دوفصلنامه محاسبات و سامانه‌های توزیع شده
سال هفتم، شماره ۲، شماره پیاپی ۱۴، صفحه ۵۶ تا ۷۰، سال ۱۴۰۳

تحلیل و بررسی تخصیص وظایف به منابع ماشینهای مجازی برای ایجاد تعادل بار در رایانش ابری مبتنی بر طبقه بندهای مختلف یادگیری ماشین

امیر حسین نظری افشار[1] و حمیدرضا دوست[2]

۱- فارغ‌التحصیل کارشناسی ارشد دانشگاه آزاد تهران واحد یادگار امام خمینی(ره) شهری

۲- دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات گرایش تجارت الکترونیک از دانشگاه آزاد واحد علوم و تحقیقات

چکیده

تعادل بار یک جنبه مهم در زمینه رایانش ابری برای افزایش توزیع بار کاری و استفاده کارآمد از منابع است که در عوض زمان پاسخ کلی سیستم را کاهش می‌دهد. رویکردها و الگوریتم‌های زیادی برای حل مسائل مربوط به تعادل بار مانند: زمان‌بندی وظایف، مهاجرت، استفاده از منابع و غیره پیشنهاد شده‌اند. در این پژوهش چندین رویکرد مرتبط با چالش حیاتی در رایانش ابری که متعادل‌سازی بار است، ارائه کرد. مشکلات مربوط به تعادل بار از طریق تحلیل مقایسه‌ای الگوریتم‌های پیشنهادی توسط محققان در شش سال گذشته مورد بحث قرار گرفت. چندین رویکرد پیشنهاد شد، با این حال، هنوز برخی از مسائل در محیط ابری وجود دارد، مانند مهاجرت ماشین‌های مجازی، مسائل مربوط به تحمل خطا که هنوز به طور کامل برطرف نشده است. در این پژوهش یک مدل بهبود طبقه‌بندی ماشین‌های مجازی برای ایجاد تعادل بار کاری در برنامه‌های رایانش ابری با استفاده از شبکه‌های عصبی عمیق و الگوریتم تکاملی علف هرز ارائه می شود. در واقع هدف اصلی پژوهش استفاده بهینه از منابع محاسباتی در محیط ابری است که الگوریتم‌های زمان بندی نقش مهمی در فرآیند بهینه سازی دارند. در واقع موضوع پیشنهادی در شاخه یادگیری ماشین و طبقه‌بندی ماشین‌های مجازی برای ایجاد تعادل باردر رایانش ابر است. در شبیه سازی روش پیشنهادی نیز در ارزیابی مقایسه روش پیشنهادی و روش پایه از نظر صحت تخصیص ها روش پیشنهادی به نسبت روش پایه مقدار ۶,۶ درصد بهبود از خود به ثبت رسانده است.

واژه‌های کلیدی: بهبود طبقه‌بندی ماشین‌های مجازی، تعادل بار کاری، رایانش ابری، شبکه‌های عصبی عمیق و الگوریتم تکاملی علف هرز

تاریخچه مقاله:

تاریخ ارسال: ۱۴۰۳/۰۷/۰۲۶

تاریخ اصلاحات: ۱۴۰۳/۱۱/۲۹

تاریخ پذیرش: ۱۴۰۳/۱۲/۱۶

تاریخ انتشار: ۱۴۰۳/۱۲/۲۵

ایمیل نویسنده مسئول:
Root.afshar@gmail.com

۱- مقدمه

امروزه مساله تعادل بار در محاسبات ابری به یکی از مهم ترین چالش ها در محیط ابر برای کاهش زمان پاسخ و افزایش رضایت کاربران ابر تبدیل شده است. از گذشته تا به امروز روش‌های مختلفی برای توازن بار ارائه شده‌اند و مقالات بسیاری در این زمینه وجود دارد. هر یک، الگوریتم‌های خود را با در نظر گرفتن معیارهایی ایجاد کرده‌اند. با توجه به جایگاه مفهوم توازن بار در محاسبات ابری و روند رو به رشد استفاده از سرویس‌های این فناوری در بسترهای مختلف، که در بیان مسئله به آن پرداخته شد. رفته رفته بر اهمیت فرآیند توازن بار افزوده شده است؛ زیرا فرآیند توازن بار در واقع قدم اصلی در استفاده از سرویس‌های ابری است و لذا حصول نتایج دقیق‌تر در این فرآیند و نیاز به بهینه‌سازی هرچه بیشتر آن، به یکی از مهم‌ترین چالش‌های فرا روی این حوزه در چند سال اخیر بدل گشته و موجب شده که تکنولوژی‌ها و روش-های مربوطه، بیشتر مورد توجه هر دو بخش دانشگاهی و تجاری قرار گرفته و تحقیقات فراوانی در هر دو بخش، بر این زمینه معطوف گردد. ارائه دهندگان ابر از فناوری مجازی‌سازی بهره‌برداری می‌کنند و به مشتریان خود منابع محاسباتی را در قالب ماشین‌های مجازی عرضه می کنند. از سوی دیگر، ارائه دهندگان خدمات از این ماشین‌های مجازی برای ارائه خدمات در سطح اپلیکیشن به کاربران بهره می‌برند .برای تخصیص وظایف کاربران به ماشین‌های مجازی، کاهش زمان پاسخگویی، ارائه کیفیت امیدوارکننده خدمات و استفاده

حداکثری از منابع، ارائه دهندگان خدمات از تکنیک های زمان بندی وظایف بهره برداری می‌کنند. بنابراین، الگوریتم زمان‌بندی کار یکی از عناصر اصلی هر زیرساخت ابری است. از این رو در این پژوهش نیز مدلی جهت بهبود طبقه‌بندی ماشین‌های مجازی برای ایجاد تعادل بار کاری در برنامه‌های رایانش ابری با استفاده از شبکه‌های عصبی عمیق و الگوریتم تکاملی علف هرز ارائه می‌گردد.

۲- کارهای مرتبط

یک الگوریتم زمان‌بندی کار مبتنی بر الگوریتم ازدحام ذرات با استفاده از تکنیک متعادل‌سازی بار برای محیط محاسبات ابری بهبود یافته ارائه شد. نویسندگان معتقدند که تامین منابع بر اساس تقاضای دینامیک یکی از اهداف اولیه فرآیند زمان‌بندی وظایف محاسبات ابری است. زمان‌بندی کار، یک مشکل زمان چند جمله‌ای غیر قطعی-سخت است و مسئول تخصیص وظایف به ماشین‌های مجازی به‌گونه‌ای است که استفاده از منابع و عملکرد را افزایش می‌دهد، زمان پاسخ را کاهش می‌دهد و کل سیستم را متعادل نگه می‌دارد. در این مقاله، یک روش زمان‌بندی کار ایستا بر اساس الگوریتم بهینه‌سازی ازدحام ذرات ارائه می‌شود که در آن وظایف غیر پیشگیرانه و مستقل فرض می‌شوند. این مقاله عملکرد روش اصلی بر الگوریتم ازدحام ذرات را با استفاده از تکنیک تعادل بار بهبود می‌بخشد. آنها روش پیشنهادی خود را با زمان‌بندی کار دور روبین[1]، زمان‌بندی وظایف بر الگوریتم ازدحام ذرات بهبود یافته و تکنیک متعادل‌سازی بار مقایسه کردند. نتایج شبیه‌سازی نشان می‌دهد که این روش با افزایش استفاده از منابع ۲۲ درصدی و کاهش ۳۳ درصدی طول عمر در مقایسه با الگوریتم پایه بر الگوریتم ازدحام ذرات از این الگوریتم‌ها بهتر عمل می‌کند. نتایج نشان می‌دهد که روش پیشنهادی آنها سریع‌تر از الگوریتم اصلی بر الگوریتم ازدحام ذرات به جواب تقریباً بهینه همگرا می‌شود و با وظایف بیشتر کارآمدتر است [۱].

تحقیقی تحت عنوان "بهینه‌سازی زمان‌بندی وظایف در رایانش ابری بر اساس الگوریتم ابتکاری" ارائه دادند. آنها معتقدند که رایانش ابری یک فناوری نوظهور است و به کاربران این امکان را می‌دهد تا بر اساس نیاز خود پرداخت کنند و عملکرد بالایی دارد. رایانش ابری نیز یک سیستم ناهمگن است و حجم زیادی از داده‌های کاربردی را در خود نگه می‌دارد. در فرآیند زمان‌بندی برخی داده‌های فشرده یا محاسبه یک برنامه کاربردی فشرده، اذعان می‌شود که بهینه‌سازی زمان انتقال و پردازش برای یک برنامه کاربردی بسیار مهم است. در این مقاله به منظور به حداقل رساندن هزینه پردازش، یک مدل برای زمان‌بندی کار فرموله شد و یک الگوریتم بهینه‌سازی ازدحام ذرات پیشنهاد شد که بر اساس قانون ارزش موقعیت کوچک است. به موجب مقایسه الگوریتم بر الگوریتم ازدحام ذرات با الگوریتم بر الگوریتم ازدحام ذرات تعبیه شده در متقاطع و جهش و در تحقیقات محلی، نتایج آزمایش نشان می‌دهد که الگوریتم ازدحام ذرات نه تنها سریع‌تر همگرا می‌شود، بلکه سریع‌تر از دو الگوریتم دیگر در مقیاس بزرگ اجرا می‌شود. نتایج آزمایش ثابت می-کند که الگوریتم ازدحام ذرات برای محاسبات ابری مناسب‌تر است [۲].

