

عنوان مقاله:

ارائه روشی جدید برای کاهش ابعاد داده های بزرگ

محل انتشار:

پنجمین کنفرانس ملی محاسبات توزیعی و پردازش داده های بزرگ (سال: 1398)

تعداد صفحات اصل مقاله: 8

نویسندگان:

عاطفه عباسی - گروه علوم کامپیوتر، دانشگاه تبریز، تبریز

حبیب ایزدخواه - گروه علوم کامپیوتر، دانشگاه تبریز، تبریز

شهریار لطفی - گروه علوم کامپیوتر، دانشگاه تبریز، تبریز

خلاصه مقاله:

در طی دهه های اخیر، مجموعه های داده های با ابعاد بالا در علوم مختلف به سرعت در حال افزایش هستند. بسیاری از این مجموعه های داده های دارای تعداد زیادی ویژگی به نسبت تعداد کم الگوها هستند. بسیاری از این ویژگیها اغلب نامرتبط و دارای افزونگی هستند که منجر به کاهش عملکرد الگوریتمهای طبقه بندی میشوند. از این رو انتخاب ویژگی، برای کاهش ابعاد مسئله و افزایش کارایی الگوریتمهای طبقه بندی پیشنهاد شده است. یکی از مسائلی که انتخاب ویژگی در آن ضروری است، طبقه بندی متون است. این مسئله در دسته مسائل چندجمله ای غیرقطعی-سخت قرار دارد؛ بنابراین الگوریتم های تکاملی در این حوزه میتوانند قابل استفاده باشند. در حال حاضر، روشهای بسیاری جهت انتخاب ویژگی در متون وجود دارند. در این مقاله یک روش انتخاب ویژگی با استفاده از الگوریتم ژنتیک ارائه میشود. استفاده از تئوری اطلاعات برای محاسبه ی تابع ارزیابی، نخستین بار در این مقاله ارائه شده است. عملکرد روش پیشنهادی با شناخته شده ترین و جدیدترین روشهای انتخاب ویژگی، بر روی طبقه بند -k نزدیکترین همسایه مقایسه شده است. نتایج به دست آمده از پیاده سازی روش پیشنهادی بر روی مجموعه دادهی روبرتز بیانگر دقتی برابر 85,41% است.

کلمات کلیدی:

کاهش ابعاد، انتخاب ویژگی، خوشه بندی، الگوریتم ژنتیک و طبقه بندی متون.

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/961886>

