

تخمین تابع چگالی توزیع شده در شبکه های سنسوری نظیر به نظیر

بهروز صفری نژادیان^۱، محبوبه استخری اصطهباناتی^۲

^۱ دانشگاه صنعتی شیراز
safarinejad@sutech.ac.ir

^۲ دانشگاه صنعتی شیراز
m.estakhri@sutech.ac.ir

چکیده

این مقاله یک الگوریتم بیزین تغییراتی توزیع شده برای تخمین چگالی و دسته بندی داده ها در شبکه های سنسوری نظیر به نظیر (peer-to-peer) ارائه می کند. در این الگوریتم، مقادیر چکیده محلی در هر گره، تنها با استفاده از داده های محلی محاسبه می شود. سپس با استفاده از یک روش میانگین گیری peer-to-peer مقادیر چکیده محلی به گره های همسایه فرستاده می شود و چکیده سراسری در هر گره محاسبه می شود. پس از آن، هر گره با استفاده از چکیده سراسری پارامترهای مدل ترکیب گوسی را به روز رسانی می کند. در این الگوریتم هر گره فقط نیاز به ارتباط با گره های همسایه دارد. در نتیجه الگوریتم پیشنهاد شده مقیاس پذیر و مقاوم است. در پایان، برای ارزیابی کارایی الگوریتم ذکر شده، چند شبیه سازی ارائه شده است.

کلمات کلیدی

الگوریتم بیزین تغییراتی توزیع شده، تخمین تابع چگالی احتمال، شبکه سنسوری، شبکه peer-to-peer، مدل ترکیب گوسی،

۱- مقدمه

[۳] دسته بندی داده ها با استفاده از روش k-means برای حالت توزیع شده توسعه یافته است. روش هایی برای تخمین چگالی به صورت غیر پارامتری با استفاده از داده هایی که به صورت همگن توزیع شده اند نیز در [۴] مورد بررسی قرار گرفته است. استفاده از مدل های گرافیکی برای استنتاج توزیع شده در شبکه ای از سنسورها در مرجع [۵] بررسی شده است. در این چارچوب، تابع توزیع احتمال مورد نظر به صورت یک گراف بیان می شود و فرآیند استنتاج به صورت دنباله ای از یک سری عملیات پیام رسانی بین گره های گراف انجام می گردد. در [۶] از این روش برای توابع توزیع گوسی با چند متغیر استفاده شده است.

یک الگوریتم EM (Expectation Maximization) توزیع شده به منظور تخمین تابع چگالی و دسته بندی داده ها در شبکه های سنسوری ارائه شده است [۷]. سپس در [۸] یک الگوریتم Newcast EM که از ساختار ارتباطی متفاوتی بهره می برد ارائه گردید. یک الگوریتم EM توزیع شده افزایشی برای تخمین تابع چگالی احتمال در شبکه های نظیر به نظیر ارائه شده است [۹].

نیاز به داده کاوی توزیع شده در محیط های توزیع شده ای مانند اینترنت، اینترنت، شبکه های داخلی موسسات مختلف، شبکه های سنسوری و سایر ساختارهای غیرمتمرکز، مناسب بودن روش تحلیل داده به صورت متمرکز را زیر سوال می برد. در داده کاوی توزیع شده، توجه ویژه ای به بهینه کردن فرآیند تحلیل داده ها از طریق تحلیل آنها به صورت توزیع شده، به جای جمع آوری تمام داده ها در یک واحد مرکزی معطوف می گردد [۱]. علاوه بر آن، در صورت زیاد بودن حجم داده ها، افزایش سرعت فرآیند داده کاوی نیز لازم خواهد بود. [۲]. داده کاوی توزیع شده اخیراً به عنوان یک زمینه بسیار مهم تحقیقاتی مطرح شده است. در اینجا، هدف انجام عملیات داده کاوی (مانند دسته بندی داده ها) بر روی یک مجموعه داده توزیع شده می باشد، یعنی مجموعه داده ای که در گره های مختلف یک شبکه توزیع شده اند. هدف از تحقیقات انجام شده در این زمینه، ارائه روش هایی به منظور تحلیل اطلاعات موجود در این مجموعه داده ها و در عین حال حداقل کردن حجم ارتباط لازم بین گره ها می باشد. در