

دریافت مقاله: ۱۴۰۴/۰۳/۰۸

پذیرش مقاله: ۱۴۰۴/۰۴/۰۹

نوع مقاله: پژوهشی

مقیاس‌بندی گنجانه‌ها با استفاده از سیستم‌های توصیه‌گر آگاه از زمینه

مهرداد آشتیانی*

استادیار، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت، تهران، ایران
پست الکترونیکی: m_ashtiani@iust.ac.ir

هانیه حق‌شناس

فارغ‌التحصیل رشته مهندسی کامپیوتر، دانشگاه علم و صنعت، تهران، ایران
پست الکترونیکی: hanihaghshenas11@gmail.com

فاطمه بخشی

کارشناسی ارشد مهندسی کامپیوتر، دانشگاه علم و صنعت، تهران، ایران
پست الکترونیکی: fatemeh_bakhshi@comp.iust.ac.ir

چکیده

مقیاس‌بندی مؤثر گنجانه‌ها در رایانش ابری با توجه به بارهای کاری پویا، چالشی اساسی است. روش‌های سنتی مقیاس‌بندی محدودیت‌هایی در درک نوسانات بیدرنگ و عوامل زمینه‌ای دارند که منجر به تخصیص غیربهبوده منابع و کاهش عملکرد می‌شود. در این پژوهش با بهره‌گیری از فیلترگذاری مشارکتی، فیلترگذاری مبتنی بر محتوا و استدلال هستی‌شناختی، سیستمی ایجاد شده است که تصمیمات مقیاس‌بندی را بهبود می‌بخشد. مدل پیشنهادی از داده‌های تاریخی، اطلاعات بیدرنگ و پارامترهای زمینه‌ای برای ارائه توصیه‌های هوشمندانه بهره می‌گیرد. آزمایش‌ها در محیط شبیه‌سازی‌شده نشان دادند که سیستم پیشنهادی در مقایسه با سیستم‌های آستانه‌محور با کاهش اقدامات مقیاس‌بندی غیرضروری، ۱۹ درصد دقت

تصمیم‌گیری را بهبود بخشیده است. با این که زمان اجرا و تأخیر در سیستم پیشنهادی تقریباً دو برابر زمان اجرا و تأخیر در سیستم‌های مبتنی بر قوانین است ولی همچنان در بازه قابل قبول برای استفاده محیط‌های بزرگ ابری قرار می‌گیرد. بنابراین، با حذف برخی محدودیت‌ها مانند وابستگی به قوانین از پیش تعیین‌شده، سیستم پیشنهادی، گامی در جهت مدیریت هوشمند و خودکار منابع در زیرساخت‌های ابری برداشته است.

واژه‌های کلیدی: مقیاس‌بندی گنجانه‌ها، سیستم توصیه‌گر ترکیبی، رایانش آگاه از زمینه، مدیریت منابع ابری، فیلترگذاری.

۱. مقدمه

یکی از چالش‌های مهم در رایانش ابری مقیاس‌بندی

* نویسنده مسئول

مؤثر گنجانه‌ها^۱ در شرایط بار کاری پویا است [۱]. گنجانه‌ها به‌عنوان واحدهای سبک‌وزن و انعطاف‌پذیر اجرای سرویس‌ها و برنامه‌ها، نقش مهمی در بهره‌وری منابع و چابکی زیرساخت‌های ابری ایفا می‌کنند. با این حال، مدیریت منابع تخصیص‌یافته به این گنجانه‌ها، به‌ویژه در شرایطی که بار کاری به سرعت تغییر می‌کند، نیازمند تصمیم‌گیری‌های سریع، دقیق و زمینه‌محور است. روش‌های سنتی مقیاس‌بندی که بر پایه قوانین ایستا و از پیش تعریف شده عمل می‌کنند، فاقد توانایی درک و تحلیل پیچیدگی‌های واقعی سیستم هستند. این روش‌ها معمولاً تغییرات آنی در بار کاری، تفاوت‌های موجود در نوع درخواست‌ها، وابستگی‌های بین سرویس‌ها، و سایر عوامل زمینه‌ای مانند زمان، مکان جغرافیایی یا اولویت‌های کسب‌وکار را نادیده می‌گیرند. در نتیجه، چنین محدودیت‌هایی اغلب به تصمیم‌گیری‌های نامناسب در تخصیص منابع منجر می‌شوند که این امر می‌تواند پیامدهای متعددی مانند افزایش هزینه‌های عملیاتی، کاهش کارایی کلی سیستم، کاهش سطح کیفیت سرویس و در شرایط بحرانی حتی از دست رفتن دسترسی‌پذیری سرویس‌ها در زمان‌های اوج مصرف را به دنبال داشته باشد [۴-۲].

انگیزه اصلی این پژوهش از افزایش پیچیدگی زیرساخت‌های ابری و ضعف روش‌های فعلی در مدیریت پویا ناشی می‌شود. ترکیب مدل‌های توصیه‌گر با آگاهی از زمینه، راهکاری نوآورانه برای غلبه بر این محدودیت‌ها ارائه می‌دهد. این مدل در مقایسه با روش‌های سنتی مقیاس‌بندی، دقت و سازگاری بیشتری دارد. همچنین تمرکز آن بر حفظ عملکرد در محیط‌های بزرگ مقیاس و با تأخیر کم است. در نهایت، سیستم پیشنهادی می‌تواند نقشی کلیدی در بهینه‌سازی هزینه‌ها و افزایش بهره‌وری در سیستم‌های ابری ایفا کند. برای حل این مشکل، پژوهش حاضر با استفاده از مدل توصیه‌گر ترکیبی سیستم جدیدی را معرفی می‌کند. این مدل، با ترکیب فیلترگذاری مشارکتی، فیلترگذاری مبتنی بر محتوا و استدلال مبتنی بر هستی‌شناسی طراحی شده

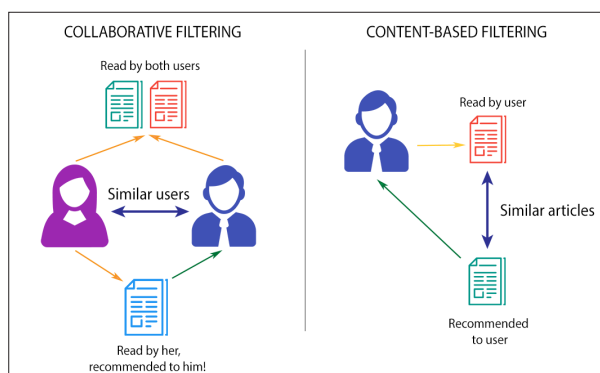
است. هدف آن، بهینه‌سازی فرآیند مقیاس‌بندی با در نظر گرفتن داده‌های تاریخی، استفاده بیدرنگ از منابع و عوامل زمینه‌ای مانند ساعات اوج مصرف یا شرایط جغرافیایی است. استفاده از هستی‌شناسی به سیستم کمک می‌کند تا روابط زمینه‌ای را درک کرده و تصمیمات هوشمندانه‌تری اتخاذ کند. برای ارزیابی سیستم پیشنهادی، شبیه‌سازی انجام گرفته که در آن فرآیند مقیاس‌بندی گنجانه‌ها طراحی شده است. در مرحله اول از این فرآیند داده‌های مصنوعی مثل هزینه پردازشی تولید می‌شوند. سپس بارهای کاری پویا برای شبیه‌سازی شرایط واقعی تولید می‌شوند. در مرحله بعد با ترکیب فیلترگذاری مبتنی بر محتوا و مشارکتی، نیازهای مقیاس‌بندی پیش‌بینی شده و عوامل مفهومی (مانند اوج مصرف) برای تصمیم‌گیری هوشمند تحلیل می‌شوند. در نهایت امتیازات ترکیبی و استدلال هستی‌شناسی ادغام شده و اقدامات مقیاس‌بندی اجرا می‌شود. سیستم پیشنهادی در مقایسه با سیستم‌هایی که از مقیاس‌بندی مبتنی بر قوانین ثابت استفاده می‌کنند در مواردی که تصمیم به مقیاس‌بندی گرفته شده در حالی که مقیاس‌بندی غیر ضروری بوده ۱۳ درصد و در مواردی که نیاز به مقیاس‌بندی به درستی تشخیص داده نشده است ۱۵ درصد بهبود داشته است. همچنین بر اساس نتایج گزارش شده، در سیستم پیشنهادی ساعات اوج مصرف، استفاده بالای شبکه و محدودیت‌های جغرافیایی از جمله تأثیرگذارترین عوامل زمینه‌ای در اتخاذ تصمیمات مقیاس‌بندی هستند.

