



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

پیشگویی طوفان در یک ابر

چکیده

پیشگویی رفتار آینده به طور قابل اتکا و کارآمد برای سیستم هایی مهم است که خدمات مجازی را مدیریت می کنند. چنین سیستم هایی باید قادر به متعادلسازی بارهای درون یک محیط ابری برای اطمینان از این امر باشد که موافقت نامه ها در سطح خدماتی با هزینه قابل قبولی برطرف شود. در اصل پیشگویی های صحیح می تواند با استخراج از انواع منابع داده ها حاصل آید که شرح رفتار تاریخی خدمات، شروط برنامه هایی که روی آنها اجرا شده است، و تقاضاهای پدید آمده روی ابر توسط کاربران نهایی می باشد. پیشگویی صحیح بارهای ماکزیمم که احتمالاً در کوتاه مدت مشاهده گردیده است، از اهمیت خاصی برخوردار است. اما رهیافتهای استاندارد برای مدلسازی رفتار سیستم، با تحلیل تمامیت داده های مشاهده شده تمایل به پیشگویی رفتار سیستمی متوسط به جای استثنایی و نادیده گرفتن الگوهای مهم تغییر طی زمان دارد. در این مقاله، ما به مطالعه توانایی یک رگرسیون خطی چندمتغیره ساده برای پیش بینی استفاده از CPU پیک (طوفانها) در یک محیط ابری صنعتی می پردازیم. نیز چندین اصلاح را در رگرسیون خطی استاندارد برای تنظیم آن برای پیشگویی طوفان مطرح می داریم.

کلیدواژه های نمایه سازی- رگرسیون، سری های زمانی، پیشگویی، ابر

1-مقدمه

فرا ساختار به شکل یک خدمات یا IaaS یک هنجار در سیستم های IT در مقیاس بزرگ شده است و مجازی سازی در این محیطها متداول است. یکی از مشکلات اصلی چنین مجازی سازی هایی قرار دادن ماشین های مجازی یا VM و متعادل سازی بار است. اگر تقاضاها روی فرا ساختارها از ظرفیت آنها خارج باشد، شکست رخ خواهد داد، زمانهای پاسخ افزایش خواهد یافت، و رضایت مشتری سقوط خواهد کرد. از اینرو ضروری است که مطمئن شویم قرار دادن و متعادلسازی به طور مناسبی صورت گرفته است.

متعادل سازی مناسب و برنامه ریزی ظرفیت مناسب در چنین محیطهای ابری نیاز به پیشگویی بار کاری آینده و مصرف های منبع دارد. بدون پیشگویی های خوب، مدیران ابر مجبور به پیکربندی بیش از حد ذخایر منابع خود

برای دستیابی به قابلیت دسترسی لازم برای احترام به توافقات سطح خدمات یا SLA می شوند. این کار گران بوده و می تواند باز در برآورده سازی SLA ها به طور مداوم شکست بخورد. بی توجه به پیش بینی های خوب ، مدیران ابر تمایل به کار در یک حالت واکنشی دارند و می توانند ناکارآمد و حتی از هم گسیخته شوند.

چندین تکنیک پیش بینی بار کاری براساس انالیز سری های زمانی طی سالیان ارائه شده است که می تواند در محیطهای ابری نیز بکار بسته شود. خط پایینی چنین متونی این است که هیچ تکنیک خوبی برای پیش بینی وجود ندارد. بسته به ماهیت داده ها و مشخصات خدمات و بار کاری، تکنیک های آماری مختلف و الگوریتم های یادگیری ماشینی می تواند بهتر از سایرین عمل کند. در برخی موارد حتی ساده ترین تکنیک ها مانند رگرسیون خطی می تواند بهتر از رقبای پیچیده تر عمل کند.

برای درک قابلیت کاربرد چنین تکنیک های پیشگویی درباره مسائل در مقیاس صنعتی، ما یک سری مطالعات موردی را تنظیم کرده ایم که در آن تکنیک های پیشگویی مختلفی را روی داده هایی که از همکار صنعتی ما CA Technologies آمده است ، بکار می بندیم. CA Technologies یک فراهم کننده ابر برای سازمانهای مقیاس کلان متعددی است. اینها IaaS را برای مشتریان خود فراهم می کنند و بر استفاده ایشان نظارت دارند. سیستم مدیریت ابر اساسا مسئول متعادل سازی بار کاری با قراردادان ماشین های مجازی روی فراساختار فیزیکی است.