تحقیقی برای زمان‌بندی کارآمد با استفاده از موازنه بار در رایانش ابری ارائه شد که نویسندگان معتقدند که زمان‌بندی گردش کار یک زمینه چالش برانگیز در محاسبات است که در آن وظایف بر اساس نیاز کاربر برنامه‌ریزی می‌شوند و به دلیل کیفیت تقاضای خدمات توسط کاربر پرهزینه می‌شوند. محیط ابری برای این کار به کار گرفته شده است تا هزینه کلی را کاهش دهد. برای حفظ و استفاده از منابع در مکانیزم زمان‌بندی رایانش ابری مورد نیاز است. الگوریتم‌ها و پروتکل‌های زیادی برای مدیریت کارهای موازی و منابعی که برای بهبود عملکرد واحد پردازش مرکزی در محیط ابری استفاده می‌شوند. بهینه سازی ازدحام ذرات و بهینه سازی گرگ خاکستری برای زمان بندی موثر استفاده می‌شود. این کار بر اساس بهینه‌سازی زمان اجرای کل و هزینه کل اجرا است. نتایج روش پیشنهادی در مقایسه با روش‌های موجود مؤثر است. بهینه‌سازی ازدحام ذرات با استفاده از توزیع پارتو مقداردهی اولیه می شود. TET و TEC هزینه و زمان به حداقل رسیده را با استفاده از الگوریتم گرگ خاکستری برای همگرایی تصمیم ماشین مجازی نشان دادند. بنابراین نتیجه نشان می‌دهد که الگوریتم گرگ خاکستری در مقایسه با الگوریتم خفاش موجود بهتر عمل می‌کند [۳].

[1] round robin

استراتژی برای زمان‌بندی ماشین مجازی بر اساس الگوریتم‌های یادگیری ماشین برای متعادل‌سازی بار ارائه شد. نویسندگان این مقاله معتقد میباشند که با افزایش سریع دسترسی کاربران، تعادل بار در مرکز داده ابری به عامل مهمی در پایداری خوشه تبدیل شده است. از نقطه نظر زمان‌بندی سبز، این مقاله یک استراتژی زمان‌بندی هوشمند ماشین مجازی مبتنی بر الگوریتم یادگیری ماشین برای دستیابی به تعادل بار مرکز داده ابری را پیشنهاد می‌کند. در مرحله اول، یک الگوریتم پیش‌بینی بار مبتنی بر الگوریتم ژنتیک[۲]، الگوریتم خوشه‌بندی کامینز[۳] بر اساس حداقل یا حداکثر بهینه و الگوریتم تکامل تفاضلی تطبیقی برای افزایش توانایی جستجوی محلی برای حل مشکل عدم تعادل بار در داده‌های ابری پیشنهاد شده‌اند. مرکز نتایج تجربی نشان داد که در مقایسه با سایر الگوریتم‌های کلاسیک، استراتژی زمان‌بندی ماشین مجازی پیشنهادی تعداد مهاجرت ماشین مجازی را تا ۹۴,۵ درصد و مصرف انرژی مرکز داده ابری را تا ۴۹,۱۳ درصد کاهش می‌دهد [۴].

تحقیقی تحت عنوان "یک رویکرد بهبود یافته برای تعادل بار در میان ماشین‌های مجازی در محیط ابری" ارائه دادند. آنها معتقدند که رایانش ابری تکامل قابل توجهی در فناوری اطلاعات نشان داده است. کاربران تنها در صورت دسترسی به اینترنت می‌توانند از خدمات مختلف فناوری ابری بهره‌مند شوند. در رایانش ابری، توازن بار به عنوان یک موضوع اساسی مطرح است که محققان این حوزه را با آن مواجه کرده است. تعادل بار اساساً با تخصیص کار منصفانه و کارآمد در بین منابع محاسباتی کار می‌کند که در نهایت رضایت کاربر را بالا می‌برد و بهره‌وری سیستم‌ها را افزایش می‌دهد. بسیاری از تکنیک‌های متعادل کننده بار تلاش‌هایی را برای حل این مشکل با استفاده از الگوریتم فراابتکاری و تقویت عملکرد و کارایی سیستم ها انجام دادند. در این مقاله، تکنیک های متعادل کننده بار موجود، مورد بحث قرار گرفته است و راجع به شکاف های تحقیقاتی مختلفی بحث شده است. همچنین تکنیک جدید[۴] الگوریتم بهینه سازی گرگ خاکستری کنجکاو ژنتیک با استفاده از ترکیب الگوریتم بهینه سازی گرگ خاکستری و الگوریتم ژنتیک پیشنهاد شده است [۵].

رویکردی مبتنی بر طبقه‌بندی ماشین مجازی برای افزایش تعادل بار کاری برای برنامه‌های رایانش ابری ارائه شد و نویسندگان معتقدند که علیرغم تحقیقات زیادی که در زمینه رایانش ابری انجام شده است، همچنان با مسائل و چالش‌هایی مانند تعادل بار مواجه است که همچنان نیازمند متدولوژی‌ها و مدل‌های بهینه‌سازی بیشتری برای بهبود عملکرد و دستیابی به رضایت بالای کاربران است. در این مقاله، تکنیک یادگیری ماشین، که طبقه‌بندی است، برای ساخت گروه‌هایی از ماشین‌های مجازی بر اساس استفاده از واحد پردازش مرکزی و حافظه و همچنین برای طبقه‌بندی مشاغل/وظایف کاربر به گروه‌های مختلف بر اساس اندازه آنها و بر اساس اطلاعات فایل‌های گزارش استفاده می‌شود. این رویکرد ماشین‌های مجازی را در گروه‌ها مرتب می‌کند و چندین کار منابع ماشین مجازی یکسان را به اشتراک می‌گذارند. هدف این مقاله اجازه دادن به منابع پویاتر و بهبود الزامات سرویس کیفیت خدمات با به حداکثر رساندن استفاده از منابع و رضایت کاربر، مانند افزایش استفاده از منابع و کاهش تعداد رد کارها است [۶].

الگوریتم‌هایی برای زمان‌بندی متعادل‌سازی بار برای آزمایشگاه‌های محاسبات مجازی در سرویس‌های رایانش ابری مبتنی بر آزمایشگاه‌های محاسبات مجازی در برنامه‌ریزی ارائه شد و این کار پژوهشی حاضر به مطالعه آزمایشگاه‌های محاسبات مجازی در برنامه‌ریزی در یک محیط محاسبات ابری می‌باشد. این موضوع از یک مطالعه واقعی ناشی می‌شود که در آن دانشگاه الکترونیک عربستان نیاز به برنامه‌ریزی کارآمد آزمایشگاه‌ها در جلسات از پیش تعریف‌شده برای یکی از کالج‌هایش در تمام شعبه‌هایش را بیان می‌کند. مشکل مورد بررسی، از یک سو، در متعادل کردن بار تخصیص آزمایشگاه‌ها به جلسات با توجه به تعداد شرکت کنندگان، و از سوی دیگر، در تعیین انواع و تعداد میزبان‌های مورد نیاز برای اطمینان از بهترین وظایف میزبان جلسات است. در حالی که تعادل بار با اندازه‌گیری فاصله بین حداکثر و حداقل تعداد بار ماشین‌های مجازی، حداکثر تعداد بار ماشین‌های مجازی و واریانس تعداد بار ماشین‌های مجازی ارزیابی می‌شود، کارایی انتخاب میزبان و تخصیص‌ها بر حسب عدم تعادل ظرفیت باقیمانده هر میزبان مورد استفاده ارزیابی می‌شوند. در این مقاله مدل‌های ریاضی خطی‌سازی‌شده را برای مسائل فرعی و کل مسئله با

[2] SVR_GA
[3] k-means
[4] IG-GWO

توابع مختلف هدف پیشنهاد می‌شود. آنها ابتدا یک رویکرد بهینه‌سازی دو مرحله‌ای را پیشنهاد می‌کنند که در آن مدل‌های ریاضی از طریق سیپلکس[5] حل می‌شوند. سپس به روش ترکیبی می‌پردازند که کل مسئله را به دو مشکل فرعی خود تقسیم می کند. در مرحله بعد، مدل ترکیبی را روی سیپلکس اجرا و دو روش اکتشافی را پیشنهاد می کنند تا یک راه حل خوب عملی در زمان معقول داشته باشند. پس از آن، نتایج به‌دست‌آمده را برای اندازه‌های متنوعی از نمونه‌هایی که با توجه به ویژگی‌های دانشگاه تولید شده، مقایسه می‌کنند. بنابراین ثابت شده است که اکتشافات پیشنهادی یک راه حل قانع کننده برای شکست نرم‌افزار قدرتمند سیپلکس در حل مشکل برای موارد در مقیاس بزرگ است. علاوه بر این، نتایج به‌دست‌آمده نشان می‌دهد که آنها از نتایجی که در حال حاضر در دانشگاه استفاده می‌شوند بهتر عمل می‌کنند [۷].

روشی برای موازنه بار در محیط‌های ابری: زمان بندی وظایف بهینه از طریق الگوریتم ترکیبی" ارائه شد نویسندگان مقاله معتقدند که در رایانش ابری، بسیاری از چالش‌ها مانند خرابی سرور، از دست دادن محرمانگی، حجم کاری نامناسب و غیره همچنان کارایی سیستم‌های ابری را در سناریوهای دنیای واقعی محدود می‌کنند. به همین دلیل کارهای تحقیقاتی زیادی برای رفع نقص سیستم‌های موجود در حال انجام است. در میان آنها، به نظر می‌رسد تعادل بار حیاتی‌ترین مسئله‌ای است که عملکرد بخش ابر را بدتر می-کند، و از این رو تعادل بار بهینه با زمان‌بندی وظایف بهینه را ضروری می‌کند. با هدف دستیابی به تعادل بار بهینه با استقرار کار موثر، این مقاله قصد دارد یک مدل متعادل کننده بار پیشرفته را با کمک به دست آمده از الگوریتم‌های فراابتکاری توسعه دهد. معمولاً رسیدگی به وظایف در سیستم ابری یک مشکل-NP سخت است و علاوه بر این، وظایف مستقل غیر پیشگیرانه در رایانش ابری بسیار مهم هستند. این مقاله با در نظر گرفتن سه هدف اصلی به ترتیب حداقل زمان، اولویت و متعادل سازی بار، یک مدل متعادل کننده بار بهینه جدید را معرفی می‌کند. علاوه بر این، یک تابع تک هدفه جدید نیز تعریف شده است که هر سه هدف ذکر شده در بالا را در بر می‌گیرد. علاوه بر این، استقرار وظایف باید بهینه باشد و برای این کار یک الگوریتم بهینه‌سازی ترکیبی جدید به نام WOA اصرار بر جنبش شب تاب[6] معرفی شده است. این الگوریتم ادغام مفهومی الگوریتم استاندارد بهینه‌سازی نهنگ و الگوریتم فایرفلای است. در نهایت، عملکرد مدل پیشنهادی با مدل‌های معمولی با هدف اثبات کارایی آن از نظر زمان ساخت، تکمیل کار (اولویت)، و درجه عدم تعادل نیز مقایسه می‌شود [۸].

تحقیقی برای بهبود الگوریتم زمانبندی وظایف بر اساس روش نامزد متعادل سازی بار در محیط محاسبات ابری" ارائه دادند. آنها معتقدند که در سال‌های اخیر، محاسبات ابری به روشی ایده‌آل برای ارائه انواع مختلف برنامه ها تبدیل شده است. می‌تواند پردازش، ذخیره‌سازی، خدمات وب، پلتفرم‌های توسعه و پایگاه‌های داده را فراهم کند تا شرکت‌ها را قادر سازد تا زیرساخت‌ها و انواع خدمات خود را به سرعت مقیاس‌بندی کنند. با افزایش تقاضا برای قدرت پردازش بالاتر، حجم زیادی از داده ها باعث ایجاد بار قابل توجهی بر روی محیط محاسبات ابری می‌شود. بنابراین، یک الگوریتم زمان‌بندی کارآمد برای وظایف تحت محیط محاسبات ابری به یک مسئله حیاتی تبدیل شده است. برخی از الگوریتم‌های معروف[7] برای بهبود عملکرد زمان‌بندی پیشنهاد شده‌اند. با این حال، الگوریتم‌های فوق مشکل تعادل بار گره های کار را در نظر نمی‌گیرند. این مشکل تعادل بار، عملکرد کلی سیستم شبکه را محدود می‌کند. از این رو، الگوریتم زمانبندی BCSV[8] برای بهبود مشکلات فوق در این مقاله پیشنهاد شده است. ایده اصلی این الگوریتم استفاده از کوچکترین ارزش رای[9]، بزرگترین ارزش حق رای[10] و معیار ارزش رای[11] به عنوان عامل زمانبندی برای افزایش عملکرد ارسال کار است. با توجه به نتایج آزمایش، الگوریتم BCSV پیشنهادی می‌تواند تعادل بار و طول عمر بهتری نسبت به الگوریتم‌های موجود تحت محیط شبکه ناهمگن HiHi، HiLo، LoHi و LoLo بدست آورد. به عبارت دیگر، الگوریتم BCSV پیشنهادی می‌تواند نتایج بهتری را در زمان‌بندی کار در حالی که مشکل تعادل بار را در نظر می‌گیرد، دریافت کند [۹].

[5] CPLEX

[6] Firefly Movement insisted WOA(FM-WOA)
[7] Suffrage ، MaxSuffrage و AMS
[8] Balanced Candidate Suffrage Value
[9] Smallest Suffrage Value
[10] Largest Suffrage Value
[11] Criteria Suffrage Value

۳- تحلیل روش پیشنهادی

رایانش ابری فناوری برجسته‌ای است که خدمات (خصوصی و عمومی) را ارائه می‌دهد، به عنوان مثال، دسترسی آسان به داده‌ها، برنامه‌ها و فایل‌ها از طریق اینترنت (ابر)، خدمات ذخیره‌سازی مقیاس‌پذیر به صورت آنلاین به جای فایل‌های ذخیره شده محلی در ماشین‌های کاربران مانند رایانه یا تلفن. در حالت استقرار هم‌مکانی هیبریدی آفلاین، معماری‌های جدید مانند کانتینرها و ریزسرویس‌ها، روند جهانی و برتری زیادی را در برنامه‌های رایانش ابری مانند داده‌های بزرگ و هوش مصنوعی نشان می‌دهند، اما پیچیدگی زیادی در مدیریت و بهینه‌سازی به همراه دارند. برای پرداختن به مشکل عدم تطابق عرضه و تقاضای منابع در زمان‌بندی ابری، در این روش، گزارش‌های محاسبات ابری به توالی‌های مکانی-زمانی نمودار تبدیل می‌شود، یک مدل پیش‌بینی عملکرد گردش کار پیشنهاد می‌شود و یک شبکه عصبی یادگیری عمیق سرتاسر جدید ارائه می‌گردد. این تحقیق راه جدیدی برای تسهیل زمان‌بندی گردش کار با پیش‌بینی‌های مبتنی بر داده ارائه می‌کند و به طور سیستماتیک وضعیت تحقیق رایانش ابری را بررسی، چالش‌هایی را برای زمان‌بندی جریان‌های کاری در محیط ابری نسل جدید، و فرصت جدید برای رویکرد پیش‌بینی مبتنی بر داده را بررسی می‌کند.

در حال حاضر، زمان‌بندی کار و تخصیص منابع، موضوعات تحقیقاتی داغ مربوط به بسیاری از مشکلات تحقیقاتی رایانش ابری است. در پلت فرم رایانش ابری، مشکل اصلی زمان‌بندی وظایف و تخصیص معقول درخواست‌های ارسالی توسط کاربر به گره‌های رایانش ابری برای پردازش است. زمان‌بندی کار در رایانش ابری یک مسئله سخت چند جمله‌ای غیر قطعی است که مستقیماً با مشکلات عملی مانند استفاده از منابع، رضایت کاربر و هزینه اجرا مرتبط است. با توجه به نیازهای مختلف کاربر، الگوریتم‌های زمان‌بندی کار بر اهداف بهینه‌سازی مختلف، از جمله زمان، تعادل بار، رضایت کاربر، اصول اقتصادی و غیره تمرکز می‌کنند.

به منظور بهبود کیفیت خدمات کاربران، دستیابی به تعادل بار بین ماشین‌های مجازی را تحت محدودیت‌های قرارداد سطح سرویس در نظر می‌گیریم و نرخ رد وظایف را کاهش می‌دهیم. تعادل بار ماشین های مجازی می تواند سرعت پاسخگویی وظایف پردازشی را بهبود بخشد. نرخ رد وظیفه به نسبت وظایفی که قرارداد سطح سرویس را در کل وظایف برآورده نمی کنند، اشاره دارد. به طور کلی، هر چه سرعت پاسخگویی ماشین مجازی سریعتر باشد و نرخ رد کار کمتر باشد، می تواند کیفیت خدمات کاربر را بهبود بخشد. قرارداد سطح سرویس یک پروتکل بین کاربران و ارائه دهندگان خدمات ابری است که در مورد حفظ حریم خصوصی، امنیت و در دسترس بودن و سایر عوامل مذاکره می کند. از آنجایی که روش پیشنهادی بر الزامات کاربر برای زمان تکمیل کار تمرکز دارد، قرارداد سطح سرویس فقط مهلت کار را در نظر می گیرد و فرض می کند که پلت فرم سیستم ایمن، قابل اعتماد و خصوصی است. اگر قرارداد سطح سرویس نقض شود، کار توسط منابع محاسباتی رد می شود، مهم است که نرخ رد کار را کاهش دهید. هرچه میزان رد کار کمتر باشد، تجربه کاربری بهتری خواهد داشت. برای بهبود ظرفیت پردازش و تحمل خطا در ماشین‌ها، هر ماشین مجازی باید تا حد امکان وظایف را به صورت موازی انجام دهد، به جای اینکه وظایف را روی یک یا چند ماشین مجازی متمرکز کند. به این روش تعادل بار ماشین های مجازی می گویند.

در این تحقیق، یک الگوریتم زمان‌بندی وظایف چندهدفه پویا مبتنی بر یادگیری عمیق و الگوریتم تکاملی علف هرز ارائه شده است که از روش شبکه عمیق Q برای انتخاب یک ماشین مجازی مناسب برای محیط فعلی استفاده می‌شود. این الگوریتم، مبتنی بر یادگیری تقویتی عمیق و الگوریتم تکاملی علف هرز ، می‌تواند مسئله زمان‌بندی وظایف را در محیط محاسبات ابری حل کند تا بار بین ماشین‌های مجازی تحت محدودیت‌های قرارداد سطح سرویس را متعادل کند. یادگیری تقویتی یک روش مهم یادگیری ماشین مبتنی بر فرآیند بهبود با الگوریتم تکاملی برای حل مسائل کنترل خودکار است. فضای حالت توسعه‌یافته که برای حل مسائل گسسته‌سازی مناسب است، حل مسائل با ابعاد بالا و یافتن راه‌حل مناسب را آسان‌تر می‌کند. به منظور حل مسئله انفجار ابعاد، روشی ترکیبی از یادگیری تقویتی و شبکه عصبی عمیق پیشنهاد شده است. در این کار، هنگامی که کاربر درخواستی را به پلتفرم ابری ارسال می‌کند، از مدل پیشنهادی برای انتخاب یک ماشین مجازی مناسب برای اجرای وظیفه تحت شرایط بار فعلی ماشین‌های مجازی استفاده میشود و پس از

انتخاب یک ماشین مجازی برای وظیفه فعلی، تعیین می‌شود که آیا وظیفه می‌تواند قبل از مهلت مقرر تکمیل شود (در صورت مطابقت با موافقت‌نامه سطح خدمات) اگر موافقت‌نامه سطح خدمات رعایت نشود، وظیفه رد خواهد شد. در غیر این صورت، وظیفه دریافت‌کننده روی ماشین مجازی انتخاب‌شده اجرا خواهد شد .الگوریتم پیشنهادی زمانی که تا حد امکان محدودیت‌های موافقت‌نامه سطح خدمات را رعایت کند، در بهینه‌سازی تعادل بار عملکرد خوبی دارد .از لحاظ تئوری، می‌توان آن را در بسیاری از حالت‌های محاسبات توزیع‌شده، از جمله محاسبات ابری، محاسبات لبه‌ای، محاسبات ابری لبه‌ای و غیره اعمال کرد.

۴- نتایج و شبیه سازی

رایانش ابری یک مدل قوی است که به کاربران و سازمان‌ها اجازه می‌دهد خدمات مورد نیاز خود را بر اساس نیاز خود خریداری کنند. این مدل خدمات بسیاری مانند ذخیره‌سازی، پلتفرم‌های استقرار، دسترسی راحت به سرویس‌های وب و غیره را ارائه می‌دهد. تعادل بار یک مشکل رایج در فضای ابری است که حفظ عملکرد برنامه‌های کاربردی مجاور اندازه‌گیری کیفیت خدمات و پیروی از سند توافقنامه سطح سرویس را که از سوی ارائه‌دهندگان ابری به شرکت‌ها نیاز است، دشوار می‌کند. ارائه دهندگان ابر برای توزیع بار کاری برابر بین سرورها تلاش می‌کنند. یک تکنیک تعادل بار کارآمد باید با استفاده بهینه از منابع ماشین‌های مجازی، رضایت بالای کاربران را بهینه و تضمین کند. در این بخش این روش شبیه‌سازی و ارزیابی می‌شود.

در این پژوهش وظایف و همچنین اطلاعات "گراف جهت‌دار غیرمدور" در مجموعه داده گردش کار مبتنی بر نمودار جمع‌آوری میشود و تعریف برچسب‌ها را بر اساس خوشه‌بندی ارائه می‌گردد. در نهایت، یک مدل پیش‌بینی توالی مبتنی بر نمودار ایجاد می‌شود، که الگویی برای مدل‌سازی مشکل پیش‌بینی‌های جریان کار تحت سناریوی مدیریت منابع کانتینری ارائه می‌دهد. این روش پیشنهادی همچنین برای مسائل پیش‌بینی کلی فرآیند صنعتی مناسب است.

گام اول روش پیشنهادی خوشه بندی مبتنی بر الگوریتم تکاملی علف هرز: در این مرحله، وظایف گروه‌بندی می‌شوند و سرخوشه‌ها با استفاده از تکنیک الگوریتم تکاملی علف هرز انتخاب می‌شوند. این روش یک رویکرد تکاملی(مبتنی بر جمعیت است که شامل حذف رقابتی، مقداردهی اولیه، تولید مثل و پراکندگی فضایی می‌شود) است که توسط رشد علف‌های هرز هدایت می‌شود. که ابتدا ایجاد جمعیت اولیه به طور قابل توجهی بر عملکرد مدل تأثیر می‌گذارد. در گردش کار مبتنی بر گراف جهت‌دار غیرمدور[12]، ترتیب وظایف در راه‌حل از جمعیت باید محدودیت اولویت را برآورده کند. به منظور توزیع کار روی راه‌حل، از یک تکنیک توزیع ارتفاع استفاده می‌شود .در ابتدا، حداقل ارتفاع h(ti)برای همه وظایف ارزیابی شده و حداکثر ارتفاع g(ti) در گراف جهت‌دار غیرمدور به شکل زیر بدست می‌آید:

$$h(t_i) = \{pred(t_i) = \emptyset\ 1 + \left(h(t_j)\right) pred(t_i) \neq \emptyset \quad (1)$$

$$g(t_{exit}) = \{h(t_i) succ(t_i) \neq \emptyset\ \left(h(t_j)\right) - 1\ succ(t_i) \neq \emptyset \quad (۲)$$

که h (ti) و g(ti) حداقل و حداکثر مقدار و $pred$ مقدار پیش فرض و در نظر گرفته شده است که به مقدار نهایی ختم می شود. همچنین $succ$ در صورت موفقیت فقط مقدار خواهد گرفت.

جمعیت اولیه با توجه به مقدار ارتفاع ایجاد می شود. در حالی که ارتفاع کار با دقت انتخاب شده در راه حل ها نباید از جانشینان و اولویت های آن فراتر رود و کمتر باشد. در نسل، ترکیب و حاصل جمعیت دور قبلی بر اساس بدترین تناسب و راه حل های انتخاب شده، بذر تولید می کند.مقدار بذر برای تولید بذر از حداقل بذر (به عنوان تابع هدف) به صورت زیر بدست می‌آید [۱۰]:

$$\text{seeds}_i = \frac{S_{max} - S_{min}}{f_{max} - f_{min}} \times [f_{our} - f_{min}) + S_{min} \quad (۳)$$

بطوریکه S نشان دهنده seed بوده و حد پایین و بالا فضای مورد نظر برای مقدار نهایی را در نظر می گیرد و f مقدار تابع استفاده شده است که باز حد پایین و بالا و استفاده شده در آن را در نظر خواهد گرفت.

در مرحله بعد، از فاز پراکندگی فضایی برای تولید فرزندان استفاده می‌شود. برای تمایز یک فرزند از بهینه‌سازی سراسری و والد آن، از یک روش فاصله کد بزرگتر برای انتشار فرزندان به یک فضای راه‌حل جداگانه استفاده می‌شود. در مقابل، از فاصله کد کوتاه‌تر برای بهینه‌سازی محلی استفاده

[12] Directed acyclic graph

می‌شود. رویکرد افزایش فاصله کد با هدف حفظ تنوع جمعیت است. برای دستیابی به این هدف، یک گره به طور تصادفی به عنوان گره عملیاتی در یک راه‌حل انتخاب می‌شود. سپس، وظایف بین جانشینان و پیشینیان آنها پیدا می‌شوند تا محدودیت تقدم تضمین شود. بنابراین، ترجمه در فواصل زمانی مشخص بین وظایف انجام می‌شود. علاوه بر این، یک جدول برچسب برای ثبت وضعیت وظیفه ترجمه ایجاد می‌شود.

برای به حداقل رساندن میزان انرژی مورد نیاز برای انتقال داده، الگوریتم تکاملی علف هرز پیشنهاد شده است. هدف از بهینه‌سازی، انتخاب بهترین گروه ممکن از گره‌ها برای خدمت به عنوان رهبران شبکه است، در نتیجه مصرف کلی انرژی به حداقل رسیده و طول عمر شبکه به حداکثر می‌رسد. تابع هدف طول عمر شبکه (δ_n^n) را گسترش می دهد که به محض از بین رفتن گره اولیه به پایان می رسد: برای تعیین طول عمر هر گره برای اجرای وظایفش، کل انرژی مورد نیاز گره به عنوان طول عمر کل اندازه‌گیری می‌شود. با جمع تمام انرژی‌های مورد نیاز، طول عمر کل گره به دست می‌آید. این طول عمر، معیار انتخاب اجرا در هر گره است.

$$\delta_n^1 = \delta_s \qquad (۴)$$

در اینجا δs نشان دهنده طول عمر گره s و S نشان دهنده مجموعه ای از گره ها است. جایی که n نشان دهنده حسگرهایی است که به طور مساوی توزیع شده اند، k نشان دهنده خوشه ها است. از این رو، n/k گره برای هر خوشه (یک CH و (n/k)-1 عضو خوشه) وجود دارد.

گام دوم ترتیب اجرا وظایف مبتنی بر تکنیک شبکه عصبی عمیق: شکل ۱ یک نمونه از وابستگر بین وظایف است که براین اساس باید زنجیره مربوط به گردش کار مشخص و از وظیفه اول به آخر اولویت بندی شود و در روش پیشنهادی از همین روش برای استخراج گردش کار و سپس اولویت آن ها استفاده می‌شود. رویکردهای موجود اهمیت پیش‌بینی را برای زمان‌بندی پیشنهاد کرده‌اند و مرحله پیش‌بینی را در الگوریتم‌های زمان‌بندی ادغام کرده‌اند. با این حال، روش‌ها محدود به مجموعه‌های آزمایشی بوده‌اند، مدل‌های پیش‌بینی خاص برای گردش کار به‌طور سیستماتیک مورد مطالعه قرار نگرفته‌اند و در زمان‌بندی اعمال نشده‌اند. پیش‌بینی‌های ابری قبلی عمدتاً بر روی یک VM/Host/Cluster/Task متمرکز می‌شوند و بیشتر آنها به‌عنوان سری‌های زمانی مدل‌سازی می‌شوند و از (t-n….t-1) برای پیش‌بینی t استفاده می‌کنند. آنها نمی توانند به طور مؤثر از روابط وابستگی وظایف استفاده کنند و هر وظیفه از یکدیگر جدا است. گردش کار به یک پارادایم برای توضیح وظایف در فضای ابری تبدیل شده است، اما هنوز پیش‌بینی‌های عملکرد کمی با جریان‌های کاری ابری روبرو هستند. دو مسیر پیش‌بینی را برای گردش‌های کاری ابری پیشنهاد کرد، یکی شبیه‌سازی و اندازه‌گیری مبتنی بر محیط آزمایش، و دیگری یادگیری ماشین مبتنی بر داده‌محور است. آنها در مورد ضرورت و امکان‌سنجی زمان‌بندی مبتنی بر گردش کار در ابر، به‌ویژه روش مبتنی بر داده توضیح می‌دهند. با این حال، هیچ اشاره ای به نحوه استفاده از ویژگی های بافتی و ساختاری، و مقایسه ای برای روش های پیش بینی مبتنی بر یادگیری عمیق وجود ندارد.

این مدل برای بازیابی دقت با حذف وزن‌های بزرگتر تنظیم شده است. کاهش ابعاد شبکه یک روش موثر برای کاهش پیچیدگی شبکه است و حل مشکل بیش از حد برازش است. دو نوع هرس به طور گسترده مورد استفاده قرار می گیرد: هرس وزنی و هرس نورونی. در نوع هرس وزنی، وزن‌های فردی در ماتریس وزن صفر تنظیم می‌شوند. راه دیگر هرس واحد/نرون است که در آن کل ستون‌های ماتریس وزن به صفر تغییر می‌کنند، بنابراین نورون خروجی مربوطه حذف می‌شود [۱۱]. گزارش‌ها و تعریف گردش کار می‌تواند عملکرد تاریخی، تخصیص منابع و رابطه خدمات را منعکس کند. همانطور که پیشنهاد شده، پیش‌بینی‌های عملکرد نسبت به گردش‌های کاری ابری کمیاب است. یکی از دلایل اصلی عدم وجود مجموعه داده است. مجموعه داده های خوشه عمومی مرتبط منتشر شده توسط فروشندگان رایانش ابری بررسی شد.

الگوریتم‌های مبتنی بر یادگیری ماشین حتی اکنون می‌توانند به طرز ماهرانه‌ای بار کاری وظایف را پخش کنند. به طور خاص، معماری ترانسفورماتور و مکانیسم توجه به خود، پتانسیل برتر را برای زمینه های مختلف نشان می‌دهد. یکی از این موارد استفاده از یادگیری عمیق در پیش بینی های وظایف ابری را نشان داده اند، اما کاربرد Transformer کمیاب است. برای این منظور، در این پژوهش از روش‌های یادگیری

عمیق و معماری‌های مبتنی بر ترانسفورماتور را برای بررسی پیش‌بینی‌های عملکرد گردش کار الهام گرفته شده است.

همچنین بسیاری از الگوریتم‌های یادگیری ماشین مبتنی بر ترانسفورماتور برای مناطق خاص، مانند پردازش زبان طبیعی و بینایی کامپیوتر طراحی شده‌اند. با این حال، گردش کارها اساساً نمودارهایی هستند که ساختار داده آنها با زبان طبیعی و تصاویر (توالی ها و وصله ها) متفاوت است. این روش‌های مبتنی بر ترانسفورماتور نمی‌توانند اطلاعات گراف جهت‌دار غیرمدور را به‌طور دقیق مدل‌سازی کنند. در نتیجه، نمی‌توان آنها را مستقیماً به جریان‌های کاری ابری منتقل کرد. برای پرداختن به این مشکل، در این پژوهش یک نوع ترانسفورماتور مبتنی بر گراف جهت‌دار غیرمدور طراحی شده تا با گردش کار ابری سازگار شود.

در این مدل، ماتریس کدگذاری موقعیت P و ماتریس داده گردش کار X به بلوک های سفارشی ResNet جداگانه تغذیه می‌شوند. ResNet به طور انتخابی ویژگی‌ها را از لایه‌های مختلف استخراج می‌کند (با اتصالات میانبر) و آنها را با ویژگی‌های اصلی ترکیب می‌کند، که به طور تجربی به طور گسترده تأیید شده است، به ویژه در شبکه‌های بسیار عمیق برای کاهش مشکل ناپدید شدن گرادیان. در بلوک ResNet سفارشی، تابع فعال‌سازی ReLU برای حفظ اطلاعات خطی بیشتر حذف شد و تعامل غیرخطی را در ماژول توجه زیر به تعویق افتاد.

۵- ارزیابی و شبیه‌سازی

مجموعه داده: برای ارزیابی روش پیشنهادی از مجموعه داده GWA-T-12 Bitbrains که در آدرس اینترنتی http://gwa.ewi.tudelft.nl/datasets/gwa-t-12-bitbrains قرار دارد استفاده شده است. در این مجموعه داده معیارهای کارایی ۱۲۵۰ ماشین مجازی که متعلق به یک دیتاسنتر توزیع شده به نام Bitbrains است، قرار دارد. این دیتاسنتر سرویس‌های خاصی برای مدیریت هاست‌ها و محاسبات حرفه‌ای تجاری فراهم می‌آورد. مشتریان این دیتاسنتر اکثر بانک‌ها، اپراتورهای کارت‌های اعتباری و موسسات بیمه‌ای هستند. هر یک از فایل‌های موجود در این مجموعه اطلاعات یک ماشین مجازی را در خود نگهداری می‌نماید و شامل ۱۱ ویژگی است که در جدول۱ این ویژگی‌ها نشان داده شده است.

برای ارزیابی روش پیشنهادی ده ماشین مجازی به صورت تصادفی از مجموعه داده انتخاب و از فایل مربوط به هریک ۵۰۰ مورد را به طور تصادفی انتخاب نموده که تشکیل تعداد ۵۰۰۰ نمونه را می‌دهند و کلیه این مجموعه را در یک فایل به عنوان مجموعه ارزیابی روش پیشنهادی گردآوری نموده‌ایم.

جدول ۱: ویژگی های مجموعه داده

ردیف	نام	توضیحات
۱	Timestamp	برچسب زمانی
۲	CPU cores	تعداد هسته‌های CPU
۳	CPU capacity	میزان ظرفیت درخواستی
۴	CPU usage	میزان استفاده از پردازنده به MHZ
۵	CPU usageP	میزان استفاده از پردازنده به درصد
۶	Memory provisioned	میزان حافظه درخواستی
۷	Memory usage	میزان استفاده از حافظه
۸	Disk read throughput	مقدار گذردهی خواندن دیسک به KB/S
۹	Disk write throughput	مقدار گذردهی نوشتن دیسک به KB/S
۱۰	Network received throughput	میزان نرخ دریافت شبکه به KB/S
۱۱	Network transmitted throughput	میزان نرخ انتقال یا ارسال شبکه به KB/S

در روش پیشنهادی از الگوریتم بهینه‌سازی علف هرز مهاجم استفاده شده است. از آنجایی که مقادیر پارامترها تأثیر قابل توجهی بر عملکرد الگوریتم دارند، باید کالیبراسیون آماری آنها انجام شود که می‌تواند منجر به کیفیت بالای الگوریتم شود. در همین حال، لازم به توضیح است که کالیبراسیون یک فرآیند تنظیم دقیق است و انتظار نمی‌رود که الگوریتم‌ها پس از کالیبراسیون کاملاً متفاوت رفتار کنند. روش پیشنهادی شامل شش پارامتر است: اندازه جمعیت اولیه گیاه $N0$، حداکثر اندازه جمعیت گیاهی $P\max$، حداقل تعداد دانه $S\min$، حداکثر تعداد دانه $S\max$، حداقل انحراف استاندارد $\sigma\min$ و حداکثر انحراف استاندارد σ . در واقع، $N0$ تاثیر کمتری بر نتایج نهایی الگوریتم دارد زیرا تعداد علف‌های هرز در کلنی به سرعت حداکثر تعداد مجاز گیاه را به دست می‌آورد (ممکن است یک تکرار زمانی که مقدار اختلاف بین $P\max$ و $N0$ زیاد نباشد). بنابراین $N0$ به سادگی روی $P\max$

تنظیم می‌شود. برای سایر پارامترها، از تکنیک طراحی آزمایش برای کالیبره کردن آنها استفاده شده است.

ما ابتدا سطوح بالقوه همه پارامترها را بر اساس برخی آزمایش‌های اولیه و ادبیات مربوط به مشکلات زمان‌بندی مشابه تعیین کرده‌ایم. سطوح در نظر گرفته شده (مقادیر بالقوه) همه عوامل (پارامترها) به صورت زیر نشان داده شده است: $P\max$ ($N0$) در سه سطح آزمایش می‌شود: {۱۰، ۱۵، ۲۰}، $S\min$ در دو سطح آزمایش می شود: {۰، ۱}، $S\max$ در سه سطح آزمایش می‌شود: {۳، ۵، ۷، ۱۰، ۱۵، ۲۰}، $\sigma\min$ در چهار سطح آزمایش می شود: {۰،۰۵، ۰،۱۵، ۰،۲۵، ۰،۳۵}، و $\sigma\max$ در سه سطح آزمایش می شود: {۳،۰، ۵،۰، ۷،۰}. در نتیجه، در مجموع ۳ × ۲ × ۶ × ۴ × ۳ = ۴۳۲ پیکربندی به دست می‌آید. علاوه بر این، لازم به ذکر است که ادبیات قبلی نشان می‌دهد که $S\min$ باید مقدار کمتری تعیین شود، یعنی ۰ یا ۱ که این به این دلیل است که اگر $S\min$ مقدار بیشتری داشته باشد، ممکن است افراد فقیر بیشتری در اطراف افراد بدتر والدین ایجاد شوند، که این امر می‌تواند کاهش کیفیت کل جمعیت فرزندان را به همراه داشته باشد و در همین حال، اگر یک جستجوی محلی جامع نیز روی این افراد ایجاد شده جدید انجام شود، زمان محاسباتی زیادی صرف خواهد شد. بر اساس دلایل بالا و مطالعات قبلی، روش پیشنهادی دو مقدار بالقوه (یعنی ۰ یا ۱) را برای $S\min$ در نظر می‌گیریم. یک طرح آزمایشی فاکتوریل کامل برای همه پیکربندی‌ها در نظر گرفته شده است. نمونه‌های آزمایشی تولید شده توسط خودمان به‌عنوان بستر آزمایشی استفاده می‌شوند که شامل ۲۵ نمونه اولیه و سه سناریوی سررسید می‌شود. در نتیجه، در مجموع از ۲۵ × ۳ = ۷۵ نمونه برای کالیبراسیون استفاده می‌شود. هر پیکربندی روی این ۷۵ نمونه آزمایش می‌شود و برای هر نمونه ۵ تکرار مستقل مختلف اجرا می‌شود. معیار خاتمه الگوریتم به عنوان زمان سپری شده ۵۰ میلی ثانیه تنظیم می‌شود. بنابراین در مجموع ۴۳۲ × ۷۵ × ۵ = ۱۶۲۰۰۰ تیمار وجود دارد. برای تکمیل فرایند حداقل به ۱۶۸،۷۵ روز واحد پردازش مرکزی نیاز است. پس از اجرای همه آزمایش ها، در مجموع ۱۶۲۰۰۰ نتیجه (مجموعه راه حل های غیر غالب) تولید می‌شود. بعد از ایجاد اولویت اجرا و تنظیم پارامترهای لازم گراف جهت‌دار غیرمدور برای تخصیص بهترین وظیفه به بهترین منبع از خوشه‌بندی استفاده می‌شود. این خوشه بندی وظیفه مورد نیاز را با اطلاعات اضافی که الگوریتم علف هرز به مجموعه اضافه می‌نماید به بهترین منبع تخصیص می‌دهد.

همانگونه که در شکل ۲ نشان داده شده است ۵۹ درصد از وظایف در ماشین مجازی اول اجرا می‌شود، و به همین ترتیب ۲۰ و ۱۰ و ۴ و ۷ درصد از وظایف در منابع دو تا پنج به اجرا در می‌آیند. این ترتیب اجرا به صورت در نظر گرفتن وابستگی بین کارها و میزان اولویت آن‌ها برای اجرا انتخاب می‌شوند. با توجه به استفاده از این وابستگی و اولویت اجرای آن‌ها با انتخاب دوباره از وظایف در شکل ۴-۶ نیز اجرایی دیگر از این روند را نشان می‌دهد که با این اجرا درصدهای اجرا در ۵ ماشین مجازی با منابع یکسان به ترتیب ۴۲، ۱۶، ۱۴، ۲۰ و ۷ بوده است.

ارزیابی بخش خوشه‌بندی: برای تست خوشه‌بندی انجام شده از تعدادی طبقه‌بند و خود یادگیری عمیق در فاز طبقه‌بند استفاده شده است. این استفاده برای نمایش میزان دقت و صحت روش پیشنهادی در تخصیص منابع است و نتایج بیان‌گر این خواهند بود که روش پیشنهادی در تخصیص وظایف به منابع به چه صورتی عمل می‌نماید. در کلیه آزمایش‌ها از تست صحت به روش K-Fold با K=10 استفاده شده است. در این نوع اعتبارسنجی داده ها به K زیرمجموعه افراز می شوند. از این K زیرمجموعه، هر بار یکی برای اعتبارسنجی و K-1 تای دیگر برای آموزش بکار میروند. این روال K بار تکرار می‌شود و همه داده‌ها دقیقا یکبار برای آموزش و یکبار برای اعتبارسنجی بکار می روند. در نهایت میانگین نتیجه این K بار اعتبارسنجی به عنوان یک تخمین نهایی برگزیده می شود. البته می‌توان از روش‌های دیگر برای ترکیب نتایج استفاده کرد. به‌طور معمول از ۱۰-Fold استفاده می‌شود. الگوریتم های منتخب یادگیری ماشین جهت تست روش پیشنهادی به شرح زیر است.

K نزدیکترین همسایه: جستجوی نزدیکترین همسایه، که همچنین با نامهای جستجوی مجاورت، جستجوی همسانی یا جستجوی نزدیک ترین نقطه شناخته می‌شود، یک مسئله بهینه سازی برای پیدا کردن نزدیک‌ترین نقطه‌ها در فضاهای متریک است [۱۲].

```
Number of iterations: 5
Within cluster sum of squared errors: 22686.0

Initial starting points (random):

Cluster 0: 1377865698;,11703.99824;,1.1833333333333333;,447392.0;,21.6;,1.4
Cluster 1: 1378436054;,11703.99824;,0.5666666666666667;,0.0;,1.0;,1
Cluster 2: 1377919102;,11703.99824;,0.55;,0.0;,0.3333333333333333;,1
Cluster 3: 1376387752;,11703.99824;,0.6833333333333332;,0.0;,1.4;,1
Cluster 4: 1378331644;,11703.99824;,0.6333333333333333;,0.0;,1.2666666666666666;,1

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                    Cluster#
Attribute                                Full Data          0               1               2               3               4
                                          (8634.0)     (5128.0)        (1760.0)        (827.0)         (353.0)         (566.0)
=============================================================================================================================
Timestamp [ms];                        1376314846;  1376314846;     1376318446;     1376322046;     1376317246;     1376327747;
CPU capacity provisioned [MHZ];        11703.99824; 11703.99824;    11703.99824;    11703.99824;    11703.99824;    11703.99824;
CPU usage [%];                 0.6166666666666667;          0.6; 0.5666666666666667;            0.55; 0.6833333333333332; 0.6333333333333333;
Memory usage [KB];                             0.0;          0.0;              0.0;            0.0;            0.0;            0.0;
Disk write throughput [KB/s];  0.6666666666666666;          1.0; 0.6666666666666666; 0.3333333333333333; 0.6666666666666666; 0.6666666666666666;
Network transmitted throughput [KB/s]          1;            1;              1;              1;              1;              1



Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      5128 ( 59%)
1      1760 ( 20%)
2       827 ( 10%)
3       353 (  4%)
4       566 (  7%)
```

شکل ۱: تخصیص وظایف به منابع در خوشه بندی یادگیری عمیق انجام شده

J48: یکی دیگر از الگوریتم‌های دسته‌بندی درخت تصمیم است. درخت تصمیم از مجموعه ای از قوانین ساده تشکیل شده است. درخت های تصمیم غیر پارامتری هستند زیرا نیاز به هیچ تخمینی از توزیع متغیرهای هر کلاس ندارند. درخت تصمیم از یک درخت برای ساخت یک مدل پیش بینی (تخمین) استفاده می‌کند که مشاهدات درباره یک آیتم را به نتیجه‌گیری‌هایی درباره مقدار هدف آن آیتم نگاشت می‌کند [۱۲].

بهینه سازی متوالی کمینه[۱۳]: ماشین بردار پشتیبان الگوریتم طبقه‌بندی بوده و به عنوان یکی از بهترین تکنیک های دسته بندی و پیش بینی و تشخیص داده‌های پرت شناخته می شود و برخلاف الگوریتم های خوشه بندی در دسته یادگیری با نظارت محسوب می شود و دو فاز آموزش و تست دارد. ماشین بردار پشتیبان ابتدا توسط آقای وپنیک برای تفکیک و دسته بندی داده هایی که جدایی پذیر خطی بودند ساخته شد ولی بعدها توسط خود ایشان و آقای کورتس برای حالت غیر خطی هم تعمیم داده شد. در واقع ماشین بردار پشتیبان برای اینکه داده های غیر خطی را از هم تفکیک کند باید از کرنل های مختلف استفاده کند. [۱۲].

الگوریتم مبتنی بر نظریه بیز[۱۴]: تئوری بیز یکی از روش‌های آماری برای رده‌بندی به شمار می‌آید. در این روش کلاس‌های مختلف، هر کدام به شکل یک فرضیه دارای احتمال در نظر گرفته می‌شوند. هر رکورد آموزشی جدید، احتمال درست بودن فرضیه‌های پیشین را افزایش و یا کاهش می‌دهد و در نهایت، فرضیاتی که دارای بالاترین احتمال شوند، به عنوان یک کلاس در نظر گرفته شده و برچسبی بر آن‌ها زده می‌شود. این تکنیک با ترکیب تئوری بیز و رابطه سببی بین داده‌ها، به طبقه‌بندی می‌پردازد [۱۲].

بگینگ[۱۵]: از مفهوم تجمیع بوت‌استرپ[۱۶] درایجاد تخمینهای مختلف استفاده نموده است. اصولاً میتوان از تکنیک فوق به منظور ارزیابی دقت تخمینهای بکار گرفته شده در روشهای داده کاوی از طریق نمونه برداری با جایگزینی از داده‌های آموزشی استفاده نمود. در این تکنیک فرض بر آنست که مجموعه‌داده های آموزشی نماینده جامعه تحت بررسی بوده و انواع حالات تحقق یافته جامعه را میتوان از این مجموعه داده شبیه سازی نمود. بنابراین با استفاده از دوباره نمونه‌گیری توسط به کارگیری مجموعه داده‌های مختلف تنوع

[13] Sequential Minimal Optimization (SMO)

[14] Bayes and Bayesian Networks
[15] **Bagging**
[16] Bootstrap Aggregating

مورد نیاز حاصل خواهد شد و زمانی که یک نمونه جدید وارد هرکدام از کلاسه بندها می‌شود، توافق اکثریتی به‌کار گرفته می‌شود تا کلاس مورد نظر تشخیص داده شود.

الگوریتم شبکه‌های عصبی عمیق: شبکه‌های عصبی عمیق از مجموعه‌ای از نورون‌ها تشکیل شده‌اند که به‌صورت لایه‌هایی سازماندهی شده‌اند. این لایه‌ها شامل لایه ورودی، لایه‌های پنهان و لایه خروجی هستند. هر نورون در یک‌لایه با نورون‌های لایه قبلی و لایه بعدی ارتباط دارد و این ارتباطات به‌وسیله وزن‌ها تنظیم می‌شوند. شبکه‌های عصبی عمیق الگوریتم‌های یادگیری ماشینی هستند که از ساختار مغز انسان الهام گرفته‌اند. این شبکه‌ها از لایه‌های متعددی تشکیل شده‌اند که هر لایه وظیفه پردازش اطلاعات ورودی را بر عهده دارد. همچنین معیارهای ارزیابی به ترتیب زیر است:

صحت: اثربخشی یک مدل با نسبت پیش‌بینی‌های دقیق تولید شده در تمام انواع پیش‌بینی‌ها ارزیابی می‌شود. فرآیند ارزیابی شامل ارزیابی صحت طبقه‌بندی است یعنی نمونه‌های طبقه‌بندی شده با تعداد کلی رخدادها. اندازه‌گیری صحت به ویژه در مواردی ارزشمند است که توزیع کلاس‌ها در متغیر هدف به طور یکنواخت در سراسر مجموعه داده پخش شود. رابطه صحت در معادله (۵) بیان شده است: [۱۳]

$$TP\ Rate = \frac{(TP)}{(TP+F\ )} \qquad (۵)$$

$$FP\ Rate = \frac{(FP)}{(FP+T\ )} \qquad (۶)$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+T}} \quad (۷)$$

یادآوری: یادآوری اندازه‌گیری نرخ مثبت واقعی در زمینه یک سیستم نقص نرم‌افزاری است. در این زمینه خاص، تعداد رخدادهایی را نشان می‌دهد که به عنوان نرم‌افزار معیوب طبقه‌بندی شده‌اند که به‌طور دقیق توسط مدل پیش‌بینی شده‌اند. معادله ۸ نشان دهنده نسبت نمونه‌های نرم‌افزاری مشکل‌ساز است که به طور دقیق توسط مدل شناسایی شده-اند: [۱۳]

$$\text{Recall} = \frac{\text{TP}}{\text{TP+F}} \quad (۸)$$

دقت: دقت پیش‌بینی‌های مثبت انجام‌شده توسط مدل را اندازه‌گیری می‌کند و نسبت نمونه‌های پیش‌بینی‌شده به‌عنوان مثبت و مثبت واقعی را نشان می‌دهد و با استفاده از فرمول زیر محاسبه می شود: [۱۴]

$$\text{precision} = \frac{\text{TN}}{\text{TN+F}} \quad (۹)$$

امتیاز F1: معیاری است که دقت و یادآوری را در یک مقدار واحد ترکیب می‌کند. این به ویژه زمانی مفید است که توزیع کلاس ناهموار (کلاس های نامتعادل) وجود داشته باشد و اغلب در مسائل طبقه بندی باینری استفاده می شود.این معیار میانگین هارمونیک دقت و یادآوری است و با استفاده از فرمول زیر محاسبه می شود: [۱۴]

$$F - \text{Measure} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision}\quad\text{recall}} \quad (۱۰)$$

معرفی ضریب همبستگی متیوز (MCC): ضریب همبستگی متیوز در واقع، همبستگی کلاس‌های واقعی c را با برچسب‌های پیش‌بینی‌شده اندازه‌گیری می‌کند[۱۵]:

$$MCC = \frac{\text{TP.TN}-\text{FP.FN}}{\sqrt{(\text{TP+F}\ ).(\text{TP+FN}).(\text{TN+FP}).(\text{TN+FN})}} \quad (۱۱)$$

ناحیه RC [17]: ناحیه منحنی دقت-یادآوری این یک عدد است که قابلیت‌های مدل را توصیف می‌کند. امتیاز PR AUC میانگین امتیازات دقت محاسبه شده برای هر آستانه یادآوری [۰،۰، ۱،۰] است. این منحنی پاسخ فاز با ترکیب ارزش پیش‌بینی مثبت و نرخ مثبت واقعی بدست می‌آید .(شکل ۱۰) برای هر آستانه، ارزش پیش‌بینی مثبت و نرخ مثبت واقعی محاسبه شده و نقطه مربوطه در نمودار رسم می‌شود .ترجیحاً، الگوریتم دارای دقت و حساسیت بالا است .این دو معیار مستقل نیستند .به همین دلیل است که بین آنها مصالحه صورت می‌گیرد. یک منحنی PRC خوب AUC بالاتری دارد. تحقیقات نشان داده است که PRC از نظر گرافیکی هنگام تخمین طبقه‌بندی‌کننده‌های دودویی روی مجموعه‌های نامتوازن، آموزنده‌تر از نمودارهای ROC است. [۱۶]

ناحیهROC: ناحیه مشخصه عملیاتی گیرنده - منحنی ROC نموداری است که بده بستان بین نرخ مثبت واقعی و نرخ مثبت کاذب را نشان می‌دهد. برای هر آستانه، نرخ مثبت واقعی و نرخ مثبت کاذب را محاسبه کرده و آنها را روی یک نمودار رسم می‌شود. هرچه نرخ مثبت واقعی بالاتر و نرخ مثبت کاذب پایین‌تر برای هر آستانه باشد، بهتر است. طبقه‌بندی‌کننده‌های بهتر منحنی‌های بیشتری در سمت چپ دارند. ناحیه زیر منحنیROC ، امتیاز AUC ROC نامیده

[17] PRC Area

جدول ۲: نمایش تست تخصیص وظایف به منابع با طبقه بندهای مختلف

TP Rate	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۹۲۳	۰,۹۴۶	۰,۹۳۴	۰,۹۰۹	۰,۹۵۴	۰,۸۹۹	۰,۹۷۴
FP Rate	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۰۷۷	۰,۰۵۴	۰,۰۶۶	۰,۰۹۱	۰,۰۴۶	۰,۱۰۱	۰,۰۲۶
Precision	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۹۰۱	۰,۹۰۹	۰,۸۹۸	۰,۸۹۵	۰,۹۳۱	۰,۸۷۷	۰,۹۵۱
MCC	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۷۷۳	۰,۷۸۸	۰,۷۷۲	۰,۷۴۶	۰,۸۱	۰,۷۳۹	۰,۸۳۴
ROC Area	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۹۴۳	۰,۹۲۱	۰,۹۴۸	۰,۹۵۹	۰,۹۶۷	۰,۹۱۷	۰,۹۶۵
Recall	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۹۲۳	۰,۹۴۶	۰,۹۳۴	۰,۹۰۹	۰,۹۵۴	۰,۸۹۹	۰,۹۷۴
F-Measure	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۹۰۵	۰,۹۱۲	۰,۹۰۱	۰,۸۹	۰,۹۳۳	۰,۸۸۳	۰,۹۵۲
Accuracy	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۹۲,۲۶۹۵	۹۴,۵۶۶۱	۹۳,۳۵۵۷	۹۰,۸۸۰۸	۹۵,۳۵۵	۸۹,۸۹۱۹	۹۷,۴۲۵۹
PRC Area	KNN	J48	Decision Table	NaiveBayes	SMO	Bagging	DNN
Data	۰,۹۰۹	۰,۸۶۲	۰,۹۰۲	۰,۹۱۳	۰,۹۴۲	۰,۸۸	۰,۹۷۶

می‌شود، عددی که تعیین می‌کند منحنی ROC چقدر خوب است [۱۶].

براساس این نتایج روش پیشنهادی می‌تواند تخصیص مناسبی برای این منظور داشته باشد زیرا در بدترین حالت نیز درصد بالای ۹۰ درصد صحت تخصیص را از خود نشان داده است. ارزیابی بعدی با روش پایه [۱۷] است که این روش نیز به مانند روش پیشنهادی از یادگیری عمیق استفاده نموده است و از آن برای اثبات روش خود نیز استفاده نموده است. این روش در تست صحت خود به مقدار ۹۰,۸ دست یافته است که روش پیشنهادی در یادگیری عمیق به مقدار ۹۷,۴ دست یافته است و بهبود ۶,۶ درصد از خود به ثبت رسانده است. مقایسه این دو روش در جدول ۳ نشان داده شده است.

جدول۳: مقایسه روش پیشنهادی و روش پایه از نظر صحت تخصیص ها

نام روش	مقدار معیار صحت
روش پیشنهادی	۹۰,۸
روش پایه	۹۷,۴

همانگونه که در جدول ۳ نشان داده شده است. روش پیشنهادی به نسبت روش پایه مقدار ۶,۶ درصد بهبود از خود به ثبت رسانده است. محاسبات با کارایی بالا اکنون یک حوزه اصلی است که در آن فناوری‌های تجاری و محاسباتی برای برآورده کردن تداوم کسب‌وکار و نیازهای زمان واقعی نیاز به عملکرد بالا انعطاف‌پذیر دارند. با این حال، بسیاری از سازمان های تجاری و فناوری سطح بالا هنوز در حال بهبود عملکرد بالا و انعطاف پذیری ترافیک هستند تا از در دسترس بودن سیستم در همه زمان ها اطمینان حاصل کنند. یادگیری ماشینی پیشرفت مهمی در فناوری کامپیوتر است که با پیش بینی و مکانیسم طبقه‌بندی بر اساس داده‌های تاریخی به تصمیم‌گیری کمک می‌کند. در روش پیشنهادی، ما مفهوم محاسبات با عملکرد بالا را با تکنیک‌های یادگیری ماشینی هوش مصنوعی در پلتفرم‌های ابری پیشنهاد و ادغام می‌کنیم. داده‌های عملکرد شبکه و محاسبات برای اعتبارسنجی، پیش‌بینی و طبقه‌بندی ترافیک و الگوهای عملکرد و اطمینان از عملکرد سیستم و تصمیم‌های انعطاف‌پذیری جریان ترافیک مداوم استفاده می‌شود. رویکرد طراحی یکپارچه پیشنهادی بر

اساس اقدامات و تصمیمات گامی مختلف بر اساس مدل‌های رگرسیون و طبقه‌بندی یادگیری ماشین، که عملکرد سیستم را در نمونه‌های زمان واقعی تصحیح می‌کند، تحلیل شده است. نتایج شبیه‌سازی‌شده طراحی یکپارچه یادگیری ماشین ما نشان می‌دهد که انعطاف‌پذیری ترافیک آن در مقایسه با مدل‌های طراحی مبتنی بر یادگیری ماشینی امروزی، با توجه به بازیابی نقطه شکست و ۷٫۵ درصد صرفه‌جویی در هزینه‌های تجاری، ۳۸٫۱۵ درصد سریع‌تر عمل می‌کند.

۶- نتیجه‌گیری

رایانش ابری یک مدل محاسباتی تجاری جدید است و از طریق اینترنت متصل می‌شود و دسترسی به اینترنت در دسترس، راحت و بر اساس تقاضا را به یک مخزن اشتراک‌گذاری منابع محاسباتی قابل تنظیم را فراهم می‌کند. مجازی‌سازی نیز یک فناوری کلیدی در رایانش ابری که از طریق این فناوری، تعداد زیادی از گره‌های محاسباتی یکپارچه می شوند تا یک منبع کلی را تشکیل دهند. قابل به ذکر است که زمان‌بندی کار و تخصیص منابع، موضوعات مهم مربوط به بسیاری از چالش های تحقیقاتی رایانش ابری است. در واقع برای تعداد زیادی از درخواست‌ها که در انتظار پردازش هستند، از زمان‌بندی وظایف برای تخصیص منطقی منابع محاسباتی به درخواست‌ها استفاده می‌شود. همچنین تعادل بار روشی برای بهینه‌سازی منابع ماشین‌های مجازی در محیط رایانش ابری بوده که یکی از تکنیک‌های مهمی است که برای اطمینان از توزیع برابر و پویا حجم کار و استفاده کارآمد از منابع مورد استفاده قرار می‌گیرد. در این پژوهش روشی برای بهبود طبقه‌بندی ماشین‌های مجازی جهت ایجاد تعادل بار کاری در برنامه‌های رایانش ابری با استفاده از شبکه‌های عصبی عمیق و الگوریتم تکاملی علف هرز ارائه شده است که با توسعه سریع سخت‌افزار و نرم‌افزار کامپیوتر، یادگیری تقویت عمیق جهت جدیدی را برای حل بهترظمسائل زمان بندی کار فراهم می‌کند. در روش پیشنهادی ابتدا از روش شبکه عصبی عمیق برای انتخاب یک ماشین مجازی مناسب برای کار استفاده شده و سپس تعیین می‌شود که آیا اجرای کار روی ماشین مجازی انتخاب شده توافق نامه سطح خدمات را نقض می کند یا خیر که اگر نقض شود، کار رد شده و بازخورد یک پاداش منفی برای آموزش شبکه عصبی عمیق دارد و در غیر این صورت، وظیفه دریافت و اجرا می‌شود و پس از اجرای کار، با توجه به تعادل بار ماشین‌های مجازی، پاداش دریافت می-شود. این الگوریتم پیشنهادی بهترین عملکرد را در متعادل کردن بار ماشین‌های مجازی و کاهش نرخ رد کار، بهبود سطح کلی خدمات رایانش ابری از خود نشان داده است.