در ادامه این مقاله، در بخش ۲ مفاهیم پایه شرح داده می‌شود، در بخش ۳ پژوهش‌های مرتبط مورد بررسی قرار می‌گیرد. رویکرد پیشنهادی در بخش ۴ معرفی شده و ارزیابی رویکرد پیشنهادی در بخش ۵ انجام می‌شود. در نهایت بخش ۶ شامل نتیجه‌گیری است.

۲. مفاهیم پایه

به منظور درک جامع از مفاهیم کلیدی، در این

این مقاله تمرکز دارد و سپس مقالاتی با ویژگی‌های مشابه را یافته و به همان کاربر پیشنهاد می‌دهد. از آنجایی که هر یک از این روش‌ها مزایا و محدودیت‌هایی دارند، مانند مشکل شروع سرد یا کمبود داده، سیستم‌های ترکیبی با ترکیب دو روش قبلی، با هدف بهبود دقت و کارایی ایجاد شده‌اند. مدل ترکیبی این پژوهش از داده‌های تاریخی بار کاری و فراداده‌های گنجانه بهره می‌برد. هدف این مدل، ارائه تصمیمات هوشمند برای مقیاس‌بندی تطبیقی در محیط‌های ابری است. با این رویکرد، بهینه‌سازی منابع در گنجانه‌های پویا و پیچیده تسهیل می‌شود [۸-۱۰].



شکل ۱: تفاوت فیلترگذاری مبتنی بر محتوا با فیلترگذاری مشارکتی [۱۱]

۳-۲- سیستم‌های زمینه‌محور

سیستم‌های زمینه‌محور با در نظر گرفتن عوامل محیطی مانند زمان، موقعیت، و شرایط منابع، تصمیم‌گیری هوشمندانه‌تری ارائه می‌دهند. این سیستم‌ها در محیط‌های پویا مانند رایانش ابری و گنجانه‌ای، به بهینه‌سازی مقیاس‌بندی و تخصیص منابع کمک می‌کنند. با بهره‌گیری از استدلال معنایی، آن‌ها اقدامات مناسبی همچون افزایش یا کاهش مقیاس را بر اساس وضعیت فعلی پیشنهاد می‌دهند و باعث بهبود عملکرد و کاهش هزینه‌ها می‌شوند [۱۲]. هستی‌شناسی به چارچوبی ساخت‌یافته برای سازماندهی اطلاعات اشاره دارد که به سیستم‌ها امکان می‌دهد داده‌ها را به صورت معناداری پردازش کرده و روابط پیچیده میان موجودیت‌های مختلف را

بخش مقیاس‌بندی گنجانه‌ها، سیستم‌های توصیه‌کننده، سیستم‌های زمینه‌محور و هستی‌شناسی معرفی می‌شوند.

۱-۲- مقیاس‌بندی گنجانه‌ها

مقیاس‌بندی گنجانه‌ها به فرایند افزایش یا کاهش تعداد گنجانه‌ها در یک محیط ابری به منظور پاسخ‌گویی به تغییرات در بار کاری اشاره دارد. این عمل به صورت پویا و خودکار، منابع محاسباتی مورد نیاز را متناسب با میزان تقاضا تنظیم می‌کند تا از یک سو عملکرد سیستم حفظ شود و از سوی دیگر منابع بدون استفاده هدر نروند. مقیاس‌بندی می‌تواند به دو صورت انجام شود. در مقیاس‌بندی افقی عملیات افزودن یا حذف گنجانه‌ها انجام می‌شود و در مقیاس‌بندی عمودی، منابع اختصاص‌یافته به هر گنجانه مانند حافظه افزایش یا کاهش می‌یابد [۵-۷]. آنچه در این پژوهش مورد بررسی قرار گرفته است در مقیاس‌بندی افقی است.

۲-۲- سیستم‌های توصیه‌کننده

سیستم‌های توصیه‌کننده ابزارهایی هوشمند برای ارائه پیشنهادها، شخصی‌سازی شده بر پایه داده‌های کاربر و زمینه هستند. این سیستم‌ها به سه گروه اصلی تقسیم می‌شوند: فیلترسازی مشارکتی که با تحلیل رفتار کاربران مشابه، موارد مناسب را پیشنهاد می‌دهد. فیلترسازی مبتنی بر محتوا که بر اساس ویژگی‌های موارد و ترجیحات قبلی کاربر عمل می‌کند. در شکل ۱ تفاوت میان دو روش نشان داده شده است. در این شکل دو کاربر برخی مقالات مشترک را خوانده‌اند که نشان‌دهنده اشتراک در علایق است و الگوریتم فیلترسازی مشارکتی با شناسایی این شباهت، کاربری را که علایق مشابهی با دیگری دارد، شناسایی می‌کند. سپس محتوایی که یکی از آن‌ها خوانده ولی دیگری هنوز نخوانده، به عنوان پیشنهاد برای فرد دیگر ارائه می‌شود. الگوریتم فیلترسازی مبتنی بر محتوا بر اساس مقاله‌ای که کاربر مطالعه کرده است و تحلیل ویژگی‌های

می‌کند و شرایط عملیاتی، سناریوهای بارکاری و روابط متنی را در تصمیم‌گیری لحاظ می‌کند. این رویکرد، دقت تصمیمات را افزایش داده و امکان تحلیل پیچیده‌تری را برای سیستم فراهم می‌کند. سیستم‌های ترکیبی معمولاً با اعمال وزن‌های مختلف بر هر یک از این روش‌ها، تعادل مطلوبی را برای دستیابی به بهترین نتیجه ایجاد می‌کنند [۱۵، ۱۶].

۳- پژوهش‌های مرتبط

در این بخش چندین مطالعه مرتبط با مقیاس‌بندی گنجانه‌ها بررسی می‌شود. ابتدا روش‌های سنتی خودکارسازی که عمدتاً مبتنی بر قوانین ایستا هستند معرفی شده و سپس، سیستم‌های پیشنهاددهنده آگاه از زمینه مورد بررسی قرار می‌گیرند که با در نظر گرفتن عوامل محیطی و ویژگی‌های کاربران، تصمیم‌گیری‌های هوشمندانه‌تری ارائه می‌دهند. در نهایت، چارچوب‌های تصمیم‌گیری مبتنی بر هستی‌شناسی تحلیل می‌شوند که با تکیه بر دانش معنایی، قابلیت تفسیر و انطباق‌پذیری سیستم‌ها را در شرایط پویا افزایش می‌دهند. این فصل به دنبال نشان دادن نقاط قوت و چالش‌های هر رویکرد و فراهم‌سازی زمینه‌ای برای ارائه یک مدل ترکیبی بهینه است.