در این مقاله، ما تجربه خود را روی بکارگیری یک تکنیک رگرسیون خطی چندمتغیره یا MVLr برای پیشگویی استفاده از CPU ی ماشین های مجازی در زمینه یکی از مشتریان CA گزارش می دهیم. اما برخلاف بسیاری از تکنیک های پیشگویی موجود ، که در آن ایشان خطاهای پیشگویی متوسط را به حداقل رسانده یا احتمالات متوسط را به حداکثر رسانده اند، ما به پیشگویی موارد فوق العاده به جای متوسط علاقمندتریم. انگیزه از نوع بار کاری ناشی می شود که ما در مطالعه موردی خود با آن روبرویم که برای سایر کاربردهای مبتنی بر ابر نیز خیلی نامتداول نیست . در مورد ما، متوسط استفاده بین کلیه VMها حداکثر 20 درصد بوده است ولی ماکزیمم استفاده تقریبا خیلی نزدیک به 100 درصد به طور تغییرناپذیر بوده است. بکارگیری MVLr در چنین داده هایی (بیشتر اوقات خیلی استفاده کمی دارد ولی گاهی اوقات به اوج می رسد)، ما درک کرده ایم که هرچند

متوسط پیشگویی ها خیلی صحیح است ولی پیشگویی برای مقادیر بزرگ (طوفانها) به طور برجسته ای ضعیف است.

برای سازگاری با این مسئله، ما چندین اصلاحات را در MVLRL اصلی برای تنظیم آن برای پیشگویی مقادیر پیک ارائه کرده ایم. نتایج نشان می دهد که کسر موسمی بودن های استخراجی توسط تغییر شکل Fourier و بعد استفاده از متوسط گیری MVLRL بهترین نتایج مشاهده شده را برای پیشگویی طوفان فراهم می کند. در بخشهای ذیل، ما به شرح جزئیات هر MVLRL اصلاح شده و گزارش نتایج آن می پردازیم.

2-موضوع مطالعه

ما با یک مجموعه اساسی از داده های عملکرد راجع به یک محیط محاسبه کننده ابر بزرگ منفرد که یک تعداد زیادی خدمات مجازی را طی یک دوره شش ماهه اجرا کرده است، روبرو شدیم. در کل، 2133 مدخل مستقل بود که عملکرد آن هر شش دقیقه را کسب می کرد. این ها شامل 1572 ماشین مجازی و 495 ماشین فیزیکی بوده است. ماشین های فیزیکی فراهم کننده حمایت برای 56 میزبان MV نرم افزاری بوده است. به طور متوسط 53 درصد خدمات نظارت شده در هر زمانی فعال بود و ماکزیمم آن 85 درصد بوده است. داده های کسب شده به طور ایده آل شرح بارهای کاری CPU، استفاده از حافظه، I/O دیسک و ترافیک شبکه را میدهد. اما در بیشتر موارد تنها بارهای کاری CPU در دسترس بود. از اینرو تنها متمرکز بر داده های بار کاری CPU بوده ایم. این داده ها با ارقام عملکرد ساعتی متوسط و ماکزیمم ترکیب و یکپارچه سازی شده است.

از لحاظ ماهیت خدمات، دست کم 423 خدمات اختصاص به فراهم سازی محیط های دسکتاپ مجازی داشته درحالیکه ابر نیز حمایتی را برای خدمات مبتنی بر وب، پردازش معامله، پشتیبانی پایگاه داده ها، قراردادن خدمات مجازی بر میزبانان، و سایر خدمات مانند نظارت و پشتیبانی بر عملکرد ثابت کرده است.

