

به نام حق



دانشکده فنی و مهندسی
کارشناسی ارشد مهندسی کامپیوتر نرم افزار
گروه مهندسی کامپیوتر و فناوری اطلاعات

گزارش درس پردازش موازی

موضوع:

کاهش و بازسازی تله متری در مراکز داده بلادرنک با استفاده از مدل های زنجیر مارکوف

نگارش:

سیما محسنی

استاد راهنما:

جناب آقای دکتر آرش قربان نیا دلاور

آبان 1399

فهرست مطالب

صفحه	فهرست مطالب
۳	چکیده
۴	مقدمه
۷	بخش دوم کارهای مرتبط
۹	بخش سوم- چارچوب کاهش داده
۱۲	A. کاهش با استفاده از روشهای PR
۱۳	B: کاهش با استفاده مدل های زنجیر مارکوف
۱۷	بخش چهارم: راه اندازی آزمایشی
۱۷	A: توصیف مجموعه داده
۱۸	B: جزئیات آزمایش
۱۹	C. معیارهای ارزیابی
۲۰	D. تست K-S دو نمونه ای
۲۰	E. چرخش زمان پویا
۲۱	بخش پنجم. نتایج تجربی
۲۱	A. آزمون ۱: کاهش و بازسازی داده ها با PR
۲۸	B. آزمایش ۲- کاهش بازسازی داده ها با استفاده از MM
۳۴	C. مقایسه PR و MM
۳۵	D. کاهش داده های به استفاده از فشرده سازی ZIP
۳۸	E. کاهش پهنای باند با استفاده از MM
۳۹	بخش ششم: مزایای استفاده از حل پیشنهادی در مراکز مختلف داده
۴۱	بخش هفتم: نتیجه گیری و تحقیقات آینده
۴۴	مراجع

چکیده

مراکز داده بزرگ از هزاران سرور در قالب رکهای به هم متصل تشکیل شده اند تا به کاربران سرویس دهی کنند. این مراکز داده به طور مداوم مقدار زیادی از جریان داده های تله متری (به عنوان مثال ، معیارهای استفاده از سخت افزار) برای اهداف مختلف، از جمله مدیریت منابع ، تشخیص حجم کار، پیش بینی استفاده از منابع، برنامه ریزی ظرفیت و تجزیه و تحلیل بلادرنگ تولید می کنند. این جریانهای تله متری به استفاده از پهنای باند پر هزینه و فضای ذخیره سازی به ویژه در بخش میان مدت - طولانی برای مراکز بزرگ داده نیاز دارند. این مقاله به این موضوع از طریق پیشنهاد و ارزیابی یک سیستم برای کاهش کارآمد پهنای باند و ذخیره سازی داده های تله متری از طریق مدل سازی بلادرنگ با استفاده از روش های مبتنی بر زنجیره مارکو می پردازد. راه حل پیشنهادی ما با استفاده از مجموعه داده های تله متری واقعی بود و با روش های چند جمله ای رگرسیون برای کاهش و بازسازی داده ها مقایسه شده است. نتایج تجربی نشان می دهد که داده ها می توانند فشرده سازی ضعیفی برای استفاده از پهنای باند به 75٪ و فضای ذخیره سازی 95/33٪ و با دقت بازسازی نزدیک به 92٪ داشته باشند.

امروزه مقادیر پتابایت از داده دیجیتال توسط دستگاههای الکترونیکی تحت شبکه که به راحتی در دسترس هستند، شبکه های اجتماعی، وسایل الکترونیکی مراقبت های بهداشتی و مراکز داده ها به طور عمده تولید می شوند. این رشد فزاینده داده های دیجیتال از لحاظ چالش های پردازش، ذخیره و تجزیه و تحلیل داده های جمع آوری شده اهمیت ویژه ای دارند. بنابراین، ما به الگوریتم ها، ابزارها و سیستم های جدیدی نیاز داریم تا داده های دارای رشد فزاینده را مدیریت کنید. محاسبات ابری به کاربران این امکان را می دهد که عمدتاً بر اساس ¹PAYG و مقیاس پذیری پویا² داده ها و برنامه های خود را در مراکز داده راه دور میزبانی کنند. این مراکز داده در مقیاس بزرگ از هزاران سرور تشکیل شده است که به صورت رک تنظیم شده اند و به منظور ارائه خدمات به مجموعه بزرگی از کاربران، به هم پیوسته اند. چنین مراکز داده ای جریانهای بزرگ و مداوم از داده های تله متری برای مقاصد مختلف از جمله مدیریت منابع، خصوصیات بار کاری، پیش بینی ثبت و تجزیه و تحلیل منابع، برنامه ریزی ظرفیت و تجزیه و تحلیل بلادرنگ ایجاد می کنند [7][4].

جریانها تله متری معمولی حاوی داده های سری زمانی در مورد معیارهای استفاده از سخت افزار، از جمله واحد پردازش مرکزی (CPU)، حافظه، I/O، پهنای باند، سوئیچ های زمینه، وقفه، خطای حافظه پنهان و چرخه در هر دستورالعمل (CPI) هستند. این معیارهای سنجش تله متری در کاربردهای مختلف استفاده می شوند. به عنوان مثال، از CPU، حافظه، پهنای باند شبکه، دستورالعمل در هر چرخه، میزان حافظه پنهان، آمار پیش بینی شاخه و داده های مصرف انرژی برای پیش بینی مصرف انرژی و کاهش استفاده می شود [8][9]. CPU، حافظه، دیسک و مصرف شبکه برای مدیریت مرکز داده و پیش بینی منابع استفاده می شود [10] - [13]. این نسل از جریانهای مداوم تله متری از همه گره های محاسباتی و ذخیره سازی، از نظر مصرف پهنای باند و نیازهای ذخیره سازی، یک چالش مهم را در مرکز داده ایجاد می کنند. به عنوان نمونه، در نظر گرفتن مجموعه تله متری در یک مرکز داده متشکل از 10000 گره محاسباتی و جمع آوری 12 اندازه گیری مختلف در هر ثانیه، در حالت اختصاص 4 بایت برای هر متریک تقریباً به 40 گیگابایت در روز نیاز دارد، این بیش از 1 TB ذخیره سازی در هر ماه به علاوه

¹ pay-as-you-go

² dynamic-scalability

سربار متا داده ها برای قابلیت ردیابی سری زمانی می باشد. راه حل های فشرده سازی داده های سنتی برای کاهش داده ها زمان جمع آوری داده های تله متری از گره های محاسباتی را افزایش می دهد. افزایش بازه زمانی اجازه گرفتن مصرف منابع دانه ریز را نمی دهد و ممکن است دقیقاً منعکس کننده استفاده از منابع نباشد. همچنین، استفاده از تکنیک های فشرده سازی روی مقادیر نقطه شناور نمی تواند اندازه را به میزان قابل توجهی کاهش دهد و دقت کامل را حفظ کند [14] - [16]. متأسفانه، جریانهای تله متری در بسیاری از مراکز داده، نرمی کم یا صفر و تغییر زیادی در داده ها از خود نشان می دهند. در چنین مواردی، حتی الگوریتم های فشرده سازی state-of-the-art floating point هم برای فشرده سازی داده ها کافی نیستند [17].

در مراکز داده در مقیاس بزرگ با میلیون ها میزبانی برنامه کاربردی، سودمندی تله متری و پروفایل باعث می شود از الگوهای رفتاری بیشتر از معیارهای دقیق استفاده شود. هنگام جمع آوری داده ها در دوره های زمانی بزرگتر، دقت کاهش می یابد. از این رو، روشهای کاهش داده های سری زمانی تله متری نیاز به حفظ خصوصیات آماری دارد که رفتارهای استفاده از سخت افزار، بخصوص تداوم، رشد سریع استفاده و الگوهای استفاده غیرعادی از سخت افزار را ضبط می کند.

در این مقاله، ما به جمع آوری و فشرده سازی داده های تله متری مراکز داده در مقیاس بزرگ می پردازیم. راه حل پیشنهادی ما شامل کاهش اندازه گیری تله متری مرکز داده بلادرنگ از طریق مدل سازی آنلاین با استفاده از زنجیره های مارکوف است [18]. سپس، چنین مدلهایی به مخازن ورود به سیستم مربوطه منتقل و ذخیره می شوند تا بازسازی داده ها برای استفاده خلفی با حداقل تلفات داده برای حفظ رفتارهای استفاده از سخت افزار امکان پذیر شود. این روش در سطح رک کار می کند تا تمام اندازه گیری ها را از گره های مستقر در رک مستقر کند و سپس از مدل زنجیره مارکوف (MM) استفاده می کند تا به طور موثر داده ها را در بلادرنگ کاهش دهد. داده های کاهش یافته وارد شده اند تا به تجزیه و تحلیل داده های سنجش از راه دور پردازند. ما همچنین یک روش کارآمد ارائه می دهیم تا ترکیب داده های کاهش یافته با فشرده سازی را به حداقل برسانیم تا نیاز به ذخیره سازی کلی برای ذخیره داده های تله متری برای مدت طولانی به حداقل برسد. بنابراین هم فضای ذخیره سازی و هم استفاده از پهنای باند برای جمع آوری تله متری مرکز داده مینیمم می شود.

روش ارائه شده با استفاده از مجموعه داده های واقعی تله متری و مقایسه با حالت روش های هنری، مانند روش رگرسیون چند جمله ای [19] (PR) و فشرده سازی مبتنی بر فرهنگ لغت (ZIP) مورد بررسی قرار گرفته است. چندین معیار مقایسه استفاده شده است، مانند محاسبه مقدار داده های ذخیره شده در هر سناریو و دقت بازسازی داده ها در صورت فشرده سازی سست. ما با محاسبه و مقایسه نیازهای ذخیره سازی برای هر روش، اثربخشی روشها را در داده های سری زمانی تله متری ارزیابی می کنیم. دقت بازسازی با مقایسه داده ها قبل از فشرده سازی و پس از بازسازی ارزیابی می شود. ابتدا، ما شباهت آماری بین داده های بازسازی شده و داده های اصلی را با استفاده از آزمون فرضیه دو نمونه $(K-S)^3$ مقایسه می کنیم [20]. از آزمون K-S برای شناسایی اینکه آیا دو توالی یک بعدی داده شده به توزیع احتمال یکسان تعلق دارند یا نه استفاده می شود. این روش شباهت داده های بازسازی شده را کم نمی کند، اما نشان می دهد که آیا بازسازی شباهت آماری دارد یا خیر. سپس ارزیابی را با تعیین کمیت این شباهت با استفاده از متریک بسته بندی زمان پویا (DTW) تکمیل می کنیم [21][22]. DTW روشی شناخته شده است که برای اندازه گیری شباهت بین دو توالی داده شده که ممکن است در سرعت متفاوت باشد، استفاده می شود [26].

مطالب اصلی مقاله به شرح زیر خلاصه می شود.

- 1) سیستمی برای کاهش و بازسازی داده های تله متری بلادرنگ برای مراکز داده طراحی کنید.
- 2) با استفاده از مدل های زنجیره مارکوف، روش تله متری کاهش و بازسازی داده ها توسعه و ارزیابی شود.
- 3) کاهش داده ها و بازسازی پیشنهادی را با روش های مبتنی بر رگرسیون چند جمله ای (PR) و فشرده سازی ZIP مقایسه کنید.
- 4) ارزیابی تجربی برای مطالعه ذخیره سازی و به حداقل رساندن پهنای باند با استفاده از راه حل پیشنهادی برای داده های تله متری برای اندازه های مختلف مرکز داده.

³ Kolmogorov-Smirnov

بقیه مقاله به شرح زیر سازماندهی می شوند. کارهای مرتبط در بخش دوم ارائه شده است. سیستم کاهش داده و بازسازی پیشنهادی ما در بخش سوم توضیح داده شده است. جزئیات راه اندازی آزمایشی در بخش چهارم آورده شده است. نتایج ارزیابی در بخش پنجم ارائه شده است. مزایای استفاده از پهنای باند و کاهش ذخیره سازی برای اندازه های مختلف مراکز داده با استفاده از راه حل پیشنهادی در بخش ششم بحث شده است. سرانجام، نتیجه گیری و کارهای آینده در بخش هفتم بحث شده است.

بخش دوم کارهای مرتبط

تلاش های بسیاری برای کاهش داده های دیجیتال در حال رشد تصاعدی برای مدیریت کارآمد صورت گرفته است [27] - [34]. روش های مختلفی از جمله کاهش ابعاد، مدل های پیش بینی و روش های فشرده سازی ارائه شده است. به عنوان مثال، Bhuiyan و همکاران [35]، چارچوبی IoT را برای شناسایی رویدادها و کاهش داده ها در زمان جمع آوری داده ها پیشنهاد کردند که به حداقل رساندن انتقال داده از طریق شبکه و همچنین کاهش مصرف انرژی کمک می کند. چارچوب پیشنهادی با استفاده از سنسورها و روشهای مبتنی بر قانون، حوادث آتش سوزی را شناسایی می کند. Wu و همکاران [36]، یک تکنیک فشرده سازی مبتنی بر فرهنگ لغت برای تقسیم جریان داده های عددی ورودی به بلوک های اندازه ثابت ایجاد کردند و آنها را با استفاده از آزمون آماری K-S برای اندازه گیری شباهت، مقایسه کردند. هنگامی که آنها به هر بلوک موجود، بلوک مشابه ورودی جدید را شناسایی کردند بلوک ورودی را دور ریختند و مرجعی از بلوک قدیمی را نگه داشتند تا بتوانند داده ها را بازسازی کنند، بنابراین این کار به کاهش چشمگیر ذخیره سازی مورد نیاز کمک می کنند. در کار دیگری، Egri و همکاران [37] از کاهش ابعادی داده های سری زمانی چند بعدی استفاده کردند. رویکرد آنها خوشه بندی مبتنی بر گراف را با استفاده از همبستگی متقابل بین داده های سری زمانی معرفی می کند. تمرکز نویسندگان بر شناسایی اتصالات بین معیارهای مختلف عملکرد به منظور کاهش تعداد معیارهای عملکرد برای ردیابی است.