۳-۱- راه‌حل‌های سنتی مقیاس‌بندی خودکار

مقیاس‌بندی خودکار سنتی معمولاً بر پایه قوانین و آستانه‌هایی برای مصرف منابع مانند پردازنده و حافظه انجام می‌شود. این روش‌ها در ابزارها و چارچوب‌های ابری مانند کوبرنیتیز^۲ و ای‌دبلی‌واس^۳ به کار می‌روند و وقتی مصرف منابع از حد مشخصی فراتر رود، به‌طور خودکار مقیاس‌بندی انجام می‌شود. این رویکردها ساده، سریع و واکنش‌گرا هستند، اما در محیط‌های بسیار پویا با بار کاری غیرقابل پیش‌بینی و وابسته به عوامل خارجی

تحلیل‌کنند. در سیستم‌های توصیه‌گر، هستی‌شناسی‌ها با تعریف روابط معنایی بین داده‌ها، دقت و قابلیت تفسیر توصیه‌ها را افزایش می‌دهند. این روش امکان ثبت و تحلیل وابستگی‌های مفهومی را فراهم می‌کند، به‌گونه‌ای که سیستم بتواند درک عمیق‌تری از ساختار اطلاعات داشته باشد. برای مثال، در سیستم‌های مقیاس‌بندی گنجانه، یک هستی‌شناسی می‌تواند روابطی مانند «یک گنجانه دارای یک بارکاری مشخص است» یا «یک زمینه خاص موجب اجرای یک اقدام مقیاس‌بندی می‌شود» را تعریف کند. این تعریف روشن از روابط بین مفاهیم، نه تنها تصمیمات را بهبود می‌بخشد، بلکه باعث افزایش شفافیت فرایندهای تصمیم‌گیری نیز می‌شود. همچنین، هستی‌شناسی‌ها با ادغام اطلاعات زمینه‌ای مانند زمان، مکان و شرایط منابع، اطمینان حاصل می‌کنند که پیشنهادهای ارائه‌شده با شرایط واقعی مطابقت داشته و امکان استدلال منطقی بر اساس الگوهای از پیش تعریف‌شده را فراهم می‌کنند. به عنوان نمونه، یک هستی‌شناسی می‌تواند رابطه‌ای بین «استفاده بالای حافظه» و «نیاز به افزایش مقیاس» ایجاد کند و در صورت تحقق شرایط موردنظر، اقدام مناسب را فعال کند [۱۳]. سیستم‌های توصیه‌گر ترکیبی با ترکیب چندین روش توصیه‌گری، محدودیت‌های روش‌های فردی را برطرف می‌کنند. این سیستم‌ها معمولاً از ترکیب فیلترسازی مشارکتی، فیلترسازی مبتنی بر محتوا و استدلال مبتنی بر هستی‌شناسی استفاده می‌کنند تا توصیه‌هایی دقیق‌تر، قوی‌تر و مرتبط‌تر ارائه دهند. در فیلترسازی مبتنی بر همکاری در مقیاس‌بندی گنجانه‌ها، این روش از الگوهای تاریخی بارکاری برای شناسایی روندها و پیش‌بینی نیازهای منابع استفاده می‌کند. در فیلترسازی مبتنی بر محتوا در محیط‌های گنجانه‌ای، این روش فراداده‌هایی مانند استفاده از پردازنده، تخصیص حافظه و ترافیک شبکه را ارزیابی می‌کند تا تصمیمات مقیاس‌بندی بهینه‌ای اتخاذ شود. در کنار این دو روش، استدلال مبتنی بر هستی‌شناسی، یک لایه معنایی به فرآیند توصیه اضافه

مانند ترافیک، زمان روز یا وابستگی‌های سیستمی، دچار محدودیت می‌شوند.

۳،۱،۱- مقیاس‌بندی خودکار مبتنی بر قوانین

این روش با وجود سادگی، با چالش‌های متعددی مواجه است. این سیستم‌ها تنها پس از عبور از آستانه‌ها واکنش نشان می‌دهند، که منجر به تأخیر در مقیاس‌بندی و کاهش عملکرد می‌شود. بنابراین، امکان پیش‌بینی وجود ندارد. علاوه بر این تنظیم صحیح آستانه‌ها دشوار، زمان‌بر و نیازمند آزمون و خطا است و ممکن است موجب نوسانات مکرر در مقیاس‌بندی شود. آستانه‌های خیلی پایین منجر به بیش‌مقیاس‌بندی و افزایش هزینه می‌شوند. آستانه‌های خیلی بالا باعث کم‌مقیاس‌بندی و کاهش عملکرد یا خرابی سیستم می‌گردند. همچنین این روش‌ها شرایط بیرونی مانند ساعات اوج مصرف یا مشکلات شبکه را در نظر نمی‌گیرند.

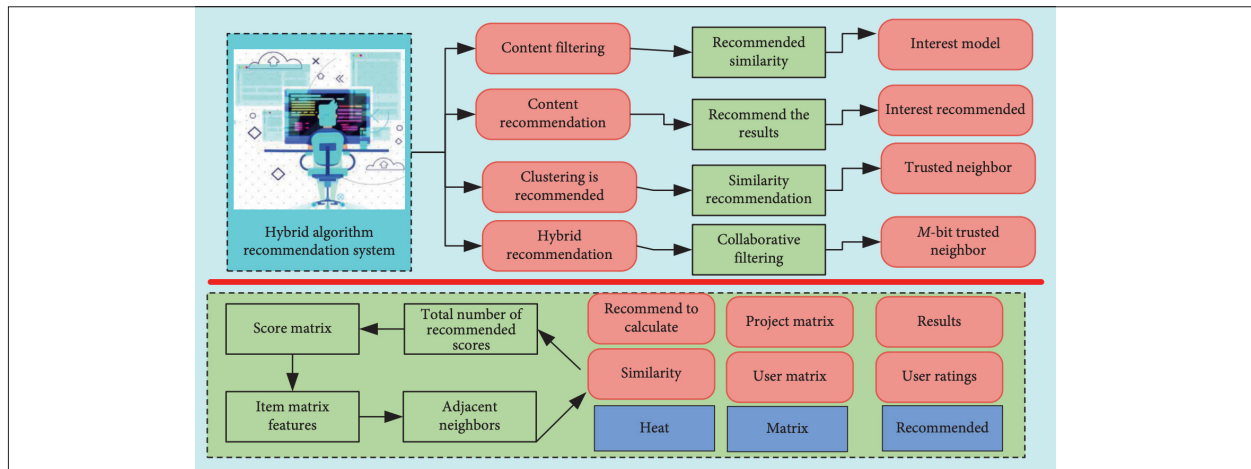
۳،۱،۲- روش‌های مبتنی بر یادگیری ماشین

برای غلبه بر محدودیت‌های مقیاس‌بندی سنتی، روش‌های مبتنی بر یادگیری ماشین معرفی شده‌اند. این روش‌ها با تحلیل داده‌های تاریخی و روندهای بار کاری، تصمیمات مقیاس‌بندی را به‌صورت هوشمندانه و داده‌محور اتخاذ می‌کنند. رایج‌ترین تکنیک‌ها شامل مدل‌های رگرسیون، سری‌های زمانی، درخت‌های تصمیم و شبکه‌های ال‌اس‌تی‌ام^۵ هستند که می‌توانند به‌صورت پیش‌بینانه منابع را تخصیص دهند و از بروز گلوگاه‌ها جلوگیری کنند. اگرچه مدل‌های یادگیری ماشین دقت پیش‌بینی را افزایش می‌دهند، اما با چالش‌هایی مانند هزینه بالای محاسباتی و نیاز به منابع پردازشی قدرتمند مواجهند که می‌تواند مقیاس‌پذیری را محدود کند. همچنین این مدل‌ها در مراحل ابتدایی اجرا، به دلیل نبود داده‌های تاریخی کافی، عملکرد ضعیفی دارند (مشکل شروع سرد). علاوه بر این، یکی از چالش‌های این رویکردها، فقدان آگاهی از زمینه

است؛ یعنی بسیاری از مدل‌های یادگیری ماشین هنوز نمی‌توانند وابستگی‌های بار کاری، اولویت‌های تجاری یا شرایط محیطی را به‌خوبی درک کنند و تصمیمات آن‌ها همچنان مبتنی بر معیارهای عددی صرف است [۲۱-۱۶].

