

یک الگوریتم جدید جهت تشخیص نقاط پرت در داده‌های عددی با استفاده از اتوماتای یادگیر و کاربرد آن در تشخیص مشترکین پر مصرف شرکت آب و فاضلاب آذربایجان شرقی

مهدی بخشی پور ، بابک اناری

چکیده

داده‌های پرت^{۵۳} داده‌هایی هستند که بطور معنی‌داری از داده‌های اصلی و واقعی انحراف داشته که می‌تواند ناشی از خطای اندازه‌گیری و جمع‌آوری داده‌ها، یا ناشی از عوامل طبیعی باشد. در سیستم قرائت کنتورهای مشترکین آب و فاضلاب، ممکن است به دلیل ثبت اشتباه کارکرد کنتور در سیستم، ترکیبگی لوله‌های داخل ملک، خرابی کنتور یا عوامل دیگر، باعث ایجاد داده‌های پرت شود که شناسایی این داده‌ها و پردازش بر روی آنها، می‌تواند از بروز مشکلاتی در سیستم قرائت جلوگیری نماید. در این مقاله برای حل این مشکل از اتوماتای یادگیر^{۵۴} استفاده می‌شود. نشان می‌دهیم، روش پیشنهادی که از اتوماتای یادگیر جهت تشخیص داده‌های پرت استفاده می‌کند قادر است داده‌های پرت را از داده‌های طبیعی جدا نماید و این کار را از طریق کاهش واریانس^{۵۵} داخلی داده‌های معمولی انجام می‌دهد. همچنین کارایی روش پیشنهادی در مقایسه با برخی روش‌های موجود مورد تحلیل و ارزیابی قرار خواهد گرفت.

کلمات کلیدی: تشخیص داده‌های پرت^{۵۶}. خوشه‌بندی^{۵۷}. داده‌کاوی^{۵۸}. اتوماتای یادگیر.

⁵³ Outlier
⁵⁴ Learning automata
⁵⁵ Variance
⁵⁶ Outlier Detection
⁵⁷ Clustering
⁵⁸ Data mining