9th SASTech Symposium on Advances in Science & Technology (2014)
نهمین سمپوزیوم پیشرفت های علوم و تکنولوژی
مشهد- ۱۳ آذر- ۱۳۹۳
کمیسیون دوم: همایش ملی علوم و مهندسی کامپیوتر
بامحوریت
امنیت ملی و توسعه پایدار
برگزار کننده: موسسه آموزش عالی خاوران
www.khi.ac.ir

انتخاب ویژگی برای داده های بزرگ با استفاده از الگوریتم ژنتیک ترکیبی موازی به کمک برنامه نویسی CUDA

محمدعلی صالح نیا[۱]، وحید ستاری نائینی[۲]، مهدی افتخاری[۳]

[۱] دانشجوی کارشناسی ارشد، دانشگاه شهید باهنر کرمان

[۲] استادیار ، بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان

[۳] استادیار ، بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان

ma.salehnia@eng.uk.ac.ir

خلاصه

انتخاب ویژگی توجه بسیاری از حوزه های تحقیقاتی در سال های اخیر بویژه در حوزه داده های با ابعاد بالا را به خود جلب کرده است. از آنجایی که تکنیک های سنتی در این زمینه کارایی کمتری برای کار با داده ها با ابعاد بزرگ دارند. لذا در این مقاله که برای انجام عمل انتخاب ویژگی برروی داده های بزرگ صورت گرفته است، از یک روش ترکیبی که در درون آن جستجوی محلی برای بالا بردن سرعت همگرایی الگوریتم با استفاده از جدا سازی ویژگی ها به دو دسته متمایز و شبیه استفاده شده است. موازی سازی این الگوریتم به روش پایه- پیرو صورت گرفته است و پیاده سازی آن روی کارت گرافیک انجام شده است. استفاده از کارت گرافیک به کمک زبان برنامه نویسی CUDA زمان اجرای الگوریتم را به حدود یک دوم کاهش می دهد. نتایج آزمایش ها که برروی ۱۶ دیتاست صورت گرفته است نشان میدهد که سرعت اجرای الگوریتم در حالت موازی حدوده دوبرابر بیشتر از حالت سری الگوریتم میباشد.

کلمات کلیدی: انتخاب ویژگی، واحد های پردازش گرافیکی(GPU)، معماری دستگاه یکپارچه ی محاسباتی(CUDA)، الگوریتم ژنتیک موازی،مدل پایه-پیرو

۱. مقدمه

در حال حاضر انتخاب ویژگی به عنوان یک عمل پیش پردازش درحوزه های تحقیقاتی در سال های اخیر مورد توجه ویژه ای قرار گرفته است. کارایی روش های انتخاب ویژگی به چگونگی جستجوی استفاده شده برای یافتن ویژگی های مطلوب بستگی دارد. الگوریتم های انتخاب ویژگی اکثراً با روش های جستجوی ترتیبی مانند [2, 1] و یا روش های جستجوی سراسری مانند [6, 5, 4, 3] در گیر هستند. برای بالا بردن سرعت کار روش انتخاب ویژگی براساس انتروپی فازی ارایه شد [7]. انتخاب ویژگی موجود را می توان به سه روش فراگیر [8, 2, 1] فیلتر [10, 9, 6, 5] و ترکیبی [4] تقسیم کرد. تمام مدل هایی که تا اینجا ذکر شدند برروی داده های با تعداد ویژگی کم کارایی مناسبی دارند. به همین دلیل محققان برای کار با داده های با تعداد ویژگی زیاد، به سمت روش های سریع تر رفته اند، تا بتوانند زمان را برای انتخاب ویژگی کم کنند. لوپز و همکارانش یک روش با جستجوی پراکنده که در درون الگوریتم ژنتیک قرار داشت ارایه کردند [11] . Zheng Zhao و همکارانش مدل موازی ارایه کردند که با استفاده از واریانس بین داده ها سرعت کار را نسبت به مدل های قبلی افزایش داد [12]. اما هنوز سرعت ارایه شده به حدی نبود که بتواند در داده های با تعداد ویژگی زیاد عمل انتخاب ویژگی را به خوبی انجام دهد.