

عنوان مقاله:

مرور و مقایسه روش های خوشه بندی شناسایی داده ی پرت در سیستم های مختلف

محل انتشار:

سومین کنفرانس ملی علوم و مهندسی کامپیوتر و فناوری اطلاعات (سال: 1396)

تعداد صفحات اصل مقاله: 9

نویسندگان:

بهناز فرضی - گروه کامپیوتر، واحد بویین زهرا، دانشگاه آزاد اسلامی، بویین زهرا، ایران

فاطمه دادرس - گروه کامپیوتر، واحد بویین زهرا، دانشگاه آزاد اسلامی، بویین زهرا، ایران

خلاصه مقاله:

در حال حاضر با افزایش روزافزون داده ها و حجم اطلاعات در مسایل دنیای پیرامون خود روبرو هستیم که چالش روبروی ما، مدل سازی و تجزیه و تحلیل داده هاست و بهره گیری از روش هایی همچون داده کاوی برای استخراج دانش و اطلاعات نهفته در داده ها، جزء مراحل ضروری شناسایی داده ها به شمار می آید. داده های پرت با عدم تطابق با سایر داده ها سبب بروز مشکل در امر تجزیه و تحلیل داده ها می گردند. بنابراین لزوم شناسایی داده های پرت امری اجتناب ناپذیر است. شناسایی داده های پرت و یا خطاها نقش مهمی در کاهش، محدود کردن حجم محاسبات دارند. داده های پرت در بسیاری از علوم کامپیوتری، پزشکی و تجارت کاربرد دارد. مسیله ی دیگری که امروزه در بحث داده کاوی وجود دارد، بحث کاهش خطا در شناسایی داده ها است. نقش شناسایی داده ی پرت و کاهش خطاها، عامل اصلی مطالعه ی تکنیک های شناسایی داده های پرت و بهبود این تکنیک ها است. در این مقاله سعی شده است تکنیک های خوشه بندی شناسایی داده های پرت و معیارهای رده بندی این روش ها بیان گردد. از جمله این تکنیک ها می توان به تکنیک مبتنی بر خوشه بندی، تکنیک مبتنی بر همسایگی، تکنیک مبتنی بر چگالی و تکنیک امتیازدهی اشاره کرد. در این مقاله، ما قصد داریم تا به مطالعه ی کارهای موجود در زمینه ی روش های شناسایی داده های پرت و کاربرد این روش ها در داده کاوی بپردازیم تا بر مبنای کاربرد این روش ها در یک سیستم، با انتخاب رویکرد متناسب با آن در یک مسیله، روند حل مسایل داده کاوی تسهیل گردد.

کلمات کلیدی:

روش مبتنی بر فاصله، شناسایی روش مبتنی بر چگالی، روش مبتنی بر خوشه بندی، روش درخت پوشای کمینه، تخمین جستجوی k نزدیک ترین همسایه

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/676868>