۳-۲- مدل‌های ترکیبی

مدل‌های ترکیبی در سیستم‌های پیشنهاددهنده به دلیل توانایی در ترکیب چند تکنیک توصیه‌گری، باعث افزایش دقت پیش‌بینی و کاهش مشکلاتی مانند پراکندگی داده‌ها و شروع سرد شده‌اند. این مدل‌ها با ادغام روش‌هایی مانند فیلترسازی مشارکتی، فیلترسازی مبتنی بر محتوا و استدلال معنایی، عملکرد بهتری دارند. در مقاله [۸] با ترکیب یک روش خوشه‌بندی^۶ با فیلترسازی مشارکتی و محتوایی دقت بهبود، پراکندگی داده‌ها کاهش و مقیاس‌پذیری افزایش یافته که برای مدیریت منابع ابری بسیار مهم است. شکل ۱ معماری سیستم را نمایش می‌دهد که شامل دو بخش فرایند پیشنهاددهی و پردازش ماتریسی و محاسبات است. در بخش اول فیلترسازی محتوا، پیشنهاد محتوایی و خوشه‌بندی انجام می‌شود و در بخش دوم ماتریس امتیازدهی و شباهت و همسایگان مجاور محاسبه می‌شود. در مقاله [۱۱] یک سیستم پیشنهاددهی ترکیبی وزنی توسعه داده شده که با ترکیب فیلترسازی مبتنی بر محبوبیت، مشارکتی و محتوایی، دقت و مقیاس‌پذیری را بهبود می‌بخشد. نتایج نشان داد که استفاده از سازوکارهای وزن‌دهی پویا موجب افزایش سازگاری سیستم شده و این ویژگی در مقیاس‌بندی منابع ابری بسیار مؤثر است. استفاده از مدل‌های پیشنهاددهی ترکیبی برای بهینه‌سازی منابع در محیط‌های ابری در حال گسترش است. پژوهش‌ها نشان داده‌اند که ادغام پیش‌بینی بار کاری، فیلترسازی مشارکتی و آگاهی از زمینه می‌تواند عملکرد مقیاس‌بندی خودکار را بهبود دهد. استفاده از الگوهای زمانی و توزیع جغرافیایی کاربران به بهبود تخصیص منابع در محیط‌های پویا کمک می‌کند.



شکل ۱: معماری سیستم مقاله [۸].

است. همچنین عوامل محیطی مانند ساعات اوج مصرف، توزیع جغرافیایی بار و محدودیت‌های سخت‌افزاری در تصمیم‌گیری‌ها در نظر گرفته می‌شود و چندین تکنیک مربوط به پیشنهاددهی برای غلبه بر مشکلاتی مانند شروع سرد و پراکندگی داده‌ها استفاده شده است. سیستم پیشنهادی با طراحی مولفه‌ای ایجاد شده تا مقیاس‌پذیری، انعطاف‌پذیری و نگهداری‌پذیری را تضمین کند و شامل چهار مؤلفه اصلی است که در شکل ۲ نشان داده شده است.

۱. مولفه آماده‌سازی داده: تولید فراداده گنجانده‌ها و شبیه‌سازی تغییرات بار کاری با ذخیره داده‌ها در قالب ساختاریافته.

۲. مولفه توصیه‌گر ترکیبی: ترکیب فیلترسازی مبتنی بر محتوا، فیلترسازی مشارکتی و استدلال معنایی مبتنی بر هستی‌شناسی برای ارائه پیشنهاد‌های هوشمند.

۳. مولفه تصمیم‌گیری مقیاس‌بندی: تحلیل امتیازها و داده‌های متنی برای اتخاذ تصمیم مناسب در خصوص افزایش، کاهش یا حفظ وضعیت مقیاس.

۴. مولفه ارزیابی: ذخیره نتایج مقیاس‌بندی، تحلیل معیارهای عملکرد و ارزیابی دقت تصمیمات.

همچنین، در مطالعه [۱۸] نشان داده شده که ترکیب تحلیل احساسات با پیشنهاددهی ترکیبی، تصمیمات مقیاس‌بندی تطبیقی را دقیق‌تر و مؤثرتر می‌کند. به گفته گلوزر و همکاران [۲۲] مدل‌سازی هستی‌شناسی نقش کلیدی در افزایش قابلیت تفسیرپذیری و انطباق‌پذیری سیستم‌های پیشنهاددهنده دارد. دلیل این اهمیت، توانایی هستی‌شناسی در سازماندهی دانش حوزه به شکل قابل فهم برای ماشین‌ها است. مطالعات آنها نشان داده‌اند که استفاده از نمایش دانش معنایی و استدلال مبتنی بر معنا می‌تواند به‌ویژه در محیط‌های پویا و در حال تغییر، دقت تصمیم‌گیری را افزایش دهد. در حوزه مقیاس‌بندی گنجانده‌ها نیز، استدلال هستی‌شناختی با شناسایی روابط میان عوامل کلیدی مانند الگوهای مصرف منابع، دوره‌های اوج بار، و محدودیت‌های جغرافیایی، آگاهی از زمینه را در فرایند تصمیم‌گیری بهبود می‌دهد.

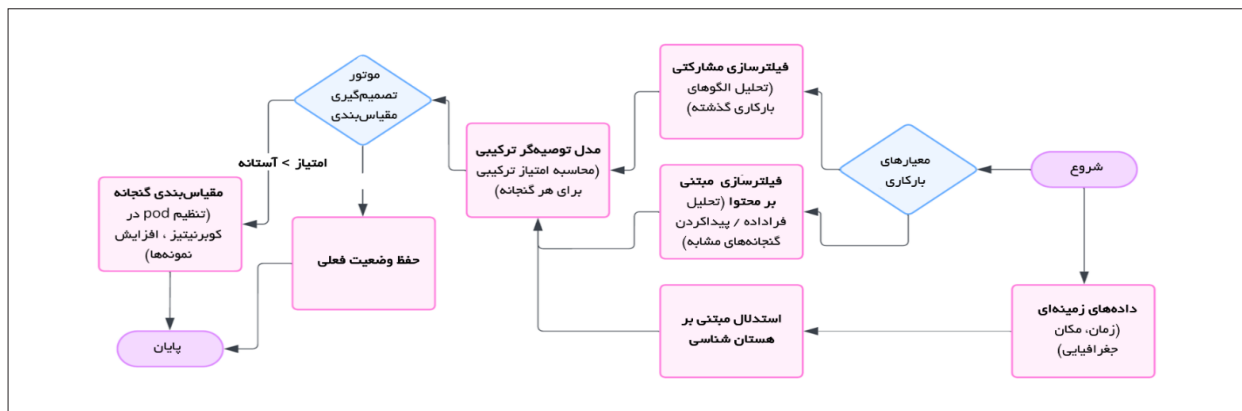
جدول (۱) رویکردهای معرفی شده را از نظر بهبود قابلیت پیش‌بینی، آگاهی از زمینه، انطباق‌پذیری، مقیاس‌پذیری، کارایی در استفاده از منابع، تصمیم‌گیری بیدرنگ و استدلال مبتنی بر هستی‌شناسی بررسی می‌کند.

۴- روش پیشنهادی

در سیستم پیشنهادی هدف مقیاس‌بندی پویا بر اساس تحلیل شرایط لحظه‌ای و الگوهای تاریخی بار کاری

۴.۱- مدل‌سازی مسئله

این پژوهش فرض می‌کند که یک سیستم توصیه‌گر



شکل ۲: مؤلفه‌های سیستم پیشنهادی

مشارکتی و $C_{context}$ میزان تأثیر عامل‌های متنی را نشان می‌دهد. α, β, γ وزن‌هایی هستند که بر اساس اهمیت نسبی هر عامل تنظیم می‌شوند. همه مقادیر در بازه ۰ تا ۱ قرار می‌گیرند.

$$S_{final} = \alpha S_{cbf} + \beta S_{cf} + \gamma C_{context} \quad (۱-۴)$$

بنابراین، سیستم پیشنهادی از سه تکنیک کلیدی فیلترسازی مبتنی بر محتوا برای تحلیل ویژگی‌های فردی گنجانه‌ها، فیلترسازی مشارکتی برای یادگیری از الگوهای تاریخی بار کاری و استدلال مبتنی بر هستی‌شناسی برای گنجاندن آگاهی متنی در تصمیمات مقیاس‌بندی استفاده می‌کند. با ترکیب این تکنیک‌ها که در ادامه شرح داده می‌شوند، سیستم بر محدودیت‌هایی مانند مشکل شروع سرد، پراکندگی داده‌ها و آستانه‌های ایستا در مقیاس‌بندی غلبه کرده و پیشنهادات مقیاس‌بندی دقیق‌تر و تطبیق‌پذیرتری ارائه می‌دهد.

