

ارائه الگوریتم موازی حل مسأله پیدا کردن نزدیک ترین همسایه

بر روی GPU با استفاده از CUDA

ندا محمدی^۱، رضا جاویدان^۲، منیژه کشتگری^۲



۱-دانشجوی کارشناسی ارشد مهندسی نرم افزار ، دانشگاه صنعتی شیراز

۲-استادیار دانشکده مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی شیراز

N.Mohammadi@sutech.ac.ir

نام ارائه دهنده:

ندا محمدی

خلاصه

الگوریتم جستجوی نزدیک ترین همسایه کاربردهای فراوانی در علوم کامپیوتر دارد. به عنوان مثال در حوزه داده کاوی تکنیک های دسته بندی مؤثر داده ها از حیاتی ترین مسائل می باشد. بالا بودن پیچیدگی محاسباتی الگوریتم جستجوی نزدیک ترین همسایه ها یک چالش برای زمان اجرا محسوب می شود. طرح این مسأله و حل آن برای مقدار داده ای اندک راحت می باشد، اما زمانی که پایگاه داده بزرگ می شود مشکل بنیادی در پردازش سریع داده بوجود می آید. در حوزه هایی مانند داده کاوی که الگوریتم جستجوی نزدیک ترین همسایه برای آن ها کاربرد دارد ، تکنولوژی های متعددی برای انجام دسته بندی داده ها معرفی شده است که با افزایش حجم داده ها انتخاب تکنولوژی مناسب برای دسته بندی آن ها اهمیت پیدا می کند. تکنولوژی کودا توسط شرکت انویدیا ارائه شده است و فرصتی را برای توسعه دهندگان فراهم می کند تا با استفاده از کارت گرافیک سخت افزار سیستم خود، به سادگی و با کم ترین هزینه، پردازش محاسباتی داده ها را به صورت موازی انجام دهند. در این مقاله از مفهوم GPGPU و تکنولوژی کودا برای جستجوی نزدیک ترین همسایه استفاده شده است و موازی سازی اجرای این الگوریتم بر روی واحد پردازش گرافیکی با دسترسی داشتن به حافظه اشتراکی را با حالتی که برنامه به صورت سریال بر روی پردازنده اجرا می شود و همچنین با حالتی که برنامه بر روی واحد پردازش گرافیکی به صورت موازی بدون استفاده از حافظه اشتراکی اجرا می شود را مقایسه می کند. نشان داده می شود که اجرای الگوریتم به صورت موازی بر روی واحد پردازش گرافیکی با استفاده از حافظه اشتراکی نسبت به دو روش دیگر بررسی شده در این مقاله بهتر عمل کرده و امکانی را برای توسعه دهندگان فراهم می کند که از راه حل های مقیاس پذیر و کم هزینه برای اجرای موازی پردازش های محاسباتی سنگین استفاده کنند.

کلمات کلیدی: نزدیک ترین همسایه ، کودا ، واحد پردازش گرافیکی، حافظه اشتراکی، موازی سازی

۱- مقدمه

الگوریتم جستجوی نزدیک ترین همسایه در مسائلی که باید برای حل آن ها اقلام مشابه را یافت کاربرد دارد. برای حل چنین مسائلی باید نزدیک ترین همسایه ها به یک شیء در یک فضای متریک پیدا شوند. این الگوریتم در حوزه هایی از جمله تخمین آنتروپی ، داده کاوی ، یادگیری ماشین و بازیابی تصویر به کار گرفته می شود [۱،۲،۳]. در این حوزه ها حجم داده هایی که نیاز است عملیات پردازش بر روی آن ها انجام شود دائما افزایش می یابد، همواره به ویژه در داده کاوی [۳] جستجو برای پیاده سازی هایی که بتوانند جواب را برای مجموعه های بزرگ داده بدست آورند انجام می گیرد. پردازش موازی یک راه حل مقیاس پذیر را برای الگوریتم پیدا کردن نزدیک ترین همسایه در حالتی که حجم داده ها بالا است ارائه می کند. واحد پردازش گرافیکی شامل صدها هسته پردازنده موازی می باشد که قادرند به طور همزمان هزاران نخ را مدیریت کنند و با این هدف بوجود آمده است تا محاسبات