همانگونه که در مورد محیطهای دسکتاپی معمولاً مصداق دارد، استفاده از کامپیوتر فردی به طور برجسته ای متفاوت است. بیشتر اوقات کار اندک شدیدی اگر باشد روی دسکتاپ مجازی در حال انجام است و خدمات مجازی به نظر تقریباً بی ارزش می باشد. اما برای هر دسکتاپ مجازی دوره هایی از فعالیت شدید وجود دارد زمانی که CPU، حافظه، I/O دیسک، و یا پیک های ترافیک شبکه وجود دارند. مشابهاً، درون محیطهای

پردازش معاملاتی، انواع وسیعی از رفتارها بسته به تقاضای کلی قرار داده شده روی چنین سیستم هایی وجود خواهد داشت.

همانگونه که اشاره گردید، توزیع فراوانی مصرف ها به شدت چولگی داشته و اکثریت وسیعی از مصرف ها (83.5%) از 25 درصد تجاوز نکرده است. از اینرو اغلب نیاز به یک تکنیک پیشگویی داریم که (با یک درجه منطقی از اطمینان) نشان می دهد که چه زمانی بارهای آتی بالا خواهد بود حتی اگر چنین پیشگویی هایی به طور ریاضی متناسب با تمامیت داده های مشاهده شده و آتی به همان دقت سایر رهیافتهای اماری نباشد.

3- پیشگویی طوفان با استفاده از رگرسیون خطی

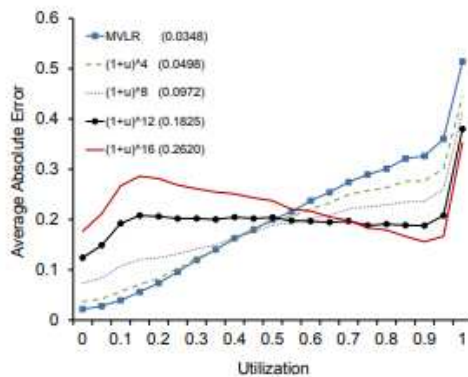
در این بخش، ما یک MVLRL اصلی و سه نوع آن را در پایگاه داده صنعتی مان بکار بسته و گزارش صحت آنها را از لحاظ متوسط خطاهای مطلق هنگام پیشگویی مقادیر پیک داده ایم.

MVLRL: برای بکارگیری یک MVLRL روی داده های استفاده از CPU، ما ابتدا ترسیم های همبستگی از داده های ارائه شده با محاسبه همبستگی خودکار هر سری زمانی با هر نسخه تاخیری از همان سری های زمانی بدست می آوریم. این امر نشان دهنده قوی ترین همبستگی خودکار در سطح ساعتی (1676 منبع)، هفتگی (247)، روزانه (106)، دو هفته یکبار (41) می باشد که این همبستگی ها تنها به کندی طی فواصل زمانی طولانی تری تنزل می یابد.

با استفاده از تاخیرات مهم کشف شده، رگرسیون خطی چندمتغیره آنگاه با استفاده از 10 تاخیر یک و دو ساعته، یک و دو روزه، 1 و 2 و 3 و 4 هفتگی و 1 و 2 ماهه برای شناسایی ضرایب همبستگی بکار بسته شد که هنگام بکارگیری در این داده های تاخیری با همبستگی قوی، تناسب خطی داده هایی را با حداقل خطای باقیمانده مربع مشاهده کرد. این امر قابلیت پیشگویی خوب کلی را در بین منابع داده ها فراهم ساخته است. معادله خطی منتهج آنگاه برای پیشگویی استفاده از ساعت بعدی بکار گرفته شد.

برای توانایی ارزیابی تکنیک های پیشگویی از لحاظ مقادیر پیک، برای هر سری های داده، استفاده های مشاهده شده به فواصل اندک در جمع های 0.05 تقسیم بندی می گردد. برای هر یک از چنین بخشهایی، متوسط تفاوت مطلق بین مقادیر مشاهده شده و پیشگویی شده بدست می آید. رسم این خطاهای مطلق متوسط به ازای

هر فاصله زمانی به درک رفتار الگوریتم پیشگویی کننده برای طیف های داده های ورودی مختلف کمک می کند.



تصویر 1- مقایسه MVLRL با رگرسیون میانگین گیری شده. پارامتر میانگین گیری شده برابر $c=4,8,12,16$ می باشد که به معنی یک نقطه داده است که دارای کاربرد u بوده و کلیه تاخیرات مرتبط به آن در $(1+u)^c$ ضرب می شود. مقادیر داخل پرانتز خطاهای مطلق متوسط بین کلیه فواصل می باشد.

