شناسایی الگو و داده کاوی نیازمند آن هستیم که شباهت یا فاصله بین داده ها را به روش مناسب اندازه گیری کنیم. بعنوان مثال کارایی الگوریتم های خوشه بندی و یا طبقه بند k نزدیک ترین همسایه به معیار فاصله/شباهت بستگی دارد. معیارهایی عمومی نظیر فاصله اقلیدسی و شباهت کسینوسی که بدون توجه به مفهوم داده ها میزان شباهت یا فاصله آن ها را مشخص می کنند، در بسیاری از کاربردها کارایی مناسبی ندارند. این مساله ضرورت یادگیری متریک را نشان می دهد. در یادگیری متریک هدف این است که با توجه به داده ها معیار شباهت یا فاصله بهینه به دست آید بطوری که داده هایی که از نظر مفهومی و منطقی شبیه به هم هستند، به یکدیگر نزدیک می شوند و داده هایی که از نظر مفهومی و منطقی شبیه نیستند از یکدیگر دور شوند. در این زمینه روش های زیادی ارائه شده است، اما همچنان یکی از چالش های مهم و جذاب، کاهش تاثیر داده پرت یا برچسب نویزی می باشد. در روش ارائه شده، مجموعه داده ورودی هم زمان با یادگیری متریک، به دو بخش داده بدون خطا و بخش داده های پرت تقسیم میشود و یادگیری معیار فاصله تنها بر روی بخش بدون خطا انجام می شود. آزمایشات انجام شده بر روی داده های واقعی (در حضور و عدم حضور داده پرت و برچسب نویزی) کارایی الگوریتم ارائه شده را تایید می کند و برتری آن را نسبت به روش های همتا در مرزهای دانش در محیط های دارای نویز برچسب نشان می دهد.

کلمات کلیدی:

یادگیری متریک، یادگیری متریک مقاوم، یادگیری متریک نزدیک ترین همسایه با حاشیه بزرگ، طبقه بند k نزدیک ترین همسایه، داده های پرت، برچسب نویزی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1193466>