یک اثر اخیر [38] از یک روش مبتنی بر همبستگی برای کاهش داده های نظارت بر مرکز داده استفاده می کند. نویسندگان ارتباط بین معیارهای اندازه گیری مختلف را با استفاده از مدل های شبکه بیزی که از داده های تاریخی آموخته شده اند شناسایی کرده و پیشنهاد دادند که از معیارهای همبسته استفاده شده است. مدل های شبکه بیزی

نمودارهای چرخه ای هستند که رابطه بین معیارها را به شکل معیارهای وابسته و مستقل نشان می دهند. در این روش، نویسندگان میزان نمونه برداری از معیارهای وابسته را کاهش داده و آنها را با استفاده از رگرسیون خطی برای مدت زمان مشخص پیش بینی می کنند، که به کاهش داده ها در مرحله جمع آوری داده ها کمک می کند. کار اخیر Yu و همکاران [39]، روشی را برای کاهش داده های ارسال شده توسط گره های لبه دستگاه های IoT به ابر برای کاهش زمان انتقال داده پیشنهاد داد. روش داده های ورودی را به عنوان توزیع عادی چند متغیره مدل سازی کرده و از فیلتر کالمن برای پیش بینی میانگین بردار و ماتریس کواریانس توزیع استفاده می کند. گره های لبه و ابر هر دو با استفاده از فیلترهای کالمن یکسان، مقادیر یکسانی را پیش بینی می کنند. اگر داده های پیش بینی شده در هر دو انتها فاصله اطمینان را برآورده نکنند، سپس داده ها از مه به لایه ابر بارگذاری می شوند، در غیر این صورت مقادیر پیش بینی شده استفاده می شود، که به کاهش حرکت داده از مه به لایه ابر کمک می کند.

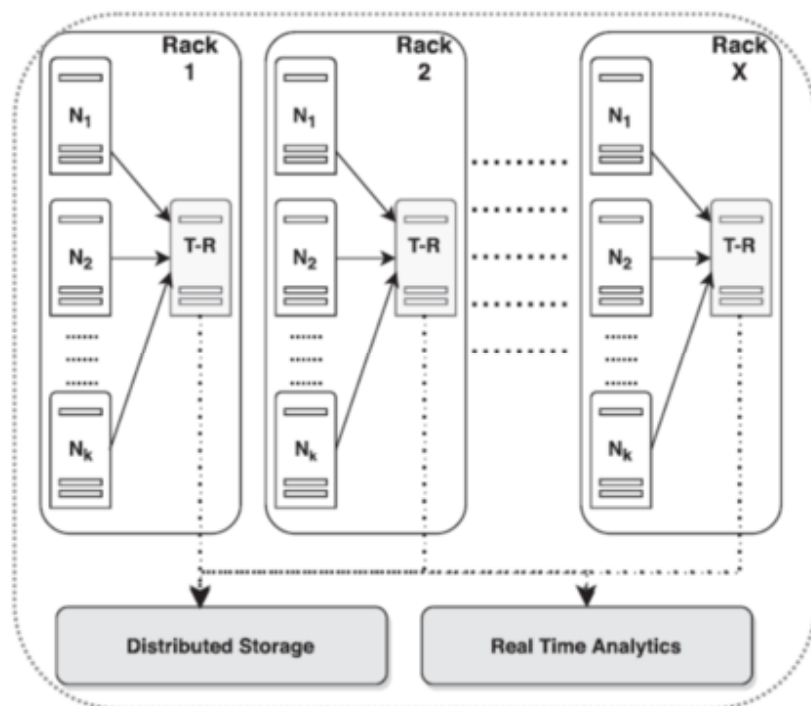
جدیدترین اثر [40]، در کاهش داده ها مشکل انتقال داده در زیرساخت های اندازه گیری انرژی هوشمند را برطرف می کند. نویسندگان چارچوبی برای نظارت بر داده های مصرف انرژی با استفاده از کنتورهای هوشمند و سپس جمع آوری داده ها برای فواصل دسته ای ثابت پیشنهاد دادند. سپس برای مقایسه داده های جدید از روش پیش بینی آموخته شده استفاده کردند. اگر داده های جدید با مدل پیش بینی قابل مقایسه باشد، داده های جدید به ابر ارسال نمی شوند. اما اگر مدل پیش بینی و داده های جدید متفاوت باشند، آنها داده ها را به ابر ارسال می کنند و همچنین روش پیش بینی را به روز می کنند. سیستم پیشنهادی مطابق با داده های ورودی تطبیقی است زیرا به روش پیش بینی منفرد تکیه نمی کند، بلکه روش خود را برای تنظیم داده های فعلی تغییر می دهد. با این حال، این کار بهینه سازی مورد نیاز ذخیره سازی داده ها را برطرف نمی کند و فقط روی داده های مصرف انرژی بعدی تمرکز دارد.

بیشتر آثار موجود براساس کاهش ابعاد، مدل های پیش بینی یا روش های فشرده سازی برای کاهش داده ها استوار است. کار ما یک سیستم کاهش داده و بازسازی داده های سنجش از تله متری مبتنی بر زنجیره مارکوف را ارائه می دهد تا بطور موثر استفاده و پهنای باند شبکه و فضای ذخیره سازی را کاهش دهد. کاهش داده ها در بلادرنگ و بدون کاهش ابعاد جریانهای تله متری ورودی و بدون استفاده از مدل های پیش بینی که هنگام تغییر جریان های

ورودی نیاز به بروزرسانی یا تغییر دارند انجام می شود. مدل های زنجیره مارکوف پیشنهادی بدون به روزرسانی مدل، داده های ورودی را به میزان قابل توجهی کاهش می دهند. علاوه بر این، سیستم هدفمند ما کاهش بیشتری در مقایسه با روشهای فشرده سازی ZIP مبتنی بر فرهنگ لغت دارد. بر اساس دانش ما، روشهای مدرن فعلی کاهش زمان واقعی را برای جریانهای تله متری مراکز داده با استفاده از مدلهای زنجیره مارکوف انجام نمی دهند. ما با استفاده از مدلهای زنجیره مارکوف بررسی کرده و آن را با PR و فشرده سازی ZIP با تنظیمات مختلف مقایسه می کنیم تا داده ها به میزان قابل توجهی کاهش یابد و داده های با دقت بالا برای جریان های مرکز سنجش داده ها بازسازی شود.

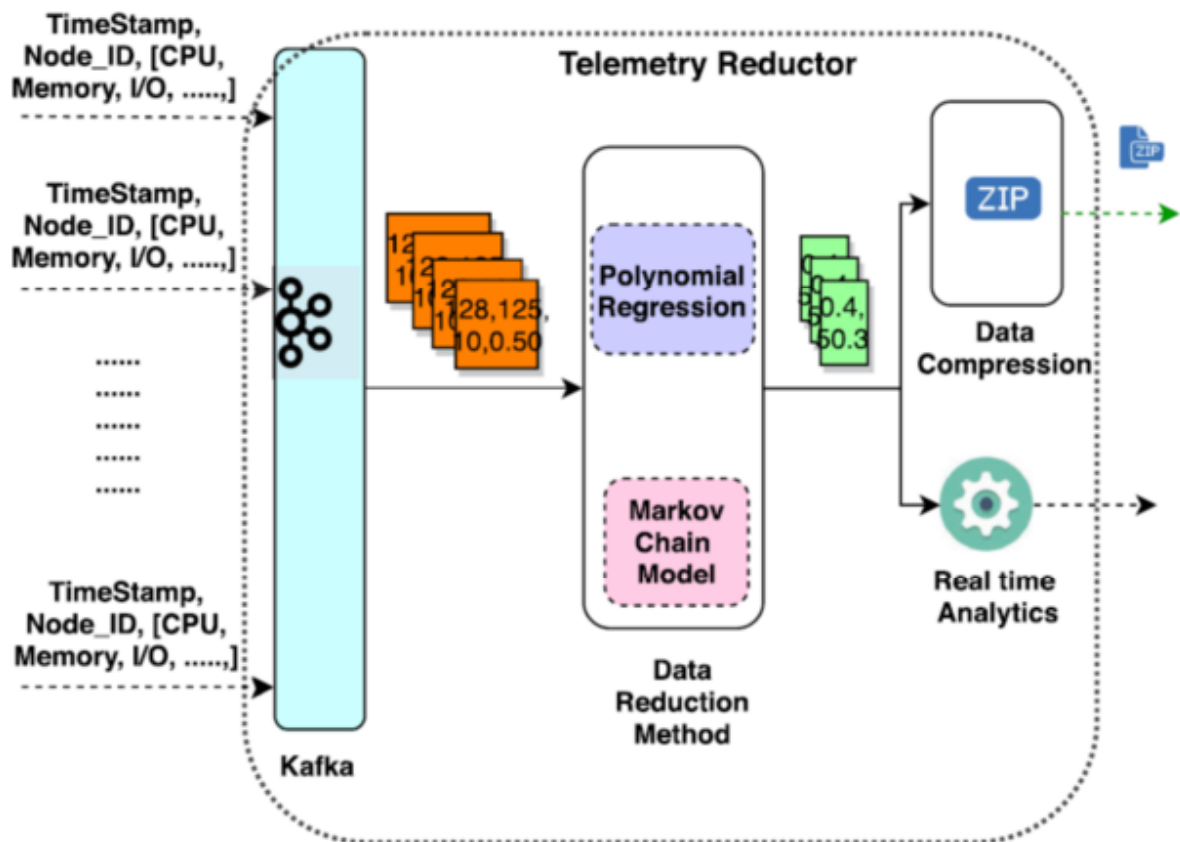
بخش سوم - چارچوب کاهش داده

سیستم پیشنهادی ما برای کاهش داده های تله متری مزایای دو برابری را ارائه می دهد. اول، فضای ذخیره سازی را به میزان قابل توجهی کاهش می دهد و دوم، استفاده از پهنای باند مورد نیاز جمع آوری داده های از راه دور را در مرکز داده به حداقل می رساند. شکل 1 معماری سیستم کاهش داده های اندازه گیری تله متری پیشنهادی را برای مراکز داده نشان می دهد. هر رک در مرکز داده میزبان یک مؤلفه تقلیل سنج تله متری (T-R) است که بطور مداوم داده های تله متری را از هر گره محاسباتی جمع می کند و قبل از انتقال آن به سیستم های آنالیز و ذخیره سازی و در بلادرنگ، روش کاهش داده را انجام می دهد. این داده ها را در سطح رک کاهش می دهد و کل داده ها را منتقل نمی کند، بلکه فقط داده های کاهش یافته را در مرکز داده قرار می دهد. داده های جریان از راه دور می تواند شامل استفاده از CPU، حافظه، دیسک، شبکه، پهنای باند حافظه و سایر معیارهای مفید باشد.



شکل 1: طرحواره چارچوب کاهش کاهش داده پیشنهادی مراکز داده. هر رک در مرکز داده میزبان یک مؤلفه **T-R** است که به طور سریع داده های سنجش از هر گره محاسباتی را جمع آوری می کند، سپس کاهش داده ها را قبل از انتقال آن به سیستم های آنالیز و ذخیره سازی و در بلادرنگ انجام می دهد.

شکل 2 روند T-R را نشان می دهد. ما از آپاچی کافکا [41][42] به عنوان یک کارگزار پیام برای دریافت داده های جریان سنج از طریق گره های محاسباتی استفاده کردیم. هر گره محاسباتی داده های تله متری را با عنوان زمان سنج به موضوع Kafka، که توسط T-R مصرف می شود، منتشر می کند. هر مصرف کننده داده هایی از موضوعات کافکا را بدست می آورد. مصرف کننده جریان های ورودی را می خواند و داده ها را در یک دسته از پیش تعریف شده (BS) تقسیم می کند و از یک روش کاهش داده برای به حداقل رساندن داده ها استفاده می کند. سپس داده های کاهش یافته در یک مرکز ذخیره سازی مرکز داده فشرده شده و ذخیره می شوند. داده های کاهش یافته نیز به موتورهای آنالیز تغذیه می شوند.



شکل 2: T-R جریان های تله متری ورودی از گره های محاسباتی را با استفاده از کافکا و روش کاهش داده پردازش می کند. از خروجی T-R می توان برای ذخیره یا تأمین ورودی برای سایر سیستمهای تحلیلی در بلادرنگ استفاده کرد.

از داده های سنجش می توان برای نظارت در بلادرنگ، خصوصیات بار کاری و اهداف تشخیص ناهنجاری استفاده کرد. داده های کاهش یافته ارسال شده به سایر مؤلفه ها قبل از استفاده باید بازسازی شوند. معماری فوق الذکر مبتنی بر نظارت خارج از باند است که در آن داده ها از دستگاه ها مانند حسگرها یا برنامه های ورود به سیستم تهیه می شود. در صورت وجود میزبان معمولی، جایی که امکان کنترل در باند وجود دارد، می توان مؤلفه T-R را بر روی دستگاه میزبان و بدون کافکا اجرا کرد تا از سربار اضافی یک دستگاه خاص کاسته شود. چارچوب پیشنهادی دو هدف اصلی را هدف قرار می دهد: (1) با کاهش و فشرده سازی داده های تله متری، میزان تقاضای ذخیره سازی برای ذخیره اندازه گیری تله متری را به حداقل می رساند. (2) استفاده از پهنای باند را با انتقال داده های کاهش

یافته از طریق شبکه مرکز داده به حداقل می‌رساند. برای دستیابی به چنین اهدافی ما دو روش مختلف، یعنی مدل‌های زنجیره روابط عمومی و مارکف را پیشنهاد می‌کنیم که برای کاهش داده‌های از راه دور و بازسازی در بلادرنگ استفاده شوند. ما هر دو روش را در زیربخش‌های زیر توضیح می‌دهیم.

A. کاهش با استفاده از روش‌های PR

ابتدا پیشنهاد می‌کنیم از روش‌های PR استفاده کنیم تا منحنی یک جریان داده تله متری داده شده در یک منحنی چند جمله‌ای قرار گیرد. فقط باید ضرایب معادله متناسب با منحنی را ذخیره کنیم. این روش با الهام از کار مشابهی در مرجع [19] است که از روش رگرسیون خطی برای کاهش داده‌ها استفاده می‌کند. از روش PR به عنوان یک پایه برای مقایسه روش پیشنهادی مبتنی بر مدل زنجیره مارکوف استفاده می‌شود. برای درک تأثیر استفاده از PR برای کاهش داده‌ها، اگر قصد داریم یک تابع جریان داده را در یک منحنی 4 چند جمله‌ای قرار دهیم، ما فقط چهار ضریب به اضافه رهگیری را ذخیره می‌کنیم. با فرض اینکه جریان داده شامل 128 نقطه داده باشد، ما داده‌ها را از 128 مقدار به 6 مقدار کاهش می‌دهیم.

اگر یک مدل PR را برای هر اندازه‌گیری تله متری آموزش دهیم، با توجه به PR درجه n ، و مشاهدات k از یک بعد تله متری خاص a ، ضرایب را با استفاده از حل شبه معکوس محاسبه می‌کنیم تا حداقل مربعات را به مینیمم برسانیم،

$$\begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = [R^T R]^{-1} R^T \begin{pmatrix} r_t \\ r_{t-1} \\ \vdots \\ r_{t-k} \end{pmatrix} \quad (1)$$

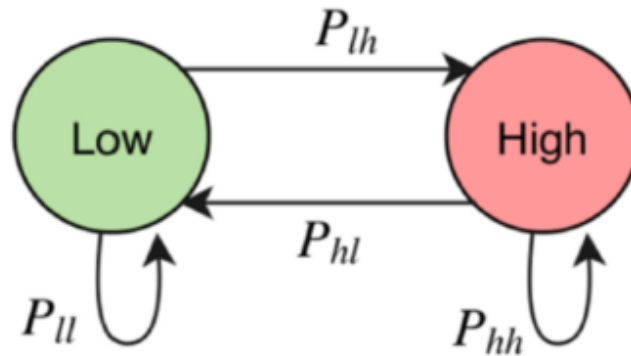
که در آن ماتریس R به صورت زیر تعریف می‌شود

$$R = \begin{pmatrix} 1 & a_t & a_t^2 & \dots & a_t^n \\ 1 & a_{t-1} & a_{t-1}^2 & \dots & a_{t-1}^n \\ \vdots & \vdots & & & \\ 1 & a_{t-k} & a_{t-k}^2 & \dots & a_{t-k}^n \end{pmatrix}. \quad (2)$$

با استفاده از مقادیر ضرائب $\alpha_0, \alpha_1, \dots, \alpha_n$ و درجه چند جمله ای n ، می توانیم منحنی چند جمله ای ها را بازسازی کنیم و k نقطه داده را به راحتی بازسازی کنیم. بنابراین، ما پیشنهاد می کنیم به جای داده های واقعی، ضرایب و اطلاعات مربوط به درجه چند جمله ای را ذخیره کنیم. در اینجا داده ها را برای یک بازه دسته ای خاص نظارت و جمع آوری می کنیم و سپس یک مدل PR را با درجه خاصی متناسب می کنیم. پس از جاگذاری مدل، ضرایب را از معادله مناسب استخراج می کنیم و به جای داده های واقعی آنها را ذخیره می کنیم. هر وقت مجبور شدیم یک دسته داده را بازسازی کنیم، مقادیر ضریب آن دسته را بار می کنیم و مدل PR مربوطه را می سازیم، سپس از آن برای بازسازی دسته نقاط داده استفاده می کنیم.

B: کاهش با استفاده مدل های زنجیر مارکوف

دوما، ما پیشنهاد می کنیم از مدل های زنجیره مارکوف (MM)، [18] برای کاهش داده ها و بازسازی داده های تله متری استفاده شود. زنجیرهای مارکوف مدل های تصادفی هستند که توالی وقایع را توصیف می کنند که در آن احتمال هر رویداد فقط به وضعیت قبلی رویداد بستگی دارد. به طور خاص، ما از زنجیره های مارکوف زمان گسسته همگن (DTMC)، [43] استفاده می کنیم زیرا داده های تله متری در فواصل زمانی گسسته مورد بررسی قرار می گیرند و احتمال انتقال حالت مستقل از زمان است. علاوه بر این، DTMC استفاده شده با داده های تله متری، غیرقابل رد شدن است زیرا سیستم پیشنهادی می تواند از هر حالت به هر حالت دیگر با احتمال مثبت عبور کند. شکل 3 زنجیره مارکوف 2 حالتی ارگونودیک غیرقابل برگشت را نشان می دهد.



شکل 3: مدل زنجیره ای مارکوف دو حالت، حالت ها و احتمال انتقال را نشان می دهد. برای احتمالات انتقال، I نمایانگر حالت کم و h حالت بالا را نشان می دهد.

ایده استفاده از MM این است که کشف کنیم که در صورت کاهش رفتار و بازسازی داده ها می توانیم با رفتارهای پشت سر هم مواجه شویم. رفتار انفجاری نشان دهنده لبه و قله های ناگهانی در داده های تله متری است. به طور کلی، بازسازی رفتار پشت سر هم چالش برانگیز است و ما با استفاده از مدل های زنجیره مارکوف به آن می پردازیم.

فرض کنید X_1, X_2, X_3, \dots متغیرهای تصادفی مستقل و یکسان توزیع شده هستند که داده های تله متری را نشان می دهند. ما از این متغیرهای تصادفی به عنوان یک مدل DTMC استفاده می کنیم که احتمال انتقال از حالت فعلی به حالت بعدی به صورت زیر تعریف می کنند.

$$\begin{aligned} Pr(X_{t+1} = s_j | X_t = s_i, \dots, X_2 = s_2, X_1 = s_1) \\ = Pr(X_{t+1} = s_j | X_t = s_i) \end{aligned} \quad (3)$$

که در آن تابع $Pr(X_{t+1} = s_j | X_t = s_i)$ مستقل از t است و احتمال حرکت از حالت s_i در زمان t به حالت s_j در زمان $t + 1$ را نشان می دهد، نماد $|$ "نمایانگر احتمال شرطی و فضای حالت (S) است.

فرض کنید P_{ij}^v , $v > 1$ در جایی که $P_{ij}^v = P(X_{u+v} = j | X_u = i)$ این احتمال را نشان دهد که پس از واحد زمان v ، زنجیره به حالت j عبور کند با توجه به اینکه وضعیت فعلی i در زمان u است. احتمال رسیدن به j از

i در n مرحله برابر مجموع تمام احتمالات از i به j از طریق نقطه میانی k است. ما از معادله چاپمن-کلموگروف [44] برای محاسبه آن به شرح زیر استفاده می کنیم:

$$P_{ij}^{u+v} = \sum_{k \in S} P_{ik}^u P_{kj}^v; \quad m \text{ and } v \geq 1, \quad i \text{ and } j \in S. \quad (4)$$

فرض کنید $P^v = (P_{ij}^v)$ یک ماتریس باشد، در این صورت معادله چاپمن-کلموگورف را می توان به صورت $P^{u+v} = P^u P^v$ بیان کرد. این کار اجازه می دهد تا محاسبه ماتریس احتمال انتقال P، که بازتاب فرکانس نسبی انتقال از یک حالت به حالت دیگر است. ماتریس P برای تعداد کل ایالات به شرح زیر است:

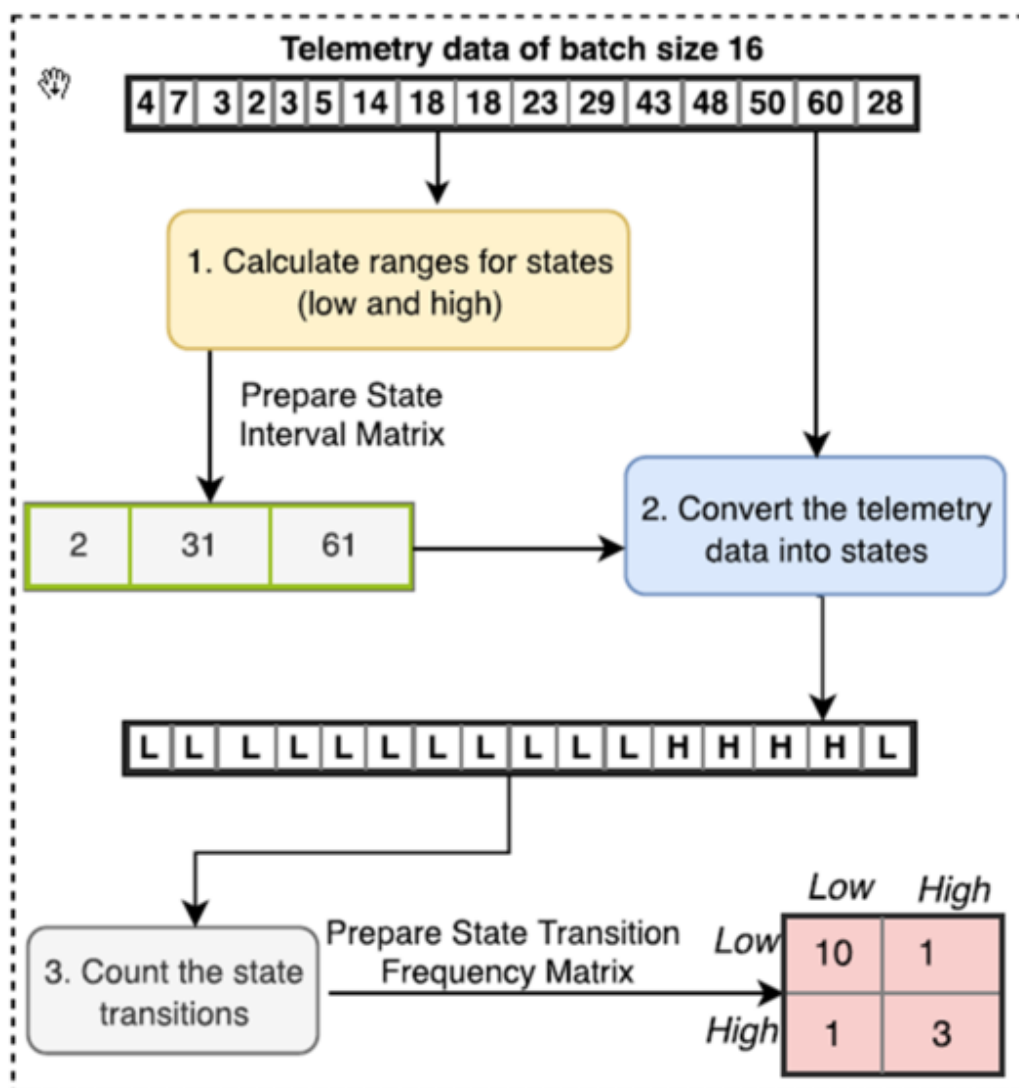
$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix}. \quad (5)$$

در روش ما، زنجیره ای می تواند از هر حالت به هر حالت دیگر منتقل شود و به عنوان یک زنجیره ergodic غیرقابل برگشت مارکوف در نظر گرفته می شود. برای زنجیره غیرقابل انعطاف مارکوف، عناصر ماتریس انتقال باید غیر منفی، $p_{ij} \geq 0$ و جمع هر سطر برابر با 1 باشد، بنابراین $\sum_{j=1}^n p_{ij} = 1$.

برای کاهش داده های تله متری داده شده، ابتدا آن را به ماتریس بازه زمانی $(n+1) \times 1$ $I = [\dots]$ تبدیل می کنیم و سپس ماتریس فرکانس انتقال حالت $n \times n$ $F = [\dots]$ را محاسبه می کنیم، جایی که n تعداد کل حالت ها است. ماتریس بازه حالت حاوی مقادیر آستانه برای تقسیم داده های داده شده به n حالت است. ماتریس فرکانس انتقال حالت شامل فرکانسهای انتقال است که با شمارش تعداد گذارها از یک حالت به حالت دیگر محاسبه می شوند.

شکل 4 فرایند محاسبه ماتریس فرکانس برای کاهش داده های تله متری را با استفاده از یک 2 MM-حالت با اندازه دسته 16 نشان می دهد. ابتدا ماتریس حالت بازه زمانی 1 که مرزهای پارتیشن را ذخیره می کند محاسبه می کنیم. برای 2 MM-حالت، ماتریس فاصله حالتی از اندازه 3×1 برای ذخیره مرزهای حالت خواهد بود. سپس،

ما با استفاده از ماتریس بازه زمانی حالت هر نقطه داده را به حالت های مربوطه تبدیل می کنیم. سرانجام، ماتریس فرکانس های انتقال حالت F را محاسبه می کنیم. فقط F و I را برای هر دسته از داده ها ذخیره می کنیم تا داده های تله متری را کاهش دهیم. در حین کاهش داده ها، ماتریس احتمال انتقال P را محاسبه و ذخیره نمی کنیم زیرا در بازسازی مورد نیاز است و می توان به راحتی با استفاده از F محاسبه شود. علاوه بر این، P از داده های شناور تشکیل شده است که همچنین سرریز ذخیره ای را معرفی می کند.



شکل 4: کاهش داده های تله متری با استفاده از مدل زنجیره ای مارکوف 2 حالت و اندازه دسته 16

برای بازسازی داده ها برای فاصله زمانی دسته ای، ابتدا ماتریس فرکانس انتقال حالت F را به ماتریس احتمال انتقال حالت P با استفاده از رابطه (6) تبدیل می کنیم.

$$p_{ij} = \begin{cases} \frac{f_{ij}}{\sum_{r=1}^x f_{ir}}, & \text{if } \sum_{r=1}^x f_{ir} \neq 0 \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

در حین بازسازی داده های دسته ای، فرض می کنیم که حالت فعلی/اولیه شناخته شده است و باید کلیه نقاط داده را با توجه به اندازه دسته ای پیش بینی کنیم. برای توضیح استراتژی بازسازی، فرض کنید یک 2 MM-حالت را همانطور که در شکل 3 نشان دادیم بررسی کنیم. احتمالات انتقال P را می توان به راحتی از ماتریس فرکانس انتقال مربوطه، که در مرحله کاهش داده ها ذخیره شده، محاسبه کرد. فرض کنیم که وضعیت فعلی زیاد است، احتمال انتقال از بالا به پایین (p_{hl}) برابر 0.3 و احتمال انتقال از بالا به بالا (p_{hh}) برابر 0.7 است. یک عدد تصادفی بین 0 تا 1 ایجاد می کنیم. اگر عدد تولید شده از 0.3 بیشتر باشد، پیش بینی می کنیم که در حالت دیگر بسیار پایین باشد. پس از شناسایی حالت بعدی، سپس مقادیر پایین و مرزی حالت پیش بینی شده را از ماتریس فاصله حالت A جستجو می کنیم. سرانجام، ما یک عدد تصادفی دیگر را در محدوده مرز حالت ایجاد می کنیم و آن را به عنوان نقطه داده بازسازی در نظر می گیریم. ما این فرآیند را تکرار می کنیم تا تمام نقاط داده را برای فاصله زمانی دسته بندی شده مشخص کنیم.

بخش چهارم: راه اندازی آزمایشی

A: توصیف مجموعه داده

برای ارزیابی روش پیشنهادی ما برای کاهش و بازسازی تله متری از داده IBM POWER8 استفاده شده است. این لاگها حاوی داده های تله متری است که با اجرای سه باره بارکردن نماینده Spark از Spark-Bench [46] ایجاد شده توسط IBM ایجاد شده و به طور گسترده با استفاده از سیستم های POWER8 مورد آزمایش قرار گرفته است. لاگهای مربوط به مجموعه داده ها از اجرای بارهای کاری "ماشینهای بردار پشتیبانی"، "PageRank" و "Spark SQL" جمع آوری می شوند. این حجم کارها در تحقیقات به خوبی شناخته شده است و ویژگی های مختلفی را برای پوشش دادن طیف وسیعی از رفتارهای مختلف استفاده از منابع ایجاد می کند. این

مجموعه داده شامل معیارهای مربوط به CPU، حافظه، سوئیچ های زمینه، پهنای باند حافظه، خطای حافظه کش L2 و L3، وقفه ها و CPI به عنوان داده های سری زمانی است.

B: جزئیات آزمایش

همانطور که قبلاً توضیح داده شد، ما دو تکنیک مختلف، یعنی مدل های PR و زنجیره مارکوف را پیشنهاد و ارزیابی می کنیم تا داده های تله متری را کاهش داده و از آن استفاده کنیم. برای ارزیابی روشهای پیشنهادی، دو آزمایش عمده انجام دادیم که به طور خلاصه در زیربخشهای زیر توضیح داده شده است.