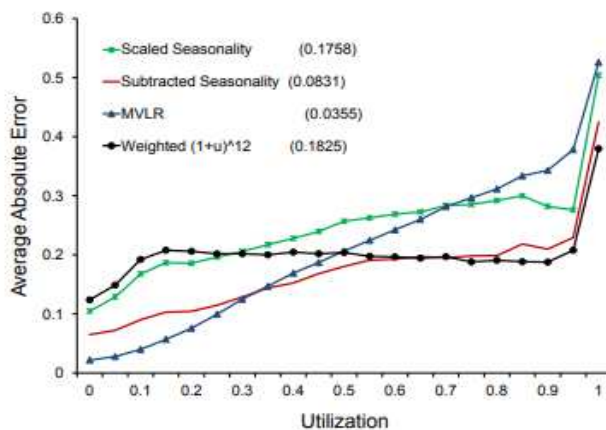
یک مسئله حداقل که با آن مواجه شدیم و نیاز به ملاحظه خاصی دارد مقادیر از دست رفته است. در این مطالعه، خلاهای کوتاه با مقادیر قبلی آنها تخمین زده شد. اما در پایگاه های ما، 769 سری زمانی دارای داده های از دست رفته بیشتری نسبت به داده های واقعی است. در چنین سناریوهایی ما منبع به شدت از دست رفته داده ها را از پایگاه داده خود دور انداختیم چون در غیراینصورت ممکن است نتایج تجربی ما را دچار چولگی آماری کند.

MVLRL میانگین گیری شده: برای تنظیم MVLRL به مقادیر بالاتر، ما ابتدا رگرسیون را به یک پنجره لغزنده 5 هفته ای محدود می کنیم. درون جمع بندی های رگرسیون، نگاه هر نقطه داده را میانگین گیری می کنیم. چون توزیع کلی استفاده ها مشاهده شده است که نمایی است، ما میانگین گیری نمایی را بکار می بندیم که در آن یک نقطه داده که دارای کاربرد u می باشد و کلیه تاخیرات مرتبط با این داده در $(1+u)^c$ ضرب شده بود. این امر به طور طبیعی استفاده های بالاتر از یک میانگین بزرگتر معنی دار را تعیین می کند با اینحساب پیشگویی هایی را به سمت مقادیر بالاتر چولگی آماری داده درحالیکه همزمان آنها را با بالاترین مقادیر محدود می سازد. همانگونه که در تصویر 1 دیده شد، افزایش c (پارامتر میانگین گیری شده)، از 4 به 16 باعث کاهش

خطاهای پیشگویی برای فواصل استفاده بالاتر حین افزایشات خطاها برای فواصل پایین تر می شود. از اینرو یک خطای مطلق متوسط ثابت تر بین کلیه فواصل می تواند بهترین انتخاب باشد. برای مثال $c=12$ به نظر یک انتخاب خوب برای پایگاه داده های ما می باشد.

در هر دو MVLRL و میانگین گیری شده، تنها متکی به تاخیرات پیش تعریف شده برای پیشگویی های خودمان بودیم. اما فرد باید نقشهای فصلی را نیز در نظر بگیرد. بکارگیری تغییر شکل های Fourier یک رهیافت معمول برای کشف الگوهای چرخه ای مشهود درون داده ها می باشد. دو رهیافت بعدی تغییر شکل های Fourier را بکار گرفته است.

موسمی بودن مقیاس دار: با بکارگیری تغییر شکل Fourier روی پایگاه داده های ما، ما جمع بندی n تای اول امواج سینوسی با بزرگترین دامنه را (اصطلاحی که بارزترین متغیرپذیری را درون داده های ورودی شرح می دهد) به داده های ورودی متناسب ساخته ایم. این کار به خوبی با موسمی بودن کلی را درون داده های فراهم شده متناسب است، ولی در متناسب سازی پیک ها در داده ها شکست می خورد. برای توجیه اصطلاحاتی که در پیشگویی مان گنجانده نشده است، منطقی است که تلاش کنیم تا تغییر شکل Fourier را برای تناسب بهتر استفاده مقیاس بندی کنیم.