۴.۱.۱. مدل پیشنهاددهنده ترکیبی

در ادامه سه بخش اصلی مدل پیشنهاددهنده ترکیبی در این پژوهش معرفی می‌شود.

الف) فیلترسازی مبتنی بر محتوا

فیلتر مبتنی بر محتوا پیشنهادات را بر اساس ویژگی‌های ذاتی گنجانه‌ها ارائه می‌دهد. هر گنجانه به عنوان یک بردار

جدول ۱: مقایسه کلی رویکردهای مقیاس‌بندی

ویژگی / رویکرد	مقیاس‌بندی مبتنی بر قوانین	مقیاس‌بندی مبتنی بر یادگیری ماشین	مدل‌های پیشنهاد دهنده ترکیبی	روش پیشنهادی
قابلیت پیش‌بینی	ندارد	دارد	دارد	دارد
آگاهی از زمینه	ندارد	محدود	دارد	دارد
انطباق‌پذیری	محدود	متوسط	دارد	دارد
مقیاس‌پذیری	متوسط	دارد	دارد	دارد
کارایی در استفاده از منابع	کم	کم	دارد	دارد
تصمیم‌گیری بیدرنگ	دارد	وابسته به پیچیدگی	متوسط	متوسط
استدلال مبتنی بر هستی‌شناسی	ندارد	ندارد	ندارد	دارد

ترکیبی که فیلترسازی مبتنی بر محتوا، فیلترسازی مشارکتی و استدلال مبتنی بر هستی‌شناسی را ادغام می‌کند، می‌تواند کارایی مقیاس‌بندی گنجانه‌ها را به‌طور قابل‌توجهی افزایش دهد. این هدف از طریق پیش‌بینی تقاضای منابع با استفاده از داده‌های تاریخی بار کاری و ویژگی‌های گنجانه‌ها، به‌کارگیری استدلال متنی بهبود دقت و مرتبط بودن تصمیمات مقیاس‌بندی و سازگاری پویا با شرایط بیدرنگ برای کاهش ناکارآمدی‌های ناشی از روش‌های مقیاس‌بندی ایستا محقق می‌شود.

فرضیه به صورت ریاضی با فرمول (۱-۴) تعریف می‌شود. که در آن S_{final} امتیاز نهایی مقیاس‌بندی، S_{cbf} امتیاز فیلترسازی مبتنی بر محتوا، S_{cf} امتیاز فیلترسازی

پنهان است که هر سطر آن الگوهای پنهان در نمونه‌های داده و V ماتریس ویژگی پنهان دیگر است که هر سطر آن نشان‌دهنده الگوهای پنهان در ویژگی‌های بارکاری است. ماتریس‌های ویژگی‌های پنهان و Σ ماتریس قطری شامل مقادیر منفرد و نشان‌دهنده اهمیت هر الگوی پنهان است.

$$M = U \Sigma V^T \quad (۳-۴)$$

فرمول پیش‌بینی CF به‌صورت زیر است که در آن k تعداد ویژگی‌های پنهان انتخاب شده است. Σ_k اهمیت هر الگو را نشان می‌دهد و مقادیر بزرگ‌تر به معنای الگوی مؤثرتر هستند.

$$M^{\wedge} = U \Sigma_k V^T \quad (۴-۴)$$

شبه کد برای محاسبه شباهت فلیترسازی مبتنی بر محتوا و مشارکتی به ترتیب در شکل ۳ نشان داده شده است.

Algorithm Name: Compute Similarity Matrix for Containers
Inputs: container_dataset (with resource usage features)
Outputs: similarity_matrix (for all containers)
BEGIN-PROCEDURE
 ForEach container A in container_dataset
 ForEach container B in container_dataset
 similarity_matrix[A][B] = compute_cosine_similarity(A, B)
 Return similarity_matrix
END-PROCEDURE

الف) محاسبه شباهت فلیترسازی مبتنی بر محتوا

Algorithm Name: Predict Scaling Scores using SVD
Inputs: M (Historical workload matrix)
Outputs: M_pred (Predicted scaling scores for containers)
BEGIN-PROCEDURE
 Perform SVD on M to obtain U, Σ , and $V^{\wedge T}$
 Reduce dimensions by selecting top-k singular values
 Reconstruct workload matrix: $M_pred = U \Sigma_k V^{\wedge T}$
 Return M_pred
END-PROCEDURE

ب) پیش‌بینی بار کاری مبتنی بر فلیترسازی مشارکتی

شکل ۳: شبه کد فلیترسازی مبتنی بر محتوا و مشارکتی

پ) استدلال متنی مبتنی بر هستی‌شناسی
 اجزای هستی‌شناسی شامل کلاس‌ها (گنجانه، متن و نوع اقدام مقیاس‌بندی) و روابط (مانند دارای متن بودن (گنجانه ← متن) و فعال کردن اقدام (متن اقدام ← مقیاس‌بندی)

ویژگی شامل مقادیر مصرف منابع مانند واحد پردازش مرکزی، حافظه و ترافیک شبکه نمایش داده می‌شود. شباهت بین گنجانه‌ها با استفاده از شباهت کسینوسی محاسبه می‌شود و توسط فرمول (۲-۴) تعریف شده است. شباهت کسینوسی معیاری برای سنجش میزان تشابه بین دو بردار (مثلاً در فضای چندبُعدی) است که با محاسبه کسینوس زاویه بین آنها تعیین می‌شود. که در آن A و B بردارهای ویژگی مربوط به دو گنجانه هستند و صورت کسر حاصلضرب داخلی دو بردار و مخرج کسر حاصلضرب بزرگی بردارها است. این روش زمانی کارآمد است که متاداده کافی برای گنجانه‌ها در دسترس باشد، اما در مواردی که اطلاعات تاریخی بار کاری محدود است، عملکرد آن ضعیف‌تر خواهد بود.

$$Similarity(A, B) = \frac{A \cdot B}{||A|| \times ||B||} \quad (۲-۴)$$

ب) فلیترسازی مشارکتی

برخلاف فیلتر مبنی بر محتوا که به ویژگی‌های گنجانه متکی است، این فیلتر از الگوهای تاریخی بار کاری برای پیش‌بینی نیازهای مقیاس‌بندی آینده استفاده می‌کند. این روش فرض می‌کند که گنجانه‌های مشابه رفتارهای مقیاس‌بندی مشابهی خواهند داشت. در این پژوهش، تجزیه ماتریس با استفاده از تجزیه مقدار منفرد کوتاه‌شده برای استخراج الگوهای پنهان از داده‌های تاریخی به کار گرفته شده است. تجزیه مقدار منفرد یک روش جبر خطی برای تجزیه ماتریس‌ها به عوامل ساده‌تر است که در کاهش ابعاد و استخراج الگوهای پنهان از داده‌های تاریخی کاربرد دارد. با در نظر گرفتن یک ماتریس کاربر-مورد که در آن ردیف‌ها نمایانگر گنجانه‌ها و ستون‌ها نمایانگر ویژگی‌های بار کاری هستند. تجزیه مقدار منفرد کوتاه شده و توسط فرمول (۳-۴) تعریف می‌شود که در آن M ماتریس بار کاری که در آن سطر در آن نمونه‌های داده و ستون ویژگی‌های بارکاری است، U ماتریس ویژگی‌های

تجمع کرده و تصمیم نهایی مقیاس‌بندی را تولید می‌کند. روش توصیه نهایی بر یک مدل آستانه‌ای متکی است، که در آن S_{final} مجموع وزن‌دار تمامی امتیازات محاسبه شده است.

$$(6-4) \quad \begin{cases} \text{Scale up} & S_{final} > 0.7 \\ \text{No Scaling} & 0.3 \leq S_{final} \leq 0.7 \\ \text{Scale Down} & S_{final} < 0.3 \end{cases}$$

۵- ارزیابی روش پیشنهادی

برای پیاده‌سازی مدل توصیه‌گر ترکیبی، پایتون^۶ به عنوان زبان برنامه‌نویسی اصلی به دلیل سادگی و اکوسیستم گسترده کتابخانه‌های آن انتخاب شده است. برای شبیه‌سازی بار کاری و مدیریت وظایف غیر هم‌زمان، از سلری^۷ همراه با ردیس^۸ استفاده شده است. پردازش داده‌ها و انجام عملیات عددی توسط پانداس^۹ و نام‌پای^{۱۰} انجام می‌شود، در حالی که اسکیت‌لرن^{۱۱} برای پیاده‌سازی فیلترگذاری مشارکتی با استفاده از تجزیه ماتریس به کار رفته است. جهت تعریف و استدلال مبتنی بر هستی‌شناسی، از آل‌ردی^{۱۲} بهره گرفته شده و برای تجسم داده‌ها و تحلیل عملکرد سیستم، مت‌پلات^{۱۳} مورد استفاده قرار گرفته است.

۵.۱- تولید جریان کاری

پیاده‌سازی سیستم پیشنهادی بر اساس یک جریان کاری ساختار یافته انجام شده است که شامل پنج مرحله اصلی است. شبه کد این پنج مرحله در شکل ۵ آورده شده است.

۱. ابتدا تولید متاداده گنجانه صورت می‌گیرد، که طی آن داده‌های مصنوعی برای شبیه‌سازی پیکربندی‌های متنوع گنجانه‌ها تولید می‌شوند. ویژگی‌هایی مانند

است. برای مثال، اگر یک گنجانه در ساعات اوج مصرف استفاده بالایی از واحد پردازش مرکزی داشته باشد، هستی‌شناسی آن را به متن «ساعات اوج مصرف» متصل کرده و اقدام «افزایش مقیاس» را فعال می‌کند. شبه‌کد یک قانون در هستی‌شناسی در شکل ۴ نشان داده شده است.

Algorithm Name: Context-Aware Scaling Recommendation
Inputs: Container workload, contextual factors
Outputs: Scaling recommendation
BEGIN-PROCEDURE
 If container.Workload > 80% and container.Context == "Peak_Hours"
 Recommend "Scale Up"
END-PROCEDURE

شکل ۴: شبه کد از یک قانون در هستی‌شناسی

ت) مقیاس‌بندی آگاه از متن

روش‌های سنتی مقیاس‌بندی فقط معیارهای عددی را در نظر می‌گیرند، در حالی که مقیاس‌بندی آگاه از متن، عوامل خارجی را نیز در فرآیند تصمیم‌گیری ادغام می‌کند. این امر راهبردهای مقیاس‌بندی تطبیقی را تضمین می‌کند که بازدهی منابع را بهبود می‌بخشند. سیستم، عوامل زمینه‌ای مانند ساعات شبانه‌روز که در آن راهبردهای مقیاس‌بندی در ساعات اوج و غیر اوج مصرف به‌طور پویا تنظیم می‌شوند، توازن بار جغرافیایی که در آن گنجانه‌های اجرا شده در مناطق مختلف ممکن است نیاز به اقدامات مقیاس‌بندی متفاوتی داشته باشند و در نهایت محدودیت‌های منابع که در آن گنجانه‌هایی که محدودیت‌های بالای حافظه یا واحد پردازش مرکزی دارند، در اولویت افزایش مقیاس قرار می‌گیرند را در تصمیمات مقیاس‌بندی در نظر می‌گیرد. محاسبه تصمیم‌گیری مقیاس‌بندی با استفاده از فرمول (۴-۵) تعریف شده است که در آن $S_{context}$ امتیاز نهایی مقیاس‌بندی و w_1, w_2, w_3 وزن‌های هر روش هستند.

(۴-۵)

$$S_{context} = w_1(CBF) + w_2(CF) + w_3(Ontology)$$

ث) نهایی‌سازی تصمیمات مقیاس‌بندی

پس از انجام محاسبات ذکر شده سیستم امتیازات را

6-Python
 7-Celery
 8-Redis
 9-Pandas
 10-NumPy
 11-Scikit-learn
 12-Owllready2
 13-Matplotlib

به عنوان یک لایه مفهومی اضافه می شود که روابط بین گنجانه ها، زمینه های اجرایی (مانند ساعات اوج مصرف یا کمبود منابع) و اقدامات مقیاس بندی را تعریف می کند. به عنوان مثال، اگر یک گنجانه در ساعات اوج مصرف دارای ۹۰ درصد استفاده از واحد پردازش مرکزی باشد، مدل هستی شناسی اقدام «افزایش مقیاس» را پیشنهاد می دهد. در حالی که مدل پیشنهاددهنده ترکیبی بر داده های عددی برای تصمیم گیری تکیه دارد، استدلال مبتنی بر هستی شناسی، عوامل مفهومی را در نظر می گیرد تا تضمین کند که تصمیمات مقیاس بندی با اولویت های عملیاتی همسو هستند. شبه کد (ت) در شکل ۵ مربوط به این مرحله است.

۵. مرحله نهایی، تولید تصمیمات مقیاس بندی است که در آن امتیازات ترکیبی محاسبه شده و استدلال هستی شناسی با یکدیگر ترکیب شده و توصیه های نهایی برای مقیاس بندی تولید می شوند. در این مرحله، اگر امتیاز ترکیبی بالاتر از یک آستانه مشخص (مثلاً ۰/۷) باشد، گنجانه برای مقیاس بندی علامت گذاری می شود. در غیر این صورت، هیچ اقدام مقیاس بندی انجام نمی شود و وضعیت فعلی حفظ می گردد. شبه کد (ث) در شکل ۵ مربوط به این مرحله است.

۵.۲- نتیجه آزمایش

سیستم بر روی یک مجموعه داده شبیه سازی شده شامل پنج گنجانه در طی ده تکرار آزمایش شد. در هر تکرار، بارهای کاری مصنوعی تولید شده، امتیازات ترکیبی محاسبه گردید و استدلال زمینه ای برای تعیین نیاز به اقدامات مقیاس بندی اعمال شد. نتایج در قالب ساختاریافته ی JSON ذخیره شدند تا امکان تجزیه و تحلیل بیشتر فراهم شود. یکی از جنبه های کلیدی ارزیابی، بررسی تعداد دفعاتی که اقدامات مقیاس بندی فعال شدند، میزان دقت مدل ترکیبی در شناسایی نیاز به مقیاس بندی، و تأثیر استدلال زمینه ای بر فرایند تصمیم گیری بود.

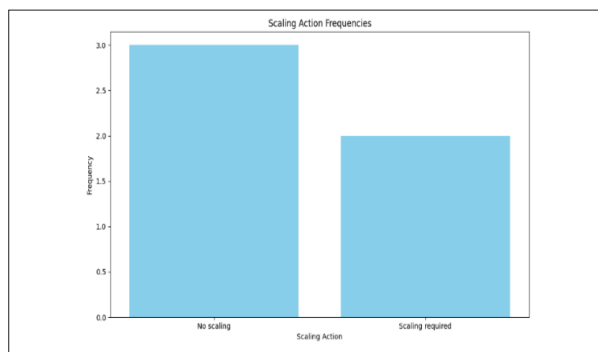
مصرف واحد پردازش مرکزی، حافظه و پهنای باند شبکه تعیین شده و در قالب یک فایل JSON ذخیره می شوند. هر گنجانه دارای سه ویژگی اصلی برای نشان دادن میزان مصرف منابع خود است. استفاده از واحد پردازش مرکزی که به صورت درصدی (۱۰ درصد تا ۱۰۰ درصد) تعریف می شود و میانگین مصرف پردازنده گنجانه را مشخص می کند. حافظه مصرفی که در محدوده ۵۰ مگابایت تا ۵۰۰ مگابایت متغیر است، تغییرات تخصیص حافظه را شبیه سازی می کند. در نهایت، میزان استفاده از شبکه که برحسب ۵ تا ۵۰ مگابیت بر ثانیه اندازه گیری می شود و نشان دهنده ترافیک شبکه گنجانه است. مصرف بالای پهنای باند می تواند نشان دهنده نیاز به نمونه های اضافی برای متعادل سازی بار شبکه باشد. شبه کد (الف) در شکل ۵ مربوط به این مرحله است.

۲. شبیه سازی بارکاری انجام می شود که بارهای کاری پویا برای بازتاب نوسانات واقعی در مصرف منابع تولید می شوند. این فرایند با استفاده از سلری برای مدیریت وظایف توزیع شده و ردیس به عنوان واسط ذخیره سازی انجام می شود. با ایجاد تغییرپذیری در الگوهای بارکاری، سیستم تضمین می کند که راهبردهای مقیاس بندی تحت شرایط واقعی مورد آزمایش قرار گیرند و عواملی مانند رفتار کاربران، افزایش ناگهانی ترافیک و شدت محاسباتی وظایف را شبیه سازی کند. شبه کد (ب) در شکل ۵ مربوط به این مرحله است.

۳. مدل توصیه گر ترکیبی اجرا می شود که ترکیبی از فیلترگذاری مبتنی بر محتوا و فیلترگذاری مشارکتی را برای ارزیابی شباهت گنجانه ها و پیش بینی نیازهای مقیاس بندی بر اساس الگوهای تاریخی بارکاری به کار می گیرد. شبه کد (پ) در شکل ۵ مربوط به این مرحله است.

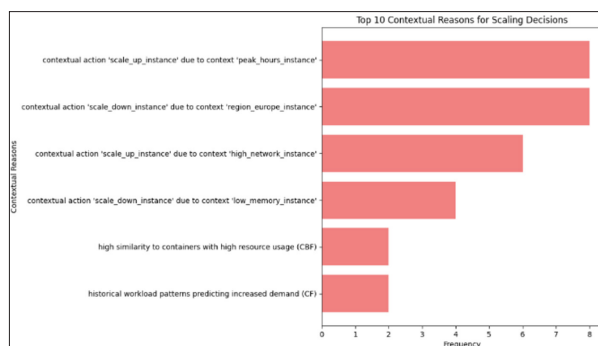
۴. سپس، استدلال مفهومی مبتنی بر هستی شناسی

نمودار شکل ۶ فراوانی تصمیمات مقیاس‌بندی را در طی چندین تکرار نمایش می‌دهد. این نمودار تصمیمات را به دو دسته «نیاز به مقیاس‌بندی» و «عدم نیاز به مقیاس‌بندی» طبقه‌بندی می‌کند، که تعادل بین مقیاس‌بندی پیشگیرانه و محافظه‌کارانه را نشان می‌دهد. این نمودار نشان می‌دهد که توصیه‌های مقیاس‌بندی بیش از حد انجام نشده‌اند، بلکه فقط در شرایطی که الگوهای بار کاری و عوامل زمینه‌ای آن را توجیه کرده‌اند، فعال شده‌اند.



شکل ۶: فراوانی اقدامات مقیاس‌بندی.

نمودار دوم شکل ۷ رایج‌ترین عوامل زمینه‌ای مؤثر بر تصمیمات مقیاس‌بندی را نمایش می‌دهد. این نمودار نشان می‌دهد که ساعات اوج مصرف، استفاده بالای شبکه و محدودیت‌های جغرافیایی از جمله تأثیرگذارترین عوامل زمینه‌ای در اتخاذ تصمیمات مقیاس‌بندی هستند.



شکل ۷: ده عامل برتر زمینه‌ای در تصمیمات مقیاس‌بندی.

این تحلیل نشان می‌دهد که تصمیمات مقیاس‌بندی تنها

Algorithm Name: Generate Container Data
Inputs: num_containers
Outputs: container_data (dictionary with resource usage attributes)
BEGIN-PROCEDURE
 DEFINE num_containers
 INITIALIZE container_data as an empty dictionary
 FOR i **FROM** 1 **TO** num_containers:
 container = {
 "CPU": RANDOM_INT(10, 100)
 "Memory": RANDOM_INT(50, 500)
 "Network": RANDOM_INT(5, 50)
 }
 ADD container **TO** container_data
 RETURN container_data
END-PROCEDURE

الف) تولید فراداده‌ی گنجانه‌ها.

Algorithm Name: Generate and Store Workload Data
Inputs: Container list
Outputs: Workload data stored in JSON file
BEGIN-PROCEDURE
 INITIALIZE Celery **WITH** Redis as broker
 DEFINE FUNCTION to generate random workload (50% - 100%)
 FOR EACH container:
 - Assign a workload value
 - Store results asynchronously **USING** Celery
 SAVE workload data **TO** JSON file
END-PROCEDURE

ب) شبیه‌سازی بار کاری.

Algorithm Name: Generate and Store Workload Data
Inputs: Container list
Outputs: Workload data stored in JSON file
BEGIN-PROCEDURE
 INITIALIZE Celery **WITH** Redis as broker
 DEFINE FUNCTION to generate random workload (50% - 100%)
 FOR EACH container:
 - Assign a workload value
 - Store results asynchronously **USING** Celery
 SAVE workload data **TO** JSON file
END-PROCEDURE

پ) محاسبه امتیاز ترکیبی

Algorithm Name: Generate and Store Workload Data
Inputs: Container list
Outputs: Workload data stored in JSON file
BEGIN-PROCEDURE
 INITIALIZE Celery **WITH** Redis as broker
 DEFINE FUNCTION to generate random workload (50% - 100%)
 FOR EACH container:
 - Assign a workload value
 - Store results asynchronously **USING** Celery
 SAVE workload data **TO** JSON file
END-PROCEDURE

ت) استدلال مبتنی بر هستان‌شناسی

Algorithm Name: Generate and Store Workload Data
Inputs: Container list
Outputs: Workload data stored in JSON file
BEGIN-PROCEDURE
 INITIALIZE Celery **WITH** Redis as broker
 DEFINE FUNCTION to generate random workload (50% - 100%)
 FOR EACH container:
 - Assign a workload value
 - Store results asynchronously **USING** Celery
 SAVE workload data **TO** JSON file
END-PROCEDURE

ث) تولید تصمیم مقیاس‌گذاری

شکل ۵: شبه کدهای مربوط به فرآیند تولید جریان کاری

جدول ۲: مقایسه دقت

معیار	مدل ترکیبی با هستی‌شناسی	مقیاس‌بندی با قوانین ثابت
مثبت اشتباه (مقیاس‌بندی غیرضروری)	۵ درصد	۱۸ درصد
منفی اشتباه (عدم تشخیص نیاز به مقایسه‌گذاری)	۷ درصد	۲۲ درصد
دقت کلی تصمیم‌گیری	۸۸ درصد	۶۹ درصد

مدل ترکیبی تعداد اقدامات مقیاس‌بندی غیرضروری را کاهش داده و در عین حال اطمینان حاصل کرده است که گنجانه‌هایی که به مقیاس‌بندی نیاز دارند، نادیده گرفته نمی‌شوند. این نتایج نشان‌دهنده برتری تحلیل تاریخچه بار کاری و استدلال مفهومی بر روش‌های مقیاس‌بندی مبتنی بر آستانه‌های ثابت است.

۶- نتیجه‌گیری

مقیاس‌بندی کارآمد گنجانه‌ها یکی از چالش‌های اساسی در رایانش ابری مدرن است، جایی که بارهای کاری متغیر نیازمند مدیریت منابع پویا و تطبیق‌پذیر هستند. روش‌های سنتی مقیاس‌بندی خودکار مبتنی بر قوانین اغلب قادر به در نظر گرفتن الگوهای تاریخی بار کاری، نوسانات بیدرنگ و عوامل زمینه‌ای نیستند. این محدودیت‌ها منجر به تخصیص غیربهبوده منابع، افزایش هزینه‌ها و افت عملکرد سیستم می‌شود. برای غلبه بر این چالش‌ها، در این پژوهش سیستمی آگاه از زمینه که از مدل توصیه‌گر ترکیبی استفاده می‌کند پیشنهاد شده است. سیستم پیشنهادی با ترکیب فیلترگذاری مشارکتی، فیلترگذاری مبتنی بر محتوا و استدلال مبتنی بر هستی‌شناسی سعی دارد در تصمیمات مقیاس‌بندی در محیط‌های گنجانه‌ای بهبود ایجاد کند. نتایج تجربی نشان می‌دهند که سیستم پیشنهادی در مقایسه با سیستم‌های سنتی در دقت ۱۹ درصد بهبود ایجاد کرده است. پیشنهاد می‌شود برای بهبود این پژوهش از مدل‌های یادگیری ماشین مولد برای پیش‌بینی دقیق‌تر بار کاری استفاده شود. همچنین، هستی‌شناسی می‌تواند به‌صورت

بر اساس مقادیر بار کاری اتخاذ نشده‌اند، بلکه به طور پویا بر اساس شرایط محیطی نیز تنظیم شده‌اند.

۵,۳- مقایسه با روش مقیاس‌بندی مبتنی بر قوانین ثابت
برای ارزیابی کارایی سیستم، معیارهای کلیدی عملکرد از جمله زمان تأخیر، زمان اجرای تصمیمات مقیاس‌بندی و دقت در تصمیم‌گیری‌های مقیاس‌بندی مورد تجزیه و تحلیل قرار گرفتند. زمان مورد نیاز برای تولید توصیه‌های مقیاس‌بندی در چندین اجرای سیستم اندازه‌گیری و نتایج در جدول (۲) گزارش شده است. استفاده از استدلال مفهومی مبتنی بر هستی‌شناسی مقداری سربار پردازشی ایجاد کرد، زیرا سیستم برای استخراج داده‌های زمینه‌ای مرتبط، پرس‌وجوهای اضافی اجرا می‌کند. با این حال، تأخیر در تصمیم‌گیری با استفاده از مدل ترکیبی در محدوده قابل قبول باقی ماند.

جدول ۲: زمان مورد نیاز برای تولید توصیه

معیار	مدل ترکیبی با هستی‌شناسی	مقیاس‌بندی با قوانین ثابت
زمان اجرا در هر تکرار	۴۵۰ میلی ثانیه	۲۷۰ میلی ثانیه
میانگین تأخیر در تصمیم‌گیری	۱۵ میلی ثانیه	۸ میلی ثانیه

اگرچه مدل ترکیبی باعث افزایش اندک در تأخیر تصمیم‌گیری شد، اما این افزایش با بهبود دقت توصیه‌های مقیاس‌بندی (که در جدول بعدی گزارش شده است) جبران شده است، که منجر به کاهش موارد مقیاس‌بندی نادرست می‌شود. دقت در تصمیم‌گیری‌های مقیاس‌بندی با مقایسه مدل توصیه‌گر ترکیبی و روش سنتی مبتنی بر آستانه‌های ثابت ارزیابی و نتایج در جدول (۳) گزارش شده است. روش سنتی از آستانه‌های ثابت برای پردازنده و حافظه هزینه پردازشی بیش از ۸۰ درصد است، در حالی که مدل ترکیبی آستانه‌های تصمیم‌گیری را به‌صورت پویا بر اساس تاریخچه بار کاری و عوامل زمینه‌ای تنظیم می‌کند.

tion platforms,” in Proceedings of the 7th South East European Doctoral Student Conference (DSC 2012), pp. 539-550.

[15] Pashdar, A., Lee, Y. C., Hassanzadeh, T. & Almi'ani, K. 2021. “Resource recommender for cloud-edge engineering,” Information, vol. 12, no. 6, p. 224.

[16] Kirchoff, D. F., Xavier, M., Mastella, J. & De Rose, C. A. 2019. “A preliminary study of machine learning workload prediction techniques for cloud applications,” in Proceedings of the 2019 27th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), IEEE, pp. 222-227.

[17] Ahamed, Z., Khemakhem, M., Eassa, F., Alsolami, F. & Al-Ghamdi, A. S. A. M. 2023. “Technical study of deep learning in cloud computing for accurate workload prediction,” Electronics, vol. 12, no. 3, p. 650.

[18] Xu, M., Song, C., Wu, H., Gill, S. S., Ye, K. & Xu, C. 2022. “esDNN: deep neural network based multivariate workload prediction in cloud computing environments,” ACM Transactions on Internet Technology (TOIT), vol. 22, no. 3, pp. 1-24.

[19] A. Soren and A. Jaiden, “Efficient Resource Reservation and Allocation in Cloud Using Energy-Aware AI Models,” doi:10.13140/RG.2.2.12117.44006, 2024.

[20] D. Alsadie and M. Alsulami, “Enhancing workflow efficiency with a modified Firefly Algorithm for hybrid cloud edge environments,” Scientific Reports, vol. 14, no. 1, p. 24675, 2024.

[21] S. Karimunnisa and Y. Pachipala, “Deep Learning-Driven Workload Prediction and Optimization for Load Balancing in Cloud Computing Environment,” Cybernetics and Information Technologies, vol. 24, no. 3, 2024.

[22] Z. Gulzar, A. A. Leema, and G. Deepak, “Pcrrs: Personalized course recommender system based on hybrid approach,” Procedia Computer Science, vol. 125, pp. 518-524, 2018.

پویا و خودکار تنظیم گردیده و اجرای موازی برای بهبود مقیاس‌پذیری در محیط‌های بزرگ پیاده‌سازی شود.

مراجع

[1] Alipour, H., Liu, Y., & Hamou- Lhadj, A. 2014. “Analyzing auto-scaling issues in cloud environments,” in Proceedings of the CASCION, vol. 14, pp. 75-89.

[2] Verma, G. & Yadav, S. ??? “Scalable Optimization Algorithm for Cloud Resource Scaling,” in Proceedings of the Advanced Computing Techniques for Optimization in Cloud: Chapman and Hall/CRC, pp. 180-197.

[3] Kumar, P. & Kumar, R. 2019. “Issues and challenges of load balancing techniques in cloud computing: A survey,” ACM Computing Surveys (CSUR), vol. 51, no. 6, pp. 1-35

[4] Zhang, H., Jiang, G., Yoshihira, K. & Chen, H. 2014. “Pro-active workload management in hybrid cloud computing,” IEEE Transactions on Network and Service Management, vol. 11, no. 1, pp. 90-100.

[5] Şenel, B. C., Mouchet, M., Cappos, J., Friedman, T., Fourmaux, O. & McGeer, R. 2023. “Multitenant containers as a service (CAAS) for clouds and edge clouds,” IEEE Access, vol. 11, pp. 144574-144601.

[6] Singh, P., Gupta, P., Jyoti, K. & Nayyar, A. 2019. “Research on auto-scaling of web applications in cloud: survey, trends and future directions,” Scalable Computing: Practice and Experience, vol. 20, no. 2, pp. 399-432.

[7] Lorido-Botran, T., Miguel-Alonso, J. & Lozano, J. A. 2014. “A review of auto-scaling techniques for elastic applications in cloud environments,” Journal of Grid Computing, vol. 12, pp. 559-592.

[8] Li, L., Zhang, Z. & Zhang, S. 2021. “Hybrid algorithm based on content and collaborative filtering in recommendation system optimization and simulation,” Scientific Programming, vol. 2021, no. 1, p. 7427409.

[9] Khalaji, M., Dadkhah, C. & Gharibshah, J. 2021. “Hybrid movie recommender system based on resource allocation,” arXiv preprint arXiv:2105.11678.

[10] Setiawan, N. V. E. B. 2024. “Movie Recommender System on Twitter Using Weighted Hybrid Filtering and GRU,” Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, vol. 9, no. 2.

[11] Sultan, U., Khan, H., Afridi, Y. S., Shah, M. I. A., Ashfaq, M. & Sultan, A. 2024. “Performance Analysis of a Hybrid Recommender System,” International Journal of Innovations in Science and Technology, pp 257-265.

[12] Vahdat-Nejad, H., Izadpanah, S. & Ostadi-Eilaki, S. 2019. “Context-aware cloud-based systems: design aspects,” Cluster Computing, vol. 22, pp. 11601-11617.

[13] Middleton, S. E., Roure, D. D. & Shadbolt, N. R. 2009. “Ontology-based recommender systems,” in Handbook on ontologies: Springer, pp. 779-796.

[14] Dautov, R., Paraskakis, I. & Kourtesis, D. 2012. “An ontology-driven approach to self-management in cloud applica-