

Predicting the workload of virtual machines in order to reduce energy consumption in cloud data centers using the combination of deep learning models

Zeinab Khodaverdian^{*}, Hossein Sadr^{**}, Mojdeh Nazari^{***}, Seyyed Ahmad Edalatpanah^{****}

^{*}Ph.D. student, Computer Department, Faculty of Mechanics, Electricity and Computer, Islamic Azad University, Science and Research Unit, Tehran, Iran

^{**}Ph.D., Department of