تصویر 2-مقایسه رهیافتهای مبتنی بر Fourier (موسمی بودن مقیاس بندی شده و کسر شده) با MVLRL و

MVLRL میانگین گیری شده

یک راه برای تناسب بهتر پیک ها بکارگیری یک تغییر شکل خطی با تغییر شکل محاسبه شده Fourier با نادیده انگاری مقادیر زیر یک مقطع مناسب (برای مثال ماکزیمم برابر با 0.05) می باشد. ما برای حداقل جهت

بدون تغییر باقی ماندن با کسر آن ترتیب بندی می کنیم، و انگاه با میانگین مقادیر مشاهده شده تقسیم بر میانگین مقادیر پیشگویی شده مقیاس بندی می کنیم قبل از اینکه می نیمم را دوباره به آن اضافه کنیم.

تصویر 2 مقایسه کننده MVLr ، MVLr میانگین گیری شده و رهیافتهای موسمی بودن مقیاس بندی شده است. MVLr فراهم کننده صحت پیشگویی کننده بهتر برای استفاده های پایین، و MVLr میانگین گیری شده برای کاربردهای بالا می باشد. موسمی بودن مقیاس بندی شده بین MVLr و MVLr میانگین گیری شده در هر دو مورد است ولی نیز احتمال پیشگویی های بلندمدتتر را به ما دارد.

برای بهبود صحت پیشگویی های ما، در رهیافت بعدی، ما آنالیزهای Fourier و رگرسیون را با هم ترکیب کرده ایم.

موسمی بودن کسر شده: در این رهیافت، ما موسمی بودن را از داده های اصلی برای حذف بیشتر متغیرپذیری در داده ها کسر می کنیم که آنرا خطی تر می سازد و با این حساب یک تناسب بهتری با مدل های پیشگویی خطی می شود. اساسا ما (1) موسمی بودن فصلی را از کاربردهای مشاهده شده کسر می کنیم (2) MVLr را روی باقیمانده منتج اجرا می کنیم (مانند قبل) و (3) موسمی بودن را دوباره به پیشگویی منتج اضافه می کنیم.

نتایج (تصویر 2) بدست آمده به طور معنی داری بهتر از استفاده از تغییرشکلهای Fourier یا MVLr یا MVLr میانگین گیری شده به تنهایی است. بکارگیری این رهیافت به پایگاه داده هایمان، به تقریب، باعث کاهش خطای مطلق متوسط بین کلیه ورودی ها برای استفاده های بزرگ توسط یک سوم می شود و متوسط کلی خطای مطلق را نصف می کند.

مهمترین نقص استفاده از تغییرشکلهای Fourier این است که برخلاف رگرسیون، که به سرعت می تواند ارائه پیشگویی ها را از نتایج مشاهده شده اولیه آغاز کند، یک مقدار اساسی از داده های قبلی باید موجود باشد تا موسمی بودن را درون یک سری زمانی ورودی کشف کند. در عمل، مطرح شده است که پیشگویی های اولیه روی خود رگرسیون پیش بینی شده است درحالی که به طور دوره ای چون داده های کافی در دسترس شده است، یک تغییر شکل Fourier سریع برای کشف مکرر موسمی بودن با داده های ورودی بکار گرفته شده است.

4- محدودیت ها و تهدیدها برای روایی

سه محدودیت اول این مطالعه که در حال حاضر داریم رویش کار می کنیم عبارتند از (1) داشتن یک پیشگویی تک بعدی مبتنی بر استفاده از CPU (2) مطالعه تنها یک رهیافت پیشگویی رگرسیون خطی (و نسخ اصلاحی اش) و (3) ارزیابی پیش بینی تنها براساس صحت پیشگویی و نه بهبود نهایی از لحاظ اثرات روی مجازی سازی و پروسه برنامه ریزی ظرفیت.

