



الگوریتم خلاصه‌سازی برای جریان داده برای سرعت بخشیدن به تشخیص داده‌های پرت

حدید ملاشاهی آ، حمید سعادت‌فر آ، حامد وحدت‌نژاد آ*

آ دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.

چکیده

اطلاعات مقاله

تاریخچه مقاله:

دریافت: 7 September 2022

اصلاح: 17 February 2023

پذیرش: 15 March 2023

انتشار آنلاین: 30 May 2023

کلمات کلیدی:

تشخیص پرت، جریان داده، یادگیری ماشینی، خوشه‌بندی، اینترنت اشیا.

تشخیص داده پرت در جریان داده‌ها مسئله‌ای مهم در پردازش داده‌ها است؛ که امروزه با توجه به رشد عظیم حجم داده‌های ایجاد شده توسط گسترش اینترنت اشیا به چالش بزرگی تبدیل شده است. پیشرفتهای زیادی اخیراً در تشخیص داده پرت مبتنی بر الگوریتم‌های تشخیص داده پرت محلی، مانند الگوریتم فاکتور داده پرت محلی مبتنی بر چگالی صورت گرفته است. گرچه این الگوریتم‌ها، برای داده‌های ایستا مناسب هستند؛ از فاکتور افزایشی این الگوریتم‌ها برای تشخیص داده‌های پرت محلی جریانی استفاده می‌شود. با این وجود، تشخیص داده پرت در داده‌های جریانی با چالش‌های محدودیت ظرفیت حافظه، عدم امکان تشخیص در زمان واقعی، عدم دسترسی به تمامی داده‌ها در یک زمان و تغییرات توزیع داده‌ای (کم و زیاد شدن نرخ ورود، عدم قطعیت و غیره) مواجه است. در الگوریتم خلاصه‌سازی پیشنهادی داده‌ها خوشه‌بندی می‌شوند، سپس خوشه‌های متراکم‌تر که سهم بیشتری از داده‌ها را دارند برای خلاصه‌سازی انتخاب می‌شوند. داده‌های درون خوشه‌های بزرگتر که چگالی بیشتری دارند، با استفاده از تابع حذف برای خلاصه‌سازی انتخاب می‌شوند. الگوریتم خلاصه‌سازی پیشنهادی، عملکرد زمان اجرا را کاهش و دقت را در مقایسه با الگوریتم‌های مورد ارزیابی افزایش می‌دهد.

© 2023. تمامی حقوق محفوظ است.

* نویسنده مسئول.

آدرس‌های رایانامه: hdmollashahi@birjand.ac.ir (ح. ملاشاهی)،

saadatfar@birjand.ac.ir (ح. سعادت‌فر)،

vahdatnejad@birjand.ac.ir (ح. وحدت‌نژاد)

تمامی حقوق محفوظ است. ISSN: 2322-4460 © 2023