1) آزمایش یک: کاهش و بازسازی داده ها با استفاده از PR.

در این آزمایش، ما اثر درجه های چند جمله ای های مختلف و اندازه های دسته ای در مدل های PR را بررسی می کنیم. درجه چند جمله ای شکل منحنی را مشخص می کند، جایی که می توان از درجه بالاتری برای جابجایی یک منحنی پیچیده استفاده کرد و از یک درجه پایین تر نیز می توان برای یک منحنی ساده استفاده کرد. اندازه دسته ای تعداد نقاط داده مورد استفاده برای متناسب بودن منحنی را مشخص می کند. ما درجه های چند جمله ای 2، 4، 6، 8 و 10 را با اندازه های مختلف از 2، 4، 8، 16، 32، 64 و 128 در نظر می گیریم تا منحنی های چند جمله ای برای کاهش داده ها و بازسازی داده ها در نظر گرفته شود. بسیاری از تنظیمات دیگر را نیز می توان مورد مطالعه قرار داد، با این وجود تنظیمات انتخاب شده برای ایجاد انگیزه استفاده از PR برای کاهش داده های تله متری و بازسازی آنها کافی است زیرا با درجات بالاتر، کاهش کمتری می گیریم و با اندازه های دسته ای بالاتر، دقت بازسازی داده ها را از دست می دهیم. هنگامی که ما یک مدل PR را با استفاده از درجه چند جمله ای خاص و اندازه دسته ای آموزش می دهیم، سپس تنها ضرایب معادله چند جمله ای را که یاد گرفته می شود در نقاط داده داده شده قرار بگیریم، ذخیره می کنیم. این امر به کاهش اندازه داده ها کمک می کند زیرا همه نقاط داده را ذخیره نمی کنیم، اما فقط مقادیر ضریب کمی دارند. بعداً از این ضرایب برای بازسازی نقاط داده به راحتی استفاده می شود. با این حال، برای دستیابی به دقت در بازسازی داده ها، باید از یک مقیاس چند کاره و مقیاس اندازه کارآمد استفاده شود. هرچه درجه بالاتر باشد، تناسب بیشتری به دست می آید، اما داده های بیشتری (ضرایب) منتقل می شوند.

(2) آزمایش دو: کاهش و بازسازی داده ها با استفاده از زنجیر مارکوف (MM)

در این آزمایش، اثر مدل های مختلف MM را با متغیر بودن تعداد حالت ها و اندازه های دسته ای بررسی می کنیم. ما 2، 3 و 4 حالت MM را در اندازه های دسته ای مختلف از 2، 4، 8، 16، 32، 64 و 128 برای بررسی اثرات در کاهش داده ها و بازسازی در نظر می گیریم. اندازه های دسته ای بالاتر دقت را کاهش می دهد زیرا بازسازی اطلاعات را که در مجموعه داده اصلی موجود نیست، حفظ نمی کند. از طرف دیگر، مدل های MM بالاتر باعث کاهش کمتری می شوند. شکل 3 یک MM 2-حالت را نشان می دهد، که در آن داده های ورودی را به دو ناحیه، کم و زیاد تقسیم می کنیم و یک ماتریس انتقال حالت برای مشاهدات تله متری را با استفاده از اندازه دسته ای خاص آموزش می دهیم. ماتریس انتقال حالت شامل احتمال انتقال از یک حالت به حالت دیگر با استفاده از نقاط داده داده ورودی معیارهای سنجش تله متری است. ما همچنین یک ماتریس بازه ای از حالت ایجاد می کنیم که محدوده منطقه کم و زیاد را تعریف می کند. این دو ماتریس به جای داده های کامل آموخته و ذخیره می شوند که باعث کاهش چشمگیر اندازه داده ها می شود. به همین ترتیب، برای MM 3 و 4، به ترتیب تعداد حالتها را به 3 و 4 افزایش می دهیم و معیارهای انتقال حالت و ماتریس فاصله زمانی را به همین ترتیب یاد می گیریم.

C. معیارهای ارزیابی

ما هر دو روش پیشنهادی را از نظر اثربخشی کاهش داده ها و دقت بازسازی ارزیابی می کنیم. برای ارزیابی اثربخشی کاهش داده ها، ما درصد کاهش داده ها را پس از اعمال روشهای پیشنهادی محاسبه می کنیم. برای دقت در بازسازی، ابتدا آزمایش K-S دو نمونه را انجام می دهیم تا تصمیم بگیریم که آیا بازسازی تولید شده داده ها را بعد از فشرده سازی، ذخیره سازی و بازسازی آن بپذیریم یا رد کنیم. این ارزیابی اول به منظور دور ریختن تنظیمات (درجه های PR و حالت های گسسته MM) که بازسازی هایی با کیفیت پایین دارند استفاده می شود. دوم، ما فاصله DTW بین داده های بازسازی شده و داده های واقعی را برای محاسبه ی خطای بازسازی محاسبه می کنیم. همه این اقدامات ارزیابی را در زیربخش های زیر توضیح می دهیم.

1) درصد کاهش داده ها

ما با کاهش داده های ذخیره شده پس از اعمال روش کاهش در برابر داده های اصلی، درصد کاهش داده ها را محاسبه می کنیم. مقدار مثبت، کاهش اندازه داده را نشان می دهد، در حالی که مقدار منفی نشان دهنده رشد در اندازه داده است. بنابراین درصد کاهش داده بالاتر بهتر و مطلوب است.

D. تست K-S دو نمونه ای

ما برای مقایسه شباهت آماری داده های واقعی با داده های بازسازی شده از آزمون K-S دو نمونه ای استفاده می کنیم. از آزمون K-S برای تعیین اینکه آیا دو توالی یک بعدی داده شده به توزیع احتمال یکسان تعلق دارند یا نه استفاده می شود. خروجی آزمایش K-S یک p -مقدار است. P -مقدار پایین تر یا مساوی 0.025 نشان می دهد که دو توالی داده شده از توزیع احتمال یکسان کشیده نشده اند. با این حال، مقدار p بالاتر از 0.025 نشان می دهد که دو دنباله داده شده از نظر آماری مشابه هستند [47]، [48]. بنابراین، در ارزیابی ما، p -مقدار را به دو ناحیه تقسیم می کنیم، یعنی قبول (A) وقتی p -مقدار از 0.025 بیشتر باشد و پذیرفته نشود (NA) وقتی p -مقدار کمتر یا مساوی 0.025 باشد.

E. چرخش زمان پویا

از آزمون K-S استفاده می شود تا ببینیم آیا توالی تولید شده از نظر آماری قابل مقایسه با نمونه اصلی است، اما شباهت پی در پی داده های بازسازی شده را کم نمی کند. بنابراین، برای اندازه گیری خطا در برابر داده های واقعی، از متریک فاصله DTW استفاده کردیم [21]، [22]. این یک روش شناخته شده است که برای اندازه گیری شباهت بین دو توالی داده شده استفاده می شود، که ممکن است در سرعت متفاوت باشد - [26] - [26]. به عنوان مثال، افزایش چشمگیر استفاده از CPU از دو مشاهده مختلف می تواند به عنوان یکسان شناخته شود حتی سرعت هر دو مشاهده برای دستیابی به یک استفاده غیر طبیعی متفاوت است. مقدار کمی از آزمون DTW خوب در نظر گرفته می شود زیرا نشان می دهد که دو دنباله داده شده به یکدیگر نزدیک هستند. با این حال، مقدار زیادی از آزمایش DTW بد تلقی می شود زیرا نشان می دهد که دو توالی داده شده نزدیک یکدیگر نیستند. بنابراین، مقدار کمتری از DTW برای در نظر گرفتن داده های بازسازی شده مشابه داده های واقعی مطلوب است.

بخش پنجم. نتایج تجربی

به منظور مطالعه و اعتبار سنجی روشهای ارائه شده، ما هر دو روش (MM و PRS) را با استفاده از مجموعه دادههای توصیف شده، با ارزیابی معیارهای فوق (K-S، DTW و بهبود کاهش) مقایسه می کنیم. همچنین رویکرد خود را در مقابل استفاده مستقیم از مکانیسم های فشرده سازی داده های کلاسیک مقایسه کردیم.

A. آزمون 1: کاهش و بازسازی داده ها با PR

در این آزمایش اول، ما کاهش داده های به دست آمده از استفاده از مدل های PR با اندازه های مختلف داده های دسته ای و درجه های چند جمله ای مختلف را محاسبه کردیم. سپس نتایج را با استفاده از آزمون K-S دو نمونه برای ارزیابی راه حل های نامناسب ارزیابی کردیم. سرانجام برای تعیین کمیت کیفیت حل، فاصله DTW را محاسبه کردیم. این ارزیابی در زیربخشهای زیر ارائه شده است.

1) تست K-S دو نمونه ای

جدول 1 نتایج حاصل از آزمایش K-S دو نمونه ای را نشان می دهد. ابتدا داده ها را با استفاده از ضرایب مربوطه، رهگیری، درجه چند جمله ای و اندازه های دسته ای برای هر مدل PR ذخیره شده در مرحله کاهش داده، بازسازی کردیم.

جدول 1: متوسط p -مقدار آزمایش $K-S$ دو نمونه ای با استفاده از PR با دستگاه های مختلف و اندازه های دسته برای کاربر CPU ، حافظه آزاد، سوئیچ های محتوا، و حافظه سخت افزاری پهنای باند با استفاده از آزمایشات 1

user cpu						context switches					
BS	PR2	PR4	PR6	PR8	PR10	BS	PR2	PR4	PR6	PR8	PR10
2	0.945	0.945	0.945	0.945	0.945	2	1	1	1	1	1
4	0.917	0.943	0.943	0.943	0.943	4	0.934	1	1	1	1
8	0.645	0.769	0.821	0.758	0.758	8	0.679	0.864	0.953	0.999	0.999
16	0.397	0.550	0.640	0.686	0.718	16	0.309	0.507	0.646	0.750	0.836
32	0.125	0.212	0.268	0.298	0.324	32	0.075	0.147	0.204	0.241	0.269
64	0.043	0.080	0.102	0.134	0.168	64	0.011	0.027	0.045	0.067	0.096
128	0.005	0.013	0.024	0.033	0.042	128	0.004	0.007	0.016	0.010	0.010
memory free						memory bandwidth					
BS	PR2	PR4	PR6	PR8	PR10	BS	PR2	PR4	PR6	PR8	PR10
2	1	1	1	1	1	2	1	1	1	1	1
4	1	1	1	1	1	4	0.989	1	1	1	1
8	0.991	0.998	0.999	1	1	8	0.826	0.914	0.974	0.988	0.988
16	0.924	0.952	0.967	0.974	0.976	16	0.500	0.684	0.831	0.873	0.901
32	0.725	0.814	0.849	0.857	0.870	32	0.210	0.389	0.473	0.515	0.544
64	0.468	0.555	0.637	0.695	0.725	64	0.063	0.132	0.215	0.297	0.371
128	0.217	0.293	0.363	0.404	0.429	128	0.012	0.022	0.047	0.075	0.103

The NA values are denoted by red line.

سپس از آزمون $K-S$ دو نمونه استفاده می کنیم تا میانگین p -مقدار برای هر اندازه گیری تله متری محاسبه شود. مقدار p پایین تر از 0.025 بد تلقی می شود و شباهت مناسبی با داده های واقعی نشان نمی دهد. با این حال، مقدار p بالاتر از 0.025 قابل قبول تلقی می شود و منعکس کننده امکان آماری داده های بازسازی شده متعلق به توزیع مشابه داده های واقعی است [47]، [48]. مقادیر NA با خط قرمز در جدول مشخص شده است.

بسیاری از مدل های PR داده ها را در یک محدوده قابل قبول بازسازی کردند. به عنوان مثال، برای متریک CPU کاربر، فقط چند جمله ای درجه 2، 4 و 6 با استفاده از اندازه دسته 128 داده ها را با دامنه قابل قبول بازسازی نکردند، در حالی که تمام مدل های PR دیگر دامنه قابل قبولی داشتند. در مورد متریک سوئیچ زمینه، همه مدل های PR با اندازه دسته 128 داده ها را در محدوده قابل قبول بازسازی نکردند، همان برای $PR2$ با اندازه دسته 64 بود، در حالی که تمام مدل های PR باقیمانده داده ها را در محدوده قابل قبول بازسازی می کردند.

(2) درصد کاهش داده ها

جدول 2 درصد کاهش داده ها را با استفاده از مدل های PR آموخته شده برای درجه های چند جمله ای مختلف و اندازه های دسته ای برای آزمایش 1 نشان می دهد. مقادیر منفی در جدول رشدی را به جای کاهش نشان می دهد. مقادیر منفی به دو دلیل مشاهده می شود. ابتدا، هر زمان که اندازه دسته ای (یعنی تعداد نقاط داده) کوچکتر یا مساوی با درجه چند جمله ای مورد استفاده برای متناسب کردن منحنی باشد، منجر به یادگیری ضرایب بیشتر از نقاط داده واقعی می شود. از این رو، اندازه داده ها در مقایسه با داده های اصلی رشد می کنند. دوم، بعضی اوقات ضرایب مدل های PR از مقادیر اعشاری با دقت بالا تشکیل شده است که در مقایسه با نقاط داده واقعی به فضای بیشتری احتیاج دارند. یک مدل درجه چند جمله ای بالاتر نسبت به مقادیر ضریب بسیار حساس است و گرد کردن مقادیر ضریب به طور قابل توجهی شکل منحنی را تغییر می دهد.

برای درک رشد داده ها با استفاده از مدل های PR، PR6 را با اندازه دسته 2 در نظر بگیرید که داده ها را 77.66٪ افزایش می دهد. این عمدتاً به این دلیل است که برای هر فاصله بین دسته ای، باید شش ضریب و یک رهگیری را ذخیره کنیم (در کل هفت نقطه داده)، در حالی که داده های واقعی فقط از دو نقطه داده تشکیل شده است. بنابراین، ما باید از قرار دادن یک منحنی در اندازه دسته ای کوچکتر از درجه چند جمله ای استفاده شده برای جاگذاری داده ها خودداری کنیم. مورد دیگر در نظر گرفتن PR6 با اندازه دسته 8 است که در آن اندازه داده ها عمدتاً به دلیل دقت بالای ضرایب با 47.74٪ افزایش می یابد.

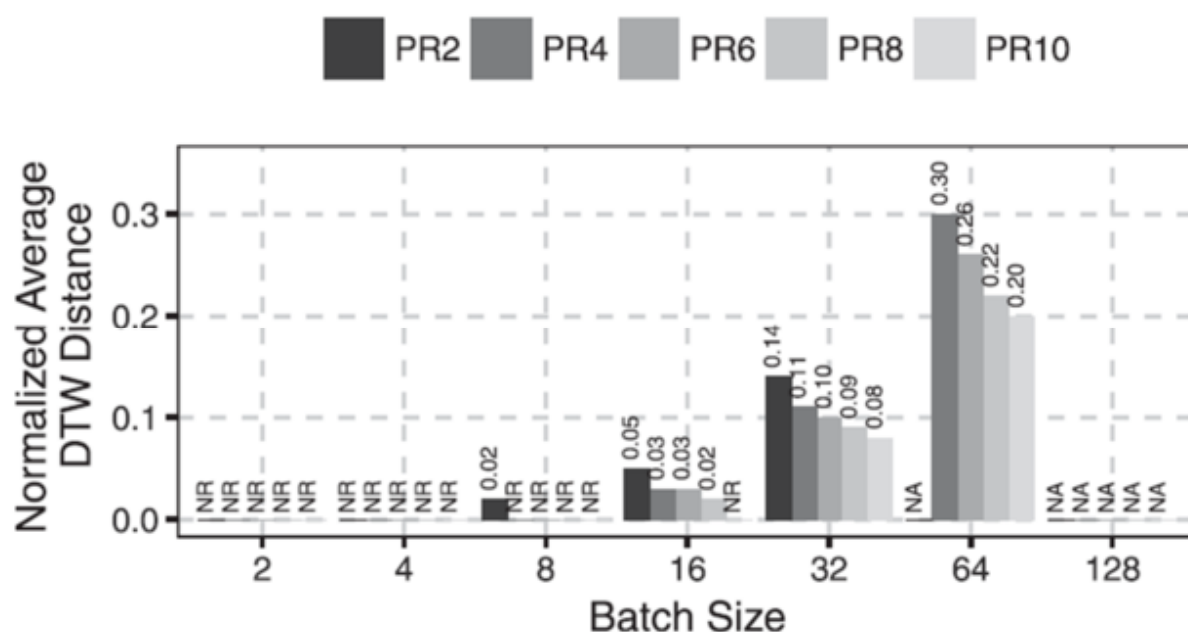
جدول 2: درصد کاهش داده ها با استفاده از PR با درجات چندجمله ای و اندازه های دسته مختلف برای آزمایش

1

Batch Size	PR2	PR4	PR6	PR8	PR10
2	-42.86	-60.36	-77.66	-94.96	-129.43
4	-3.58	-43.35	-51.82	-60.51	-77.89
8	39.21	-3.15	-47.72	-69.19	-78.22
16	68.06	46.78	25.54	4.39	-18.92
32	82.95	75.64	69.69	65.95	60.43
64	90.08	84.90	78.62	74.28	67.39
128	93.08	91.80	90.11	87.18	86.26

The red line values are not accepted and taken from table I. The grey shaded negative values represent data growth.

مشاهده کردیم که مدل های PR با اندازه دسته ای بزرگ و درجه چند جمله ای کوچک، در کاهش معنی داری داده های تله متری کمک می کنند. به عنوان مثال، اندازه دسته 64 با PR4 داده ها را به 84/90 درصد کاهش می دهد. با این حال، چنین مدل های PR ممکن است داده ها را با دقت مناسب به طور خاص برای مشاهده های تله متری پراکنده و پر نویز بازسازی نکنند. بنابراین، برای کاهش اندازه داده ها با دقت بیشتر در بازسازی داده ها، باید یک ترکیب مناسب از اندازه دسته ای و درجه چند جمله ای را شناسایی کنیم.



شکل 5: میانگین فاصله DTW نرمالیزه برای اندازه های تله متری با استفاده از مدل های PR آموزش دیده در اندازه های مختلف دسته ای و درجات چند جمله ای برای آزمایش 1.

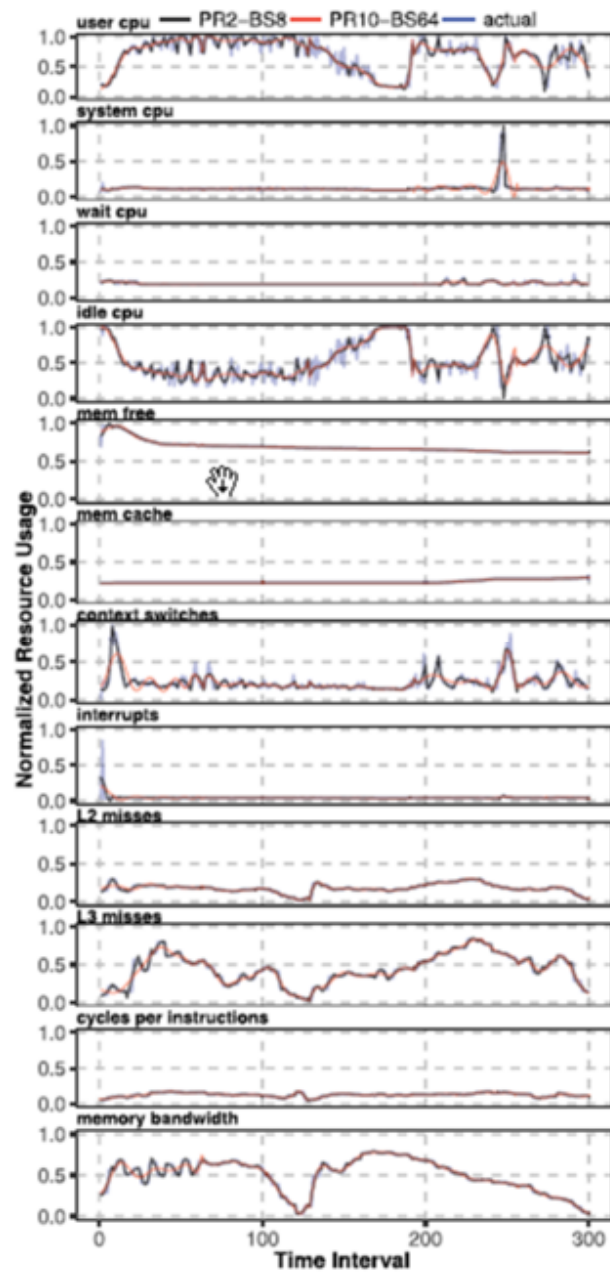
(NA = کاهش نیافته، پذیرفته نمی شود)

3) فاصله DTW

شکل 5 فاصله طبیعی DTW عادی را برای مدل های PR آموزش دیده با استفاده از درجه های مختلف و اندازه های دسته ای برای آزمایش 1 نشان می دهد. اندازه های دسته بندی بدون کاهش توسط NR نشان داده می شود و اندازه های دسته ای که در مناطق قابل قبول نیست توسط NA مشخص می شوند. ما مشاهده کردیم که در اندازه های دسته بزرگ، فاصله DTW افزایش می یابد و با افزایش درجه PR کاهش می یابد. همچنین مشاهده کردیم که در اندازه دسته 8، فقط درجه 2 (PR2) داده ها را با دقت 98٪ کاهش می دهد. دقت داده های بازسازی شده در اندازه دسته 16 از 95٪ تا 98٪ متغیر است و در سایز 64 به 70٪ تا 80٪ کاهش می یابد.

شکل 6 بازسازی 300 نقطه داده اولیه از بار کاری PageRank از مجموعه داده های ما با استفاده از مدل های مختلف رگرسیون آموزش داده شده در درجه 2 با اندازه دسته پایین تر (BS8) و در درجه 10 با اندازه دسته بالاتر (BS64) ارائه می دهد. ما مشاهده کردیم که BS 8 با درجه 2 بازیافت بهتری برای لبه و پیشرفت در مقایسه با

BS 64 به دست می دهد، اما این امر فقط باعث کاهش داده ها 39.21٪ شد. با این حال، BS 64 با درجه 10 بازده 67.39٪ کاهش داده ها، آزمون K-S نیز قابل قبول است، اما فاصله DTW برای مدل BS 64 بالاتر از مدل BS 8 برای تمام معیارهای سنجش تله متری است. این تأیید می کند که PR با درجه بالاتر و اندازه دسته ای بالاتر، دقت کمتری را در داده های بازسازی شده نشان می دهد.



شکل 6: بازسازی داده های تله متری بار کاری PageRank با استفاده از درجه چند جمله ای 2 با اندازه دسته 8 (PR2-BS8) و درجه چند جمله ای 10 با اندازه دسته 64 (PR10-BS64).

B. آزمایش 2- کاهش بازسازی داده ها با استفاده از MM

در آزمایش دوم، ما کاهش داده های به دست آمده با استفاده از MM با اندازه های مختلف داده های مختلف و تعداد متفاوت حالت های مارکوف را محاسبه کردیم. سرانجام برای تعیین کمیت کیفیت حلها، فاصله DTW را محاسبه کردیم. بخش های زیر نتایج ارزیابی را ارائه می دهند.

1) آزمایش دو نمونه ای K-S

جدول 3 نتایج آزمون K-S دو نمونه ای را نشان می دهد. بسیاری از داده های بازسازی شده MM در محدوده قابل قبول هستند مگر محدود مواردی که معیارهای CPU کاربر بازسازی شده با استفاده از MM 2-حالت با اندازه دسته 128 از آنها در محدوده قابل قبول نیست. معیارهای بازسازی شده از حافظه با استفاده از مدل های MM با اندازه دسته 128 در محدوده قابل قبول نیست. نتایج این واقعیت را نتیجه می گیرد که کلیه معیارهای تله متری بازسازی در هنگام استفاده از MM با 2، 3 و 4 حالت و تا اندازه 64 دسته در محدوده قابل قبول هستند.

جدول 3: p - مقدار متوسط در آزمایش $K-S$ دو نمونه ای با استفاده از MM با وضعیت ها و اندازه های لبه مختلف برای کاربر CPU ، حافظه آزاد، سوئیچ های محتوا و پهنای باند سخت افزار حافظه در آزمایش 2

user cpu				context switches			
BS	2MM	3MM	4MM	BS	2MM	3MM	4MM
2	0.999	0.999	0.999	2	0.996	0.996	0.996
4	0.663	0.674	0.663	4	0.745	0.766	0.775
8	0.403	0.401	0.408	8	0.579	0.622	0.656
16	0.281	0.285	0.294	16	0.408	0.485	0.543
32	0.181	0.203	0.215	32	0.237	0.331	0.396
64	0.076	0.113	0.126	64	0.107	0.189	0.246
128	0.023	0.037	0.060	128	0.013	0.053	0.092
memory free				memory bandwidth			
BS	2MM	3MM	4MM	BS	2MM	3MM	4MM
2	1	1	1	2	1	1	1
4	1	1	1	4	0.800	0.843	0.892
8	0.998	0.998	0.998	8	0.271	0.328	0.354
16	0.937	0.927	0.906	16	0.115	0.139	0.129
32	0.658	0.580	0.545	32	0.082	0.068	0.072
64	0.124	0.088	0.063	64	0.057	0.053	0.051
128	0.002	0.003	0.004	128	0.026	0.028	0.034

The NA values are denoted by red line.

2) درصد کاهش داده ها

جدول 4 درصد کاهش داده ها را برای اندازه های مختلف دسته ای با MM 2، 3 و 4 نشان می دهد. مقادیر منفی در جدول رشد داده را به جای کاهش نشان می دهد. مقادیر منفی هر زمان مشاهده می شود که تعداد نقاط داده آموخته شده به عنوان بخشی از مدل MM بیشتر از نقاط داده واقعی شود.

جدول 4: درصد کاهش داده ها با استفاده از MM با وضعیت ها و اندازه ای لبه مختلف

Batch Size	2MM	3MM	4MM
2	-30.15	-106.06	-198.15
4	39.00	2.38	-42.06
8	70.05	53.08	31.50
16	83.09	75.13	63.77
32	90.28	84.64	79.38
64	92.81	91.30	87.30
128	93.95	93.16	92.16

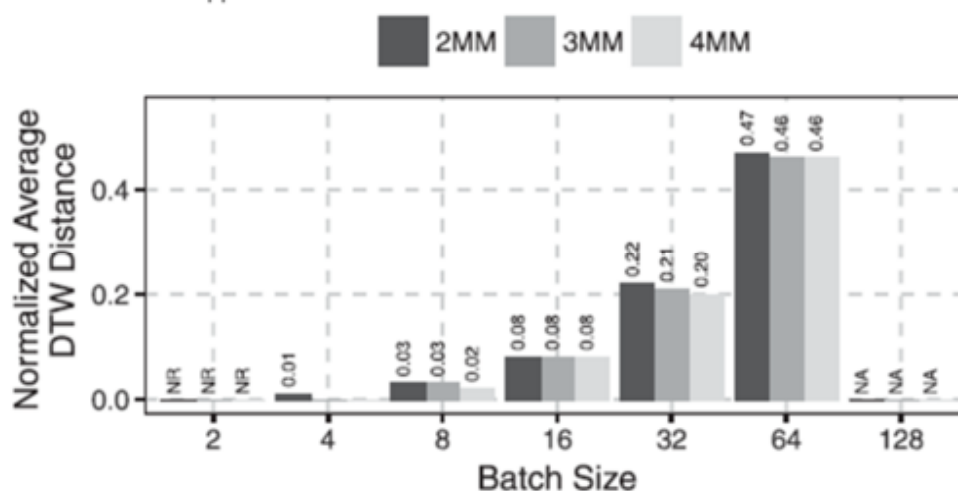
The red line values are NA and taken from Table III. The gray shaded negative values represent data growth.

مشاهده کردیم که اندازه های دسته بزرگ و مقادیر حالت کوچک باعث کاهش داده های بالاتر می شوند. به عنوان مثال، اندازه دسته 64 با 2 حالت بازده 92.81٪ داده را نشان می دهد. برای 2 و 3 حالت MM، ما شروع به مشاهده کاهش داده ها بعد از اندازه گروه بیشتر از 2 می کنیم. در حالی که برای MM، 4 حالت، کاهش داده ها را برای اندازه های دسته ای بیشتر از 4. مشاهده می کنیم. از این رو، کاهش داده به اندازه ماتریس احتمال انتقال و ماتریس فاصله زمانی بستگی دارد. به عنوان مثال، برای ذخیره احتمالات انتقال حالت و 3 نقطه داده به 2 ماتریس نیاز داریم تا ماتریس بازه حالت را نشان دهیم. بنابراین، برای MM دو حالت، حداقل هفت نقطه داده نیاز داریم تا یک دسته داده شده را نمایان کنیم. ما 39٪ کاهش داده ها را در اندازه گروه 4 مشاهده کردیم که نقاط اصلی داده 4 است و نقاط داده MM، 7 است. دلیل این کاهش این است که برخی از معیارهای سنجش تله متری در داده های واقعی خود حداکثر نه رقم دارند، به عنوان مثال سوئیچ های متن، وقفه را قطع می کند در حالی که ماتریس احتمال انتقال حالت فقط تعداد جابجایی از یک حالت به حالت دیگر را شامل می شود. بنابراین حتی اگر نقاط داده MM از اندازه دسته اصلی بالاتر باشد، به فضای کمتری نیاز دارد. این تعداد بعداً در صورت نیاز به بازسازی، به احتمالات تبدیل می شود.

(3) فاصله DTW

شکل 7 میانگین فاصله طبیعی DTW نرمالیزه را برای MM آموخته شده با استفاده از درجه ها و اندازه های دسته مختلف برای آزمایش 2 نشان می دهد. اندازه های دسته بندی بدون کاهش توسط NR نشان داده شده و اندازه های دسته ای که در مناطق قابل قبول نیست توسط NA مشخص می شود. مشاهده کردیم که در اندازه های دسته بزرگ فاصله DTW افزایش می یابد. ما همچنین مشاهده کردیم که در اندازه دسته 4، تنها مدل MM، حالت 2 (MM2) داده ها را با دقت 99٪ کاهش می دهد. دقت داده های بازسازی شده در اندازه دسته 8 از 97٪ به 98٪ و در اندازه ی دسته 16 متغیر است، که 92٪ است زمانی که در اندازه دسته 64 به 53٪ تا 54٪ کاهش می یابد. مدل های MM اثر خاصی را روی اندازه های تله متری که شامل لبه، توالی یا نویز نیست را نشان نمی دهند. به عنوان مثال، اندازه گیری تله متری حافظه آزاد (MEM آزاد) از مجموعه داده های ما با استفاده از مدل های مختلف MM که در حالت ها و اندازه های دسته ای مختلف MM آموزش دیده اند، هیچ تاثیری در DTW ندارد.

بنابراین، چنین نوع اندازه گیری تله متری می تواند با استفاده از تعداد کمی از حالت های MM بازسازی شود.

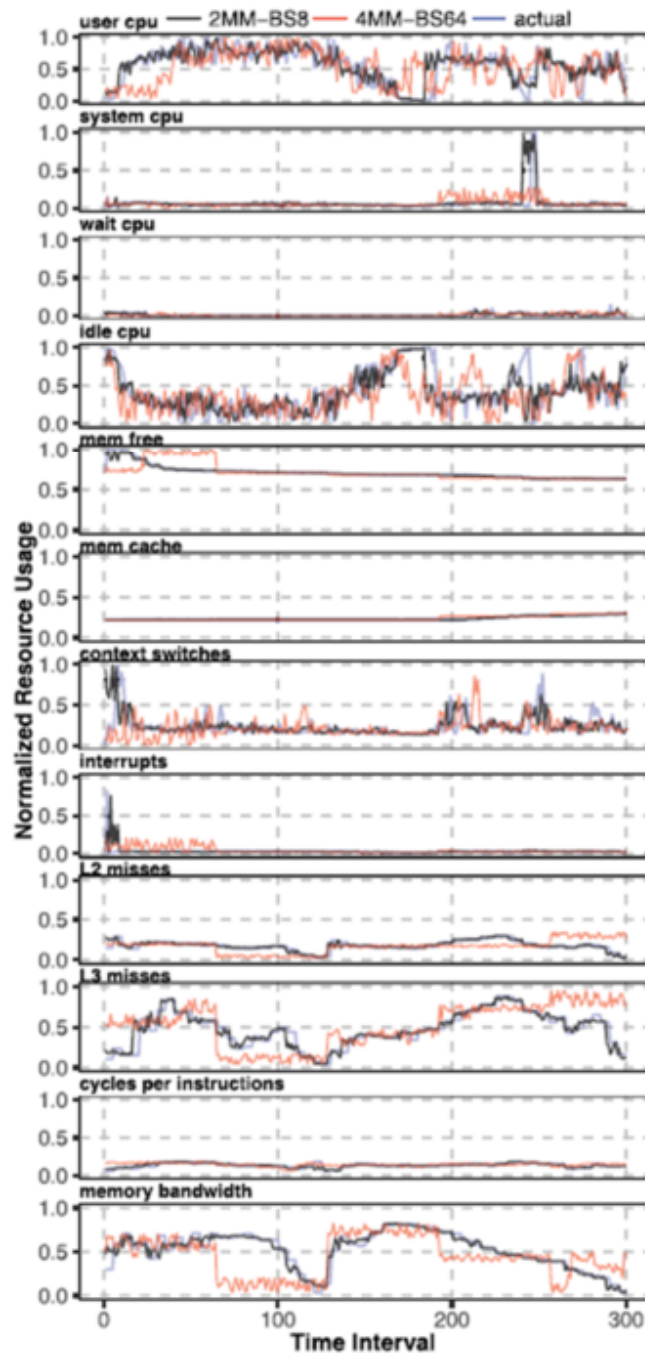


شکل 6: میانگین فاصله DTW نرمالیزه برای MM آموزش داده شده در اندازه های مختلف دسته ای و حالت های مربوط به اندازه گیری تله متری. (NR=کاهش نیافته ، NA=پذیرفته نمی شود).

شکل 8 بازسازی 300 نقطه اولیه داده از بار کاری PageRank از مجموعه داده های ما با استفاده از MM های مختلف آموزش داده شده با استفاده از 2 حالت و اندازه دسته ای 8 (2MM-BS8) و 4 حالت و اندازه دسته ای

64 (4MM-BS64) را نشان می دهد. ما مشاهده کردیم که BS 8 با MM 2 حالت داده ها را بطور مناسب بازسازی کرد. با این حال ، با مدل 4 حالت با اندازه دسته 64 بازسازی خوبی را نشان نمی دهد. کاهش داده ها با استفاده از مدل MM-BS82 فقط 70.05٪ داده را کاهش می دهد، در حالی که MM-BS644 باعث کاهش 87/30٪ از داده ها می شود. آزمون K-S برای هر دوی این مدلها نیز قابل قبول است، در حالی که فاصله DTW برای مدل BS64 بالاتر از BS8 برای تمام معیارهای سنجش تله متری است.

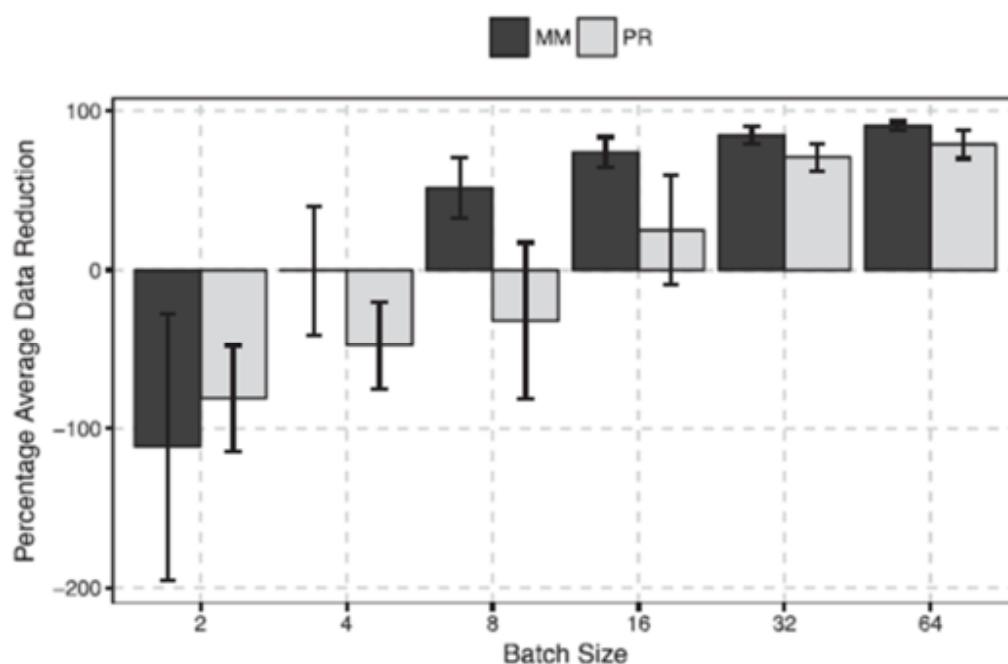
استفاده از یک اندازه دسته کوچک (BS8) به گرفتن الگوهای داده کمک می کند، از جمله لبه ها و ترکیدگی در مقایسه با یک دسته بزرگ (BS64). این نشان می دهد که لبه ها و نويز در داده های اصلی با استفاده از اندازه های دسته بزرگتر نمی توانند به خوبی ضبط شوند. به عنوان مثال، سیستم CPU لبه های کمی در داده های واقعی حدود 270 ثانیه داشت. این لبه ها با استفاده از اندازه دسته 8 به خوبی ضبط می شوند، اما اندازه 64 این دسته ها به خوبی جذب نمی شود. دلیل این رفتار عمدتاً به این دلیل است که MM به ماتریس انتقال حالت بستگی دارد، که شامل احتمال انتقال از یک حالت به حالت دیگر است. بنابراین، اگر در یک دسته خاص، توالی وجود داشته باشد، احتمال وقوع پشت سر هم برای کل بازه دسته ای باقی می ماند. بنابراین عملکرد روش MM در بازسازی داده ها همیشه با استفاده از اندازه دسته های بزرگ قوی نیست. نتیجه می گیریم که اگر داده ها حاوی لبه و ترکیدگی باشند، باید از متدهای MM با اندازه کوچکتر استفاده کنیم. با این حال، اگر بتوانیم تشخیص دهیم که داده ها حاوی لبه و شکنندگی نیستند، می توانیم از روش MM با اندازه های بالاتر استفاده کنیم. این رفتار به خوبی در معیارهای سنجش تله متری بدون حافظه و حافظه نهان مشاهده شده در شکل 8 مشاهده می شود.



شکل 8: بازسازی داده های تله متری بار کاری PageRank با استفاده از MM دو حالت با اندازه دسته 8 (2MM-BS8) و 4 حالت MM با اندازه دسته 64 (4MM-BS64)

C. مقایسه PR و MM

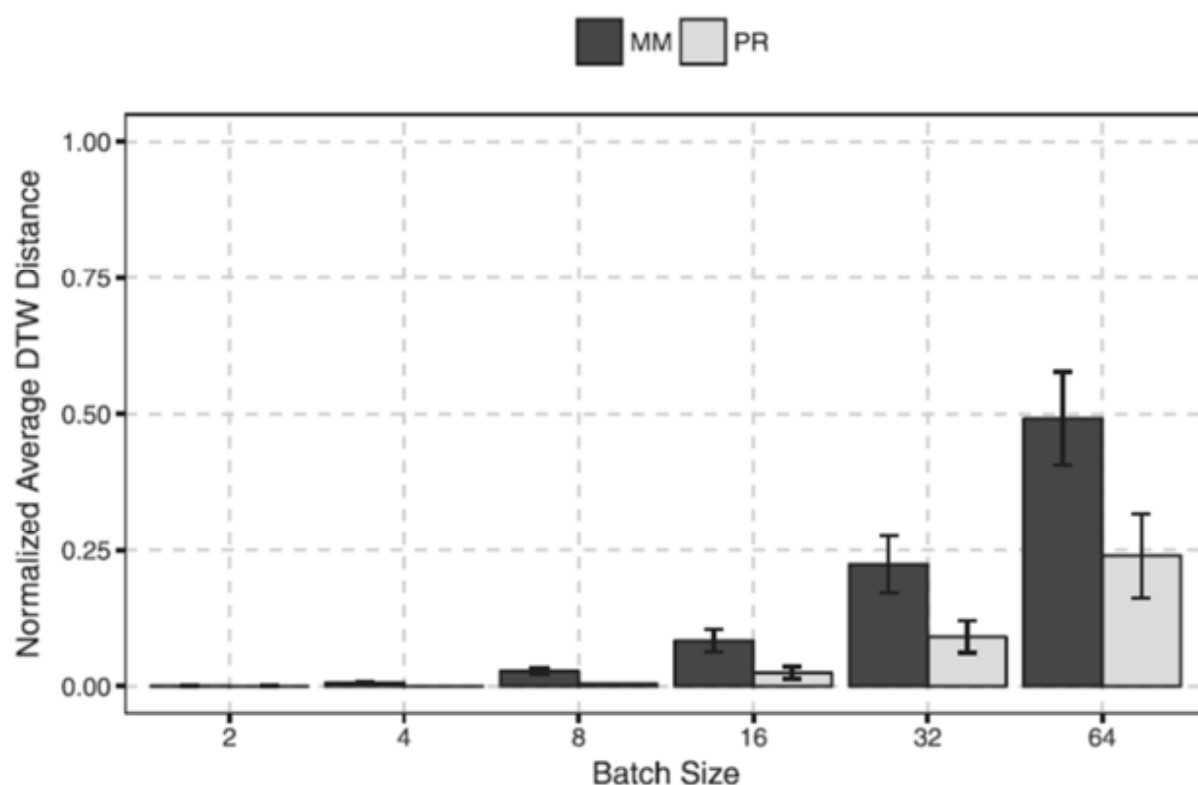
برای مقایسه روشهای PR و MM، میانگین درصد کاهش داده ها و میانگین فاصله DTW را برای کلیه تنظیمات (درجه چند جمله ای و حالت) این روش ها در اندازه های مختلف دسته ای محاسبه می کنیم. در این بخش، مقایسه روش های PR و MM برای کاهش داده ها و بازسازی داده ها را نشان می دهیم. شکل 9 میانگین داده ها را در اندازه های مختلف دسته ای برای کلیه تنظیمات با استفاده از روش های PR و MM نشان می دهد. به طور متوسط، اندازه های گروه کمتر از 8 هیچ گونه کاهش داده ای را به همراه نمی آورد، اما در عوض باعث رشد داده ها می شوند. با این حال افزایش قابل توجهی از 51.54 با استفاده از روش MM در اندازه دسته 8 نسبت به 31.81٪ برای روش PR مشاهده شده است. برای اندازه های دسته ای بالاتر از 8، MM همیشه در کاهش داده از PR بهتر عمل می کند.



شکل 7: مقایسه میانگین کاهش درصد داده ها با استفاده از روش های PR و MM برای کاهش داده های تله متری.

شکل 10 میانگین فاصله DTW نرمالیزه برای روش PR و MM را در اندازه های مختلف دسته نشان می دهد. به طور متوسط، اندازه دسته ای کمتر از 16 بازده فاصله DTW بسیار کم، کمتر از 0.2 برای هر دو روش PR و MM

دارد که این نشان دهنده شباهت خوبی از داده های بازسازی شده با داده های واقعی است. برای اندازه های بزرگ دسته، به عنوان مثال، 32 و 64، روش PR از بازسازی داده از MM بهتر است. اندازه های دسته بزرگ، داده ها را به میزان قابل توجهی کاهش می دهد، اما از نظر شباهت بازسازی داده ها، عملکرد ضعیفی دارند. بنابراین، از شکل های 10 و 9 نتیجه می گیریم که اندازه دسته 16 برای استفاده با هر دو روش PR و MM مناسب است زیرا ما برای هر دو روش PR و MM به طور متوسط 25.17٪ و 74٪ کاهش داده با فاصله DTW زیر 0.2 برای هر دو روش را بدست می آوریم. با این حال، ما ترجیح می دهیم از مدل MM استفاده کنیم، عمدتاً به دلیل کاهش بیشتر داده ها در اندازه دسته 16، اگرچه روش PR دارای DTW کمی بهتر است.