همانگونه که در تحقیقات صنعتی متداول است، مطالعه به داده هایی محدود است که برای گروه تحقیقاتی موجود می باشد. مشهود است که داشتن دانش درباره سایر اندازه گیری های عملکردی مانند حافظه ، I/O دیسک، و مصرف های ترافیک شبکه احتمالا به بهبود قدرت پیش گویی منجر می شود. بعلاوه، دانش درباره نوع بار کاری و حتی زمینه کسب و کار پشت بار کاری در میان متغیرهایی است که می تواند بر مصرف CPU آتی اثر داشته باشد. اما در این مطالعه ما تنها به داده های استفاده از CPU از سیستم های مشتری CA دسترسی داشتیم. از اینرو هدف به حداکثرسانی صحت پیشگویی (بخصوص از لحاظ مقادیر پیک) با استفاده از داده های موجود می باشد. اما در آینده ما مشغول برنامه ریزی برای دسترسی به چندین منبع داده های عملکرد و گسترش رهیافت تک بعدی مان به چنین پایگاه داده های غنی می باشیم.

درحالیکه رگرسیون خطی چندمتغیره می تواند بنا به انتظار به طور مناسبی به روندهای در حال تغییر پاسخ دهد، پیش فرض ما (که طبق مطالعه داده ها پیشگویی شده است) آن بود که هیچ روندی درون موسمی بودن طولانی مدت وجود ندارد. اگر روندها درون موسمی بودن مشاهده شده وجود داشتند، لازم است که تلاش داریم تا موسمی بودن را با استفاده از کاری پیچیده تر از یک معادله خطی ساده مقیاس بندی کنیم. رهیافتهای رگرسیون غیرخطی در میان اولین تکنیک هایی است که ما مشغول برنامه ریزی برای بکارگیری روی پایگاه های داده کنونی و اتی مان هستیم. بعلاوه، تکنیک های یادگیری ماشینی، برای مثال شبکه های عصبی نیاز به ارزیابی برای یافتن بهترین رهیافت پیشگویی دارد.

در این مرحله از مطالعه مشکل است که یافته های تحقیقاتی را روی مجازی سازی کامپیوتر و پروسه برنامه ریزی ظرفیت بکار بندیم. اما در کوتاه مدت ما مشغول برنامه ریزی برای بررسی پایگاه های داده و تکنیک های بیشتری و افزایش شواهد حمایت کننده حول ایده های پیش بینی طوفان می باشیم به نحوی شرکت خواهان بکارگیری آنها در پروسه مجازی اش شود.

از لحاظ روایی سازه، ما بهترین تلاش را برای تامین داده های از دست رفته نمودیم ولی فرضیات درباره اینکه مقادیر از دست رفته چه می تواند بوده باشد، الزاما الگوریتم پیشگویی کننده را به مخاطره می اندازد. بعلاوه، از لحاظ روایی خارجی، این تحقیق روی یک داده مشتری منفرد طی یک فاصله زمانی کوتاه مدت مقایسه ای شش ماهه پیشگویی گردید. حتی با گنجاندن مقدار زیادی خدمات فیزیکی و مجازی، رفتار سیستم و الگوهای داده ها نمی تواند درون کلیه محیطهای محاسبه شده ابری معمول و متداول باشد.

5- کار مرتبط

در کل، متون مرتبط با این کار می تواند در سه دسته بگنجد: (1) مشخصات بار کاری (2) پیش بینی بار کاری و (3) تکنیک های پیشگویی. اولین طبقه بیشتر متمرکز بر خصوصیات بار کاری است که می تواند به تحلیل و احتمالا پیشگویی آن کمک کند. دومین طبقه به بررسی داده های مختلف و تکنیک های پیشگویی برای پیشگویی بار کاری اتی میپردازد ولی باز تمرکزش بیشتر روی بررسی داده ها نسبت به خود پیشگویی است.

ولی در این مقاله تمرکز ما بیشتر روی سمت پیشگویی یعنی طبقه بندی سوم است. ما از داده هایی استفاده می کنیم که برای ما همکار صنعتی مان را قابل دسترس ساخته و روی احتمالاتی برای به حداکثرسانی صحت پیشگویی ها مطالعه می کنیم. از اینرو ما به اختصار به برخی مقالات مرتبط در این جهت اشاره می کنیم.

