



بخشی از ترجمه مقاله

عنوان فارسی مقاله :

Sprint: یک طبقه بندی موازی مقیاس پذیر برای داده کاوی

عنوان انگلیسی مقاله :

SPRINT: A Scalable Parallel Classifier for Data Mining



توجه !

این فایل تنها قسمتی از ترجمه میباشد. برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.



بخشی از ترجمه مقاله

5 نتیجه گیری

5 Conclusion

With the recent emergence of the field of data mining, there is a great need for algorithms for building classifiers that can handle very large databases. The recently proposed SLIQ algorithm was the first to address these concerns. Unfortunately, due to the use of a memory-resident data structure that scales with the size of the training set, even SLIQ has an upper limit on the number of records it can process. In this paper, we presented a new classification algorithm called SPRINT that removes all memory restrictions that limit existing decision-tree algorithms, and yet exhibits the same excellent scaling behavior as SLIQ. By eschewing the need for any centralized, memory-resident data structures, SPRINT efficiently allows classification of virtually any sized dataset. Our design goals also included the requirement that the algorithm be easily and efficiently parallelizable. SPRINT does have an efficient parallelization that requires very few additions to the serial algorithm. Using measurements from actual implementations of these algorithms, we showed that SPRINT is an attractive algorithm in both serial and parallel environments.

با ظهور اخیر زمینه داده کاوی، نیاز زیادی برای الگوریتمها به منظور ایجاد طبقه بندی کننده هایی وجود دارد که می توانند پایگاههای داده بسیار بزرگ را کنترل نمایند. الگوریتم SLIQ که اخیراً پیشنهاد شده اولین الگوریتمی بود که در این ارتباط بیان شد. متأسفانه، به دلیل استفاده از ساختار داده ی ساکن در حافظه که با اندازه مجموعه آموزشی مقیاس بندی می شود، حتی SLIQ نیز محدودیت بالایی برای تعداد رکوردهایی که می تواند پردازش کند، دارد.

در این مقاله، ما یک الگوریتم طبقه بندی کننده جدید به نام SPRINT را ارائه کردیم که همه محدودیتهای حافظه موجود در الگوریتم های درخت تصمیم حذف می کند و هنوز هم رفتار مقیاس بندی عالی مشابه مانند SLIQ نشان می دهد. با اجتناب از نیاز برای هر گونه ساختار داده ساکن در حافظه، متمرکز، SPRINT به طور موثری به طبقه بندی مجموعه داده های مجازی با هر اندازه ای اجازه می دهد. اهداف طراحی ما همچنین این نیاز را شامل می شود که الگوریتم به راحتی و به طور موثر قابل موازی سازی باشد. SPRINT موازی سازی موثری ندارد که نیازمند افزایش های اندکی به الگوریتم های مجموعه ای می باشد.

با استفاده از اندازه گیری ها از اجراهای واقعی این الگوریتم ها، نشان دادیم که SPRINT یک الگوریتم جذاب در محیطهای موازی و مجموعه ای می باشد.



توجه!

این فایل تنها قسمتی از ترجمه میباشد. برای تهیه مقاله ترجمه شده کامل با فرمت

ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.

برای جستجوی جدیدترین مقالات ترجمه شده، [اینجا](#) کلیک نمایید.