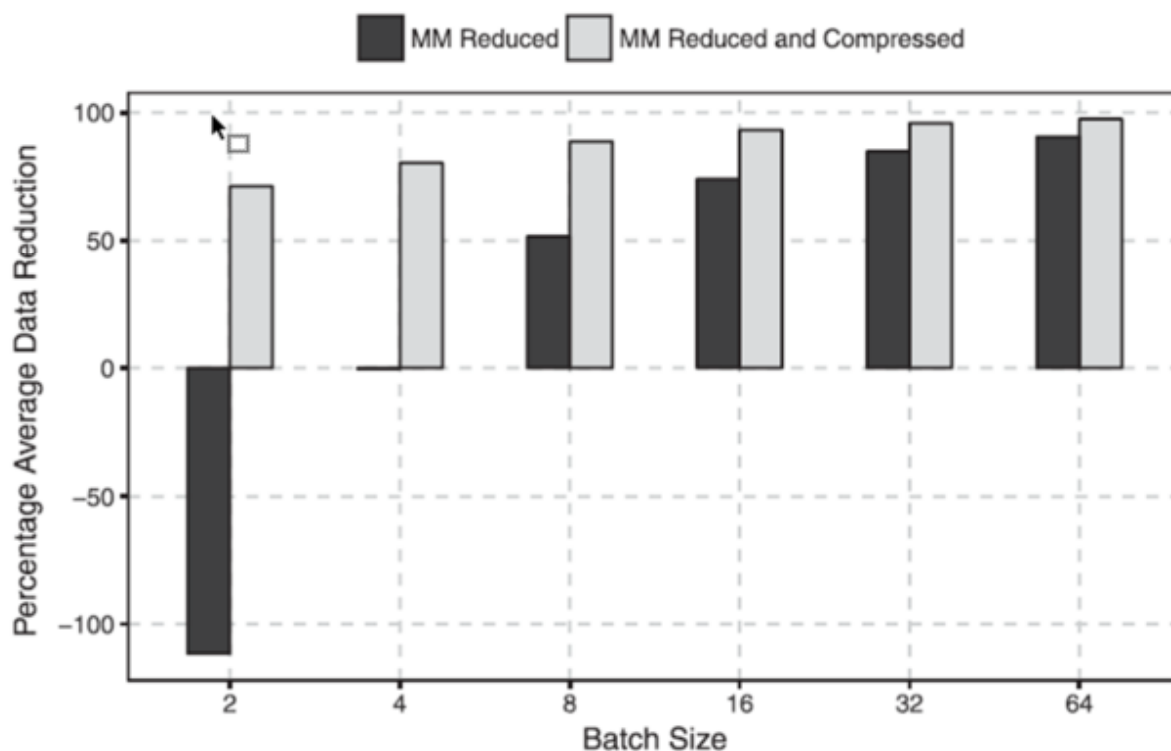


شکل 10: مقایسه فاصله نرمالیزه DTW برای بازسازی داده ها با استفاده از روش PR و MM.

D. کاهش داده های به استفاده از فشرده سازی ZIP

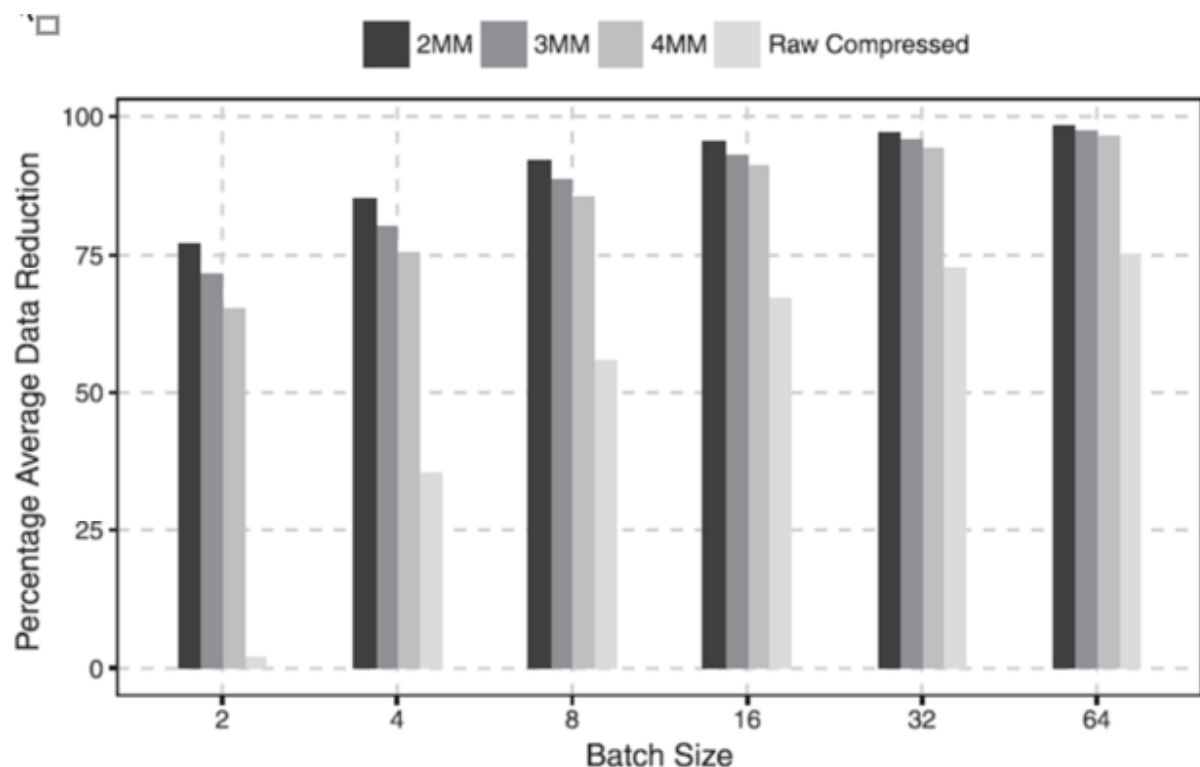
پس از مقایسه دو روش پیشنهادی، ما آنها را در برابر استفاده از روش های فشرده سازی بدون ضرر کلاسیک، مانند الگوریتم های ZIP ارزیابی می کنیم. ما الگوریتم ZIP را بر روی داده های خام با استفاده از الگوریتم ZIP در

داده های کاهش یافته مقایسه کردیم تا شاهد بهبود یا سربار باشیم. شکل 11 مقایسه داده های کاهش یافته با استفاده از MM (MM کاهش یافته) و داده های بدست آمده پس از اعمال فشرده سازی ZIP بر روی داده های کاهش یافته (MM کاهش یافته و فشرده شده) برای اندازه های مختلف دسته ای را نشان می دهد. ما مهمترین کاهش داده را به دلیل فشرده سازی ZIP در اندازه های دسته پایین به ویژه 2، 4 و 8 مشاهده کردیم زیرا این است که داده های کاهش یافته شامل تعداد انتقال در ماتریس انتقال هستند که در صورت 2 یا 0 یا 1 است. حالات 0، 1 یا 2 در صورت 4 حالت و 0 در 7 مورد در مورد 8 حالت. با این وجود، بیشتر مقادیر در این ماتریس ها به دلیل عدم انتقال از یک حالت به حالت دیگر از 0 تشکیل شده است، از این رو فشرده سازی فشرده سازی داده ها را بیشتر می کند. بعد از اندازه دسته ای 32، مشاهده کردیم که تفاوت بین داده های MM کاهش یافته و MM کاهش یافته فشرده شده به طور متوسط کمتر از 8 درصد است. حداکثر کاهش داده ها با فشرده سازی فشرده با مدل MM 2- حالت مشاهده شده است، که در اندازه دسته 64، 98.24 درصد است. فشرده سازی ZIP به کاهش بیشتر داده ها کمک می کند، اما برای اندازه های بزرگ دسته ای، اثر فشرده سازی ZIP معنی دار نیست.



شکل 8: مقایسه داده های MM کاهش یافته با داده های MM کاهش یافته و فشرده شده. MM کاهش یافته داده های بدست آمده با استفاده از روش MM است و MM کاهش یافته فشرده شده داده ای است که پس از اعمال فشرده سازی ZIP بر روی داده های کاهش یافته حاصل از روش کاهش MM بدست می آید.

سپس با استفاده از یک الگوریتم استاندارد فشرده سازی ZIP بدون ضرر به عنوان پایه، شکل 12 مقایسه فشرده سازی ZIP در داده های خام با داده های کاهش یافته و فشرده شده (MM فشرده شده کاهش یافته) را با استفاده از روش MM در اندازه های مختلف دسته نشان می دهد. فقط با اعمال فشرده سازی ZIP بر روی داده های ردیف، ما برای داده های دسته ای از سایز 2 تا 64 به 1.88٪ به 75.02٪ دست یافتیم، اما راه حل پیشنهادی ما منجر به کاهش از 76.95٪ به 96.48٪ با مدل MM 2 و 4 از اندازه دسته به ترتیب 2 و 64 می شود.



شکل 9: مقایسه فشرده سازی ZIP در داده های خام و فشرده سازی ZIP بر روی داده هایی که با استفاده از روش MM کاهش می یابد.

E. کاهش پهنای باند با استفاده از MM

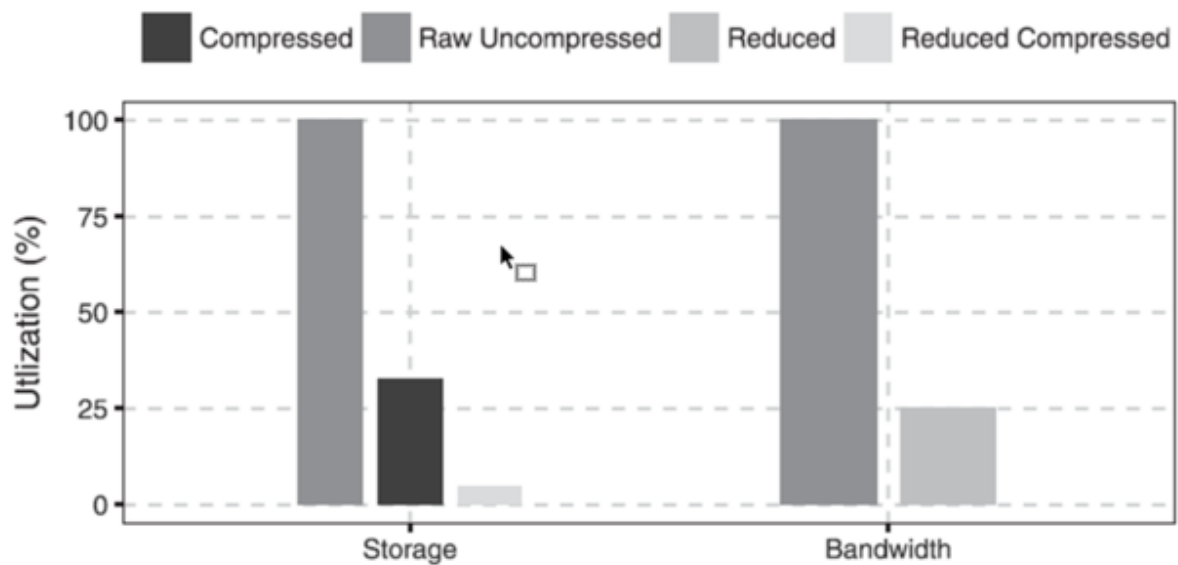
جدول 5 درصد کاهش استفاده از پهنای باند را با استفاده از راه حل پیشنهادی با 2، 3 و 4 حالت MM برای جمع آوری داده های تله متری در مرکز داده نشان می دهد. ما نوع داده شناور 4 بایت را در نظر می گیریم تا مقدار داده های واقعی و ماتریس فاصله زمانی حالت را نشان دهیم. مشاهده شد که برای بالاترین دسته قابل قبول (اندازه دسته 64)، ماتریس انتقال حالت می تواند حداکثر مقدار 63 را داشته باشد زیرا شامل تعداد انتقال از یک حالت به حالت دیگر است. این نوع داده ها را می توان با استفاده از 1 بایت از طریق شبکه ارسال کرد که باعث کاهش قابل توجهی در استفاده از پهنای باند می شود. علاوه بر این، اندازه های دسته بالاتر باعث کاهش قابل توجه در استفاده از پهنای باند در مرکز داده می شود. به عنوان مثال، اندازه دسته 32 و 64 بازده پهنای باند قابل توجهی دارد که از تعداد 71.88٪ تا 93/75٪ برای تعداد مختلف حالات MM متغیر است.

جدول 5: کاهش درصد پهنای باند محتوا در مرکز داده ها با استفاده از روش MM

Batch Size	2MM	3MM	4MM
2	-100	-212.50	-350
4	0	-56.25	-125
8	50	21.88	-12.50
16	75	60.94	43.75
32	87.50	80.47	71.88
64	93.75	90.23	85.94

بخش ششم: مزایای استفاده از حل پیشنهادی در مراکز مختلف داده

معیارهای معمولی تله متری متشکل از پیشخوان های عملکرد سخت افزاری تا صد رویداد [49] مربوط به CPU، حافظه، شبکه، دیسک، دما و غیره است. در این بخش، ما تأثیر استفاده از پهنای باند و فضای ذخیره سازی را با استفاده از راه حل پیشنهادی برای نظارت بررسی می کنیم و داده های تله متری را در اندازه های مختلف مراکز داده جمع آوری می کنیم. سه مرکز داده مختلف یعنی کوچک، متوسط و بزرگ را در نظر گرفتیم که فرض می کنیم 100 تله متری مختلف پس از یک بازه زمانی گسسته از هر گره محاسباتی جمع آوری می شود. یک نمونه از متریک تله متری به 8 بایت حافظه برای ذخیره اطلاعات از جمله زمان سنج نیاز دارد. ما مدل MM دو حالت را با اندازه دسته 16 به عنوان حل هدفمند خود انتخاب کردیم زیرا کاهش 95/33 درصد در فضای ذخیره سازی، 75 درصد کاهش در پهنای باند با 92 درصد دقت به دست می آوریم. شکل 13 مقایسه درصد فضای ذخیره سازی مورد نیاز برای داده های خام فشرده نشده، فشرده شده و کاهش یافته را نشان می دهد. همچنین درصد استفاده از پهنای باند برای داده های فشرده نشده و کاهش یافته خام برای حل مورد نظر را نشان می دهد.



شکل 10: مقایسه درصد استفاده از ذخیره سازی برای داده های خام فشرده سازی نشده، فشرده و کاهش یافته (سمت چپ). مقایسه درصد استفاده از پهنای باند برای داده های خام فشرده نشده و داده های کاهش یافته (سمت راست).

جدول 6 میزان ذخیره و پهنای باند را برای هر سه نوع مرکز داده نشان می دهد و شایستگی راه حل پیشنهادی را برجسته می کند. یک مرکز داده کوچک با 3000 گره محاسباتی که در 100 رکورد مستقر شده است. جایی که هر قفسه میزبان 30 گره محاسباتی است. برای ذخیره یک روز داده های تله متری، ما به ذخیره سازی 193.11 گیگابایت نیاز داریم و در یک ماه که نیاز به ذخیره سازی به 5.65 TB افزایش می یابد. برای ذخیره یک روز داده های تله متری در یک مرکز داده با اندازه متوسط، به عنوان مثال، شمارش با 10 گره محاسباتی، مستقر در 250 قفسه (با در نظر گرفتن 40 گره در هر قفسه)، به 643.73 گیگابایت حافظه نیاز داریم و در یک ماه، میزان ذخیره سازی افزایش می یابد. تا 18/85 ترابایت. به طور مشابه، برای یک مرکز داده بزرگ، به عنوان مثال، با 30 گره محاسباتی مستقر در 500 قفسه، ما به یک فضای ذخیره سازی 1931.19 گیگابایت برای یک روز نیاز خواهیم داشت و در یک ماه، نیازهای ذخیره سازی به 56.57 TB افزایش می یابد.

روش پیشنهادی MM دو حالت با 16 عدد لبه، میزان ذخیره سازی را به ترتیب برای مراکز داده کوچک، متوسط و بزرگ به ترتیب با 0.26، 0.88 و 2.64 ترابایت با دقت بازسازی 92 درصد کاهش داد. میزان استفاده از پهنای باند

واقعی رک برای مراکز داده کوچک، متوسط و بزرگ به ترتیب 187.50، 250 و 375 Kbps است که به طور قابل توجهی به 46.88، 62.50 و 93.75 Kbps کاهش می یابد. به طور مشابه، استفاده از پهنای باند سطح مرکز داده های واقعی برای اندازه های کوچک، متوسط و بزرگ به ترتیب 18750، 62500 و 187500 Kbps است. سیستم هدفمند ما این استفاده از پهنای باند را به ترتیب برای مراکز داده های کوچک، متوسط و بزرگ به ترتیب به 4687.50، 15625 و 46875 Kbps تبدیل می کند.

جدول 6: مقایسه ذخیره خام و کاهش یافته (GB) به مدت 30 روز و پهنای باند (KBPS) با استفاده از 2 MM - حالتی با اندازه 16

Data Center Type	Total Racks	Nodes per Rack	Total Telemetry Sources	30 days Storage (GB)			Bandwidth (Kbps)			
				Raw Uncompressed	Compressed	Reduced Compressed	Raw Uncompressed at Rack	Reduced at Rack	Raw Uncompressed at Datacenter	Reduced at Data Center
small	100	30	300,000	5,793.57	1,158.71	270.56	187.50	46.88	18,750	4,687.50
medium	250	40	1,000,000	19,311.90	3,862.38	901.87	250	62.50	62,500	15,625
large	500	60	3,000,000	57,935.71	11,587.14	2,705.60	375	93.75	187,500	46,875

بخش هفتم: نتیجه گیری و تحقیقات آینده

مراکز داده تعداد زیادی از داده های تله متری را تولید می کنند، که برای بسیاری از اهداف از جمله مدیریت منابع، تجزیه و تحلیل و بهینه سازی استفاده می شود. با این حال، اندازه داده های تله متری بطور چشمگیری رشد می کند و فضای ذخیره سازی و استفاده از پهنای باند در مرکز داده را به میزان قابل توجهی افزایش می دهد. در این مقاله، ما یک روش مبتنی بر زنجیره مارکوف را برای کاهش داده های تله متری برای به حداقل رساندن استفاده از پهنای باند و فضای ذخیره سازی مورد نیاز برای ذخیره آن در مرکز داده پیشنهاد داده ایم. راه حل ما از روش پایه بر اساس روش PR نسبت به کاهش و بازسازی داده های تله متری بهتر است. ما به طور گسترده اثر اندازه های دسته ای، تعداد حالات در MM و درجه چند جمله ای در PR را ارزیابی کردیم. مشاهده کردیم که یک اندازه دسته بزرگتر داده ها را کاهش می دهد، اما دقت بازسازی کمتر است. بنابراین، مشخص کردیم که یک اندازه دسته ای بین 16 تا 64 برای کاهش داده با دقت بهتر بازسازی استفاده شود مناسب است. ارزیابی تجربی ما نشان می دهد که روش مبتنی بر PR به دلیل دقت بالای ضرایب، فضای ذخیره سازی بیشتری را در مقایسه با روش مبتنی بر مدل زنجیره مارکوف انجام می دهد. ما همچنین مشاهده کردیم که فضای ذخیره سازی 95/33 درصد و استفاده از پهنای باند 75 درصد با استفاده از دقت 92 درصد کاهش می یابد. راه حل پیشنهادی برای کار بعدی، ما بر روی

شناسایی تطبیقی اندازه دسته و تعداد حالات در MM متمرکز هستیم تا بیشتر فضای خود را کاهش داده و دقت بازسازی را افزایش دهیم. ما همچنین قصد داریم از یک MM در هر متریک در سطح مرکز داده برای تکرار بارهای کاری که دارای الزامات مصرف منابع مشابه هستند استفاده کنیم. علاوه بر این، ما قصد داریم مدل‌های ترکیبی برای کاهش داده‌ها و بازسازی داده‌های از راه دور را بررسی کنیم.

- [1] D. A. Reed, J. Dongarra, “Exascale computing and big data,” *Commun. ACM*, vol. 58, no. 7, pp. 56–68, 2015.
- [2] A. Labrinidis and H. V. Jagadish, “Challenges and opportunities with big data,” in *Proc. VLDB Endow.*, vol. 5, no. 12, 2012, pp. 2032–2033.
- [3] W. Iqbal, M. N. Dailey, and D. Carrera, “Unsupervised learning of dynamic resource provisioning policies for cloud-hosted multitier web applications,” *IEEE Syst. J.*, vol. 10, no. 4, pp. 1435–1446, Dec. 2016.
- [4] F. Seracini, X. Zhang, T. Rosing, and I. Krüger, “A proactive customeraware resource allocation approach for data centers,” in *Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl.*, 2014, pp. 26–33.
- [5] R. Birke, L. Y. Chen, and E. Smirni, “Usage patterns in multi-tenant data centers: A temporal perspective,” in *Proc. 9th Int. Conf. Autonomic Comput.*, 2012, pp. 161–166.
- [6] M. Kutare, G. Eisenhauer, C. Wang, K. Schwan, V. Talwar, and M. Wolf, “Monalytics: Online monitoring and analytics for managing large scale data centers,” in *Proc. 7th Int. Conf. Autonomic Comput.*, 2010, pp. 141–150.
- [7] L. Zhan, T. Z. Fu, D. M. Chiu, and Z. Lei, “A framework for monitoring and measuring a large-scale distributed system in real time,” in *Proc. 5th ACM Workshop HotPlanet*, New York, NY, USA, 2013, pp. 21–26.
- [8] M. S. Karabinoglou and T. Gözel, “Load forecasting modelling of data centers and IT systems by using artificial neural networks,” in *Proc. 10th Int. Conf. Elect. Electron. Eng.*, 2017, pp. 62–66.
- [9] L. Duan, D. Zhan, and J. Hohnlerlein, “Optimizing cloud data center energy efficiency via dynamic prediction of CPU idle intervals,” in *Proc. IEEE 8th Int. Conf. Cloud Comput.*, 2015, pp. 985–988.
- [10] J. Xue, R. Birke, L. Y. Chen, and E. Smirni, “Spatial-temporal prediction models for active ticket managing in data centers,” *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 1, pp. 39–52, Mar. 2018.
- [11] F.-H. Tseng, X. Wang, L.-D. Chou, H.-C. Chao, and V. C. Leung, “Dynamic resource prediction and allocation for cloud data center using the multiobjective genetic algorithm,” *IEEE Syst. J.*, vol. 12, no. 2, pp. 1688–1699, Jun. 2018.
- [12] M. Duggan, K. Mason, J. Duggan, E. Howley, and E. Barrett, “Predicting host cpu utilization in cloud computing using recurrent neural networks,” in *Proc. 12th Int. Conf. Internet Technol. Secured Trans.*, 2017, pp. 67–72.

- [13] O. Rolik, E. Zharikov, A. Koval, and S. Telenyk, "Dynamic management of data center resources using reinforcement learning," in *Proc. 14th Int. Conf. Adv. Trends Radioelectronics, Telecommun. Comput. Eng.*, 2018, pp. 237–244.
- [14] S. Di and F. Cappello, "Fast error-bounded lossy HPC data compression with SZ," in *Proc. IEEE Int. Parallel Distrib. Process. Symp.*, 2016, pp. 730–739.
- [15] M. Burtscher and P. Ratanaworabhan, "FPC: A high-speed compressor for double-precision floating-point data," *IEEE Trans. Comput.*, vol. 58, no. 1, pp. 18–31, Jan. 2009.
- [16] J. Iverson, C. Kamath, and G. Karypis, "Fast and effective lossy compression algorithms for scientific datasets," in *Proc. Eur. Conf. Parallel Process.*, 2012, pp. 843–856.
- [17] K. Sayood, *Introduction to Data Compression*. Burlington, Massachusetts, San Francisco, CA: Morgan Kaufmann, 2017.
- [18] S. Karlin, *A First Course in Stochastic Processes*. Cambridge, Massachusetts: Academic Press, 2014.
- [19] C. Carvalho, L. Leal, M. Lemos, and R. Holanda, "Avoiding data traffic on smart grid communication system," in *Proc. Int. Electron. Conf. Sensors Appl.*, 2014, pp. 1–16.
- [20] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [21] P. Senin, "Dynamic time warping algorithm review," Information and Computer Science Department, University of Hawaii at Manoa, Honolulu, USA, Tech. Rep. 855, 2008.
- [22] M. Müller, *Information Retrieval for Music and Motion*, vol. 2. Salmon Tower Building, New York: Springer, 2007.
- [23] J. Gu and X. Jin, "A simple approximation for dynamic time warping search in large time series database," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.*, 2006, pp. 841–848.
- [24] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. KDDWorkshop*, vol. 10, Seattle, WA, USA, 1994, pp. 359–370.
- [25] T. Rakthanmanon *et al.*, "Searching and mining trillions of time series subsequences under dynamic time warping," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2012, pp. 262–270.
- [26] T. Oates, L. Firoiu, and P. R. Cohen, "Clustering time series with hidden Markov models and dynamic time warping," in *Proc. IJCAI Workshop on Neural, Symbolic and Reinforcement Learn. Methods for Sequence Learn.*, 1999, pp. 17–21.

- [27] M. Kim, "Time-series dimensionality reduction via Granger causality," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 611–614, Oct. 2012.
- [28] D. Littau and D. Boley, "Streaming data reduction using low-memory factored representations," *Inf. Sci.*, vol. 176, no. 14, pp. 2016–2041, 2006.
- [29] G. Gawde and J. Pawar, "Shape based time series reduction using PCA," in *Proc. Int. Conf. Innov. Inf., Embedded Commun. Syst.*, 2017, pp. 1–4.
- [30] T. Sun, H. Sun, and W. Chen, "Dimensionality reduction for interval time series," in *Proc. World Congr. Inf. Commun. Technol.*, 2012, pp. 1115–1120.
- [31] J. Wang, S. Yue, X. Yu, and Y. Wang, "An efficient data reduction method and its application to cluster analysis," *Neurocomputing*, vol. 238, pp. 234–244, 2017.
- [32] S. Tripathi and S. De, "An efficient data characterization and reduction scheme for smart metering infrastructure," *IEEE Trans. Ind. Inf.*, vol. 14, no. 10, pp. 4300–4308, Oct. 2018.
- [33] Z. C. Dagdia, C. Zarges, G. Beck, and M. Lebbah, "A distributed rough set theory based algorithm for an efficient big data pre-processing under the spark framework," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, 2017, pp. 911–916.
- [34] J. Yan *et al.*, "Effective and efficient dimensionality reduction for largescale and streaming data preprocessing," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 3, pp. 320–333, Mar. 2006.
- [35] M. Z. A. Bhuiyan, G. Wang, W. Tian, M. A. Rahman, and J. Wu, "Contentcentric event-insensitive big data reduction in Internet of Things," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.
- [36] K. Wu, D. Lee, A. Sim, and J. Choi, "Statistical data reduction for streaming data," in *Proc. Scientific Data Summit*, New York, NY, USA, 2017, pp. 1–6.
- [37] A. Egri, I. Horváth, F. Kovcs, R. Molontay, and K. Varga, "Crosscorrelation based clustering and dimension reduction of multivariate time series," in *Proc. IEEE 21st Int. Conf. Intell. Eng. Syst.*, 2017, pp. 000241–000246.
- [38] X. Peng and B. Pernici, "Correlation-model-based reduction of monitoring data in data centers," in *Proc. 5th Int. Conf. Smart Cities Green ICT Syst.*, 2016, pp. 1–11.
- [39] T. Yu, X. Wang, and A. Shami, "A novel fog computing enabled temporal data reduction scheme in IoT systems," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–5.
- [40] M. F. Mohamed, M. El-Gayyar, A. E.-R. Shabayek, and H. Nassar, "Data reduction in a cloud-based AMI framework with service-replication," *Comput. Elect. Eng.*, vol. 69, pp. 212–223, 2018.
- [41] K. Goodhope *et al.*, "Building LinkedIn's real-time activity data pipeline," *IEEE Data Eng. Bull.*, vol. 35, no. 2, pp. 33–45, 2012.

- [42] J. Kreps *et al.*, “Kafka: Adistributed messaging system for log processing,” in *Proc. NetDB*, 2011, pp. 1–7.
- [43] L. Liu, O. Hasan, and S. Tahar, “Formal reasoning about finite-state discrete-time Markov chains in HOL,” *J. Comput. Sci. Technol.*, vol. 28, no. 2, pp. 217–231, 2013.
- [44] K. S. Trivedi and A. Bobbio, *Reliability and Availability Engineering: Modeling, Analysis, and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2017.
- [45] J. P. D. C. Shuja-ur Rehman Baig and Marcelo Amaral, “Performance characterization of spark workloads on shared NUMA systems,” in *Proc. IEEE 4th Int. Conf. Big Data Comput. Service Appl.*, 2018, pp. 41–48.
- [46] Spark-bench. [Online]. Available: <https://github.com/SparkTC/sparkbench>
- [47] One and two-tailed tests. [Online]. Available: https://en.wikipedia.org/wiki/One_and_two-tailed_tests
- [48] R. A. Fisher, *Statistical Methods for Research Workers*. Guildford, U.K: Genesis Publishing Pvt Ltd, 1925.
- [49] D. Terpstra, H. Jagode, H. You, and J. Dongarra, “Collecting performance data with PAPI-C,” in *Tools for High Performance Computing 2009*. Springer, 2010, pp. 157–173.