اگرچه چندین روش زمان‌بندی برای انواع محیط‌های محاسباتی وجود دارد، اما اگر به‌طور خاص برای آن طراحی نشده باشند، باید برای محیط ابری تنظیم شوند. به عنوان مثال، در یک خوشه کامپیوترهای مستقل و همگنی وجود دارد که با هم به عنوان یک سیستم عمل می‌کنند و این سیستم‌ها از منابع محلی خود استفاده می‌کنند، در حالی که یک محیط ابری شامل مجموعه ای از منابع ناهمگن توزیع شده است و کاربران می‌توانند از همه آنها استفاده کنند. بنابراین، یک الگوریتم زمان‌بندی که می‌تواند برای یک خوشه خوب باشد ممکن است برای یک محیط ابری مناسب نباشد. فضای مورد نظر(چالشی) باید در عناصر الگوریتم نگاشت شود تا الگوریتم بتواند برای ویژگی‌های ساختاری محیط ابر اعمال شود. افزایش ماشین‌های مجازی ناهمگن و تنوع اندازه وظایف منجر به تعداد زیادی ترتیبات کار می‌شود. یافتن جایگشت مناسب با حداقل زمان تکمیل در بین همه جایگشت ها یک مسئله سخت است. بسیاری از مطالعات مرتبط با استفاده از الگوریتم‌های فراابتکاری در ارتباط با زمان‌بندی در یک محیط ابری انجام شده‌اند.

در این پژوهش علاوه بر ارایه روشی بر پایه الگوریتم تکاملی بهینه علف هرز مهاجم تکنیک‌های متوازن بار متعدد را در سه محیط مختلف استاتیک، دینامیک و الهام‌گرفته از طبیعت را توصیف و بررسی کرد. محققان می‌توانند به تحقیقات بیشتر در مورد چگونگی طراحی الگوریتم‌هایی که پویاتر و هوشمندتر هستند و می‌توانند مشکلات تحمل خطا را برای افزایش بیشتر عملکرد محاسبات ابری حل کنند، ادامه دهند. در آینده، بررسی الگوریتم‌های الهام‌گرفته از طبیعت و هوشمند مانند کاربرد تکنیک‌های یادگیری ماشین یا خوشه‌بندی خواهد بود. بر اساس این ادبیات، نتیجه گیری می شود که هنوز موضوعات تحقیقاتی باز وجود دارد که در آینده باید به آنها پرداخت. در این روش زمینه های بهبود الگوریتم های متعادل کننده بار را برای محققان شناسایی کرده است تا در تحقیقات آینده خود در این زمینه برای بهینه سازی بیشتر

منابع در رایانش ابری در نظر بگیرند. این موارد در زیر ذکر شده است:

اکثر الگوریتم‌های بررسی شده در اینجا تحمل خطا ندارند. این یکی از عوامل مهمی است که می توان برای طراحی یک الگوریتم پایدار برای عملکرد خوب در محیط ابری پویا در نظر گرفت. اشکال در الگوریتم ممکن است به دلایل زیادی مانند اضافه شدن ناگهانی گره ها در مرکز داده ابری، کار با اولویت بالا در انتظار اجرا، تغییر ناگهانی در حجم کار و تنظیمات پیکربندی ماشین های مجازی وجود داشته باشد.

با این حال، چالش‌های باز در انتقال وظیفه وجود دارد، زیرا تعداد کمی از نویسندگان بر پارامتر زمان مهاجرت تمرکز کرده‌اند.

برخی از رویکردها به دلیل الگوریتم‌های استاتیک زیربنایی مورد استفاده مانند RR یا FCFS، همچنان منجر به زمان انتظار بالاتری می‌شوند. برای حل این مشکل، محققان می توانند بررسی کنند که چگونه الگوریتم های هوشمند می توانند برای تقلید از محیط ابری مانند الگوریتم های الهام گرفته از طبیعت که می توانند مسائل پیچیده بهینه سازی را حل کنند، استفاده کنند.

نکات فوق منجر به یک شکاف تحقیقاتی جدید می‌شود که می‌تواند توسط محققان علاقه‌مند به این حوزه برای بهینه-سازی بیشتر و بهبود عملکرد برنامه های رایانش ابری در نظر گرفته شود. از آنجایی که توازن بار برای بهینه‌سازی ابر مفید است، تخصیص وظیفه آگاه از انرژی، که به برنامه‌های مختلف ابری، عمدتاً برای برنامه‌های کاربردی حیاتی مرتبط با سلامت کمک می‌کند.

۶- منابع

1. *Ebadifard, F. and S.M. Babamir, A PSO-based task scheduling algorithm improved using a load-balancing technique for the cloud computing environment. Concurrency and Computation: Practice and Experience, 2018. 30(12): p. e4368.*
2. *Guo, L., et al., Task scheduling optimization in cloud computing based on heuristic algorithm. Journal of networks, 2012. 7(3): p. 547.*
3. *Kaur, R. and K.S. Dhindsa, Efficient Task Scheduling using Load Balancing in Cloud Computing. International Journal of Advanced Networking and Applications, 2018. 10(3): p. 3888-3892.*
4. *Sui, X., et al., Virtual machine scheduling strategy based on machine learning algorithms for load balancing. EURASIP Journal on Wireless Communications and Networking, 2019. 2019: p. 1-16.*
5. *Sansanwal, S. and N. Jain, An Improved Approach for Load Balancing among Virtual Machines in Cloud Environment. Procedia Computer Science, 2022. 215: p. 556-566.*
6. *Elrotub, M. and A. Gherbi, Virtual machine classification-based approach to enhanced workload balancing for cloud computing applications. Procedia computer science, 2018. 130: p. 683-688.*
7. *Jarraya, M. and S. Elloumi, Load balancing scheduling algorithms for virtual computing laboratories in a Desktop-As-A-Service Cloud Computing Services. Computer Communications, 2022. 192: p. 343-354.*
8. *Deshmukh, S.R., S. Yadav, and D. Kyatanvar, Load balancing in cloud environs: Optimal task scheduling via hybrid algorithm. International Journal of Modeling, Simulation, and Scientific Computing, 2021. 12(02): p. 2150008.*
9. *Chiang, M.-L., et al., Improvement of tasks scheduling algorithm based on load balancing candidate method under cloud computing environment. Expert Systems with Applications, 2023. 212: p. 118714.*
10. *Ananth, C.A. and N. Krishnaraj, Detection of intrusions in clustered vehicle networks using invasive weed optimization using a deep wavelet neural networks. Measurement: Sensors, 2023: p. 100807.*
11. *MohiEldeen Alabbasy, F., A. Abohamama, and M.F. Alrahmawy, Compressing medical deep neural network models for edge devices using knowledge distillation. 2023.*
12. *Islam, M.A., Uddin, M. A., Aryal, S., & Stea, G. , An ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes. Journal of Information Security and Applications, 78, 103618., 2023.*
13. *Kene Tochukwu Anyachebelu, S.H.H., Muhammad Umar Abdullahi, Maimuna Abdullahi Ibrahim, Comparative Analysis of Machine Learning Algorithms for Breast Cancer Prediction Dutse Journal of Pure and Applied Sciences (DUJOPAS), Vol. 9 No. 4b December 2023, 2023.*
14. *Islam, M.R., M.S. Islam, and S. Majumder, Breast cancer prediction: a fusion of genetic algorithm, chemical reaction optimization, and machine learning techniques. Applied Computational Intelligence and Soft Computing, 2024. 2024(1): p. 7221343.*
15. *Chicco, D., N. Tötsch, and G. Jurman, The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData mining, 2021. 14(1): p. 13.*

Analysis and investigation of task allocation to virtual machine resources for load balancing in cloud computing based on various machine learning classifiers

A.H Nazari Afshar[1] and H Doost[2]
[1]Tehran Azad University, Yadegar Imam Khomeini Branch, Shahr-e Ray
[2]Tehran Azad University, Science and Research Branch

Abstract

Load balancing is an important aspect in the field of cloud computing to increase workload distribution and efficient resource utilization, which in turn reduces the overall system response time. Many approaches and algorithms have been proposed to solve load balancing problems such as task scheduling, migration, resource utilization, etc. This study presents several approaches related to the critical challenge in cloud computing, which is load balancing. The problems related to load balancing are discussed through a comparative analysis of the algorithms proposed by researchers in the last six years. Several approaches have been proposed, however, there are still some issues in the cloud environment, such as virtual machine migration, fault tolerance issues that have not been fully resolved yet. In this study, an improved virtual machine classification model is presented for workload balancing in cloud computing applications using deep neural networks and evolutionary weeding algorithm. In fact, the main objective of the study is to optimize the use of computing resources in the cloud environment, where scheduling algorithms play an important role in the optimization process. In fact, the proposed topic is in the field of machine learning and virtual machine classification for load balancing in cloud computing. In the simulation of the proposed method, in the evaluation of the comparison of the proposed method and the base method in terms of the accuracy of the assignments, the proposed method has recorded an improvement of 6.6 percent compared to the base method.

16. *Vujović, Ž., Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 2021. 12(6): p. 599-606.*
17. *Kamila, N.K., et al., Machine learning model design for high performance cloud computing & load balancing resiliency: An innovative approach. Journal of King Saud University-Computer and Information Sciences, 2022. 34(10): p. 9991-10009.*

امیر حسین نظری افشار فارغ التحصیل کارشناسی ارشد رشته مهندسی کامپیوتر گرایش نرم افزار از دانشگاه آزاد واحد یادگار امام (ره) و نشانه رایانامه ایشان عبارتند از :
root.afshar@gmail.com

حمیدرضا دوست دانشجوی کارشناسی ارشد مهندسی فناوری اطلاعات گرایش تجارت الکترونیک از دانشگاه آزاد واحد علوم و تحقیقات و نشانه رایانامه ایشان عبارتند از:
dhamidreza@rocketmail.com

روش ارجـاع: ا.ح نظـری افشـار و ح. دوسـت. تحلیـل و بررسـی تخصیص وظایف به منابع ماشینهای مجازی برای ایجاد تعادل بـار در رایانش ابری مبتنـی بـر طبقـه بنـدهای مختلـف یـادگیری ماشـین. دوفصلنامه محاسبات و سامانه‌های توزیع شده، سال هفتم، شـماره۲، شماره پیاپی ۱۴، صفحه ۵۶ تا ۷۰ سال ۱۴۰۳.

How to cite: A.H. Nazari Afshar and H. Doost. Analysis and investigation of task allocation to virtual machine resources for load balancing in cloud computing based on various machine learning classifiers. Journal of Distributed Computing and Systems (JDCS), Vol 7, Issue ۱۴, Pages ۵۶ – ۷۰, 2025.