تکنیک های رگرسیون خطی در میان مشهورترین رهیافتهای پیشگویی بار کاری است. برای مثال Andreolini و همکارانش استفاده از متوسطهای در حال تغییر را برای روان سازی سری های زمانی ورودی و بعد استفاده از بهره برداری خطی را روی دو تا از مقادیر روان سازی شده برای پیشگویی بارکاری آتی مطرح ساخته اند.

روان سازی نمایی، مدل های رگرسیون خودکار و ARIMA پراستفاده ترین رهیافتهای دیگر در این حیطه است. برای نمونه Dinda و همکارانش توانایی انواع مدل های ARIMA مانند را برای پیشگویی آینده مقایسه کردند. Nathuji و همکارانش ارزیابی ماشین های مجازی را در جداسازی پیشگویی کرده و بعد رفتار آنها را هنگامی که با هم کار می کنند با استفاده از سیگنالهای ورودی متعدد برای تولید خروجی های پیشگویی کننده متعدد با استفاده از معادلات تفاوت (روان سازی نمایی) پیشگویی کردند.

استفاده از تکنیک های یادگیری ماشینی برای پیشگویی بار کاری دسته بزرگ دیگری را از متون مرتبط تشکیل می دهد. برای نمونه Istin و همکارانش از شبکه های عصبی برای پیشگویی بار کاری استفاده کردند و Khan و

همکارانش مدل‌های مارکوف پنهان را برای کشف همبستگی‌ها بین بارهای کاری بکار بستند که می‌تواند انگه برای پیشگویی تنوعات در الگوهای بار کاری استفاده گردد.

برخلاف کار موجود، مقاله ما از تکنیک‌های پایه (رگرسیون خطی و نسخه اصلاحی آن در ترکیب با تغییر شکل Fourier) به شکل یک نقطه شروع کننده استفاده کرده است و آنها را به داده‌های استفاده از یک مشتری تکنولوژی CA با یک هدف خاص پیشگویی استفاده‌های پیک بکار بسته است.

6- نتیجه‌گیری‌ها و کاراتی

استفاده از سیستم می‌تواند هم به شکل نتیجه ملاحظات موسمی بودن منظم و هم به شکل نتیجه انواع بدشکلی‌ها که به طور ذاتی پیشگویی سختی دارند، به اوج برسد. روشن نیست که شیوه بهینه پیشگویی چنین فعالیت سیستم اوجی از طریق رهیافتهایی مانند رگرسیون خطی چندمتغیره هم باشد یا خیر، چون چنین پیشگویی‌هایی طبق تمامیت داده‌های مشاهده شده می‌باشد و تمایل دارد تا نتایج روانتری را نسبت به نتایجی که بر احتمال استفاده از سیستم خارج از ظرفیت تاکید می‌کند، ایجاد کند.

ما یک تعداد اصلاحات را در رگرسیون خطی چندمتغیره استاندارد و در تغییر شکل‌های Fourier ارائه کرده ایم که به طور منفرد و احتمالاً به طور جمعی توانایی رگرسیون خطی چندمتغیره را برای پیشگویی استفاده‌های پیک با خطای مطلق متوسط کوچک منطقی بهبود می‌دهد.

بهترین تغییر شکل مطرح شده یک موسمی بودن مقیاس بندی شده را کسر می‌کند که با آنالیز Fourier از استفاده‌های مشاهده شده استخراج شده است. و بعد یک رگرسیون خطی چندمتغیره میانگین‌گیری شده را روی باقیمانده‌های منتج اجرا می‌کند و دست‌آخر موسمی بودن را دوباره به پیشگویی‌های منتج اضافه می‌کند.

در آینده، ما برنامه ریزی می‌کنیم تا این مطالعه را با استفاده از متغیرهای پیشگویی کننده تری مانند حافظه، I/O دیسک، و مصرف‌های ترافیک شبکه و مشخصات بار کاری و داده‌های کسب و کار بسط دهیم. بعلاوه برنامه ریزی می‌کنیم تا به ارزیابی چندین تکنیک پیشگویی مانند رگرسیون غیرخطی و تکنیک‌های یادگیری ماشینی برای بهبود صحت پیشگویی طوفان پردازیم.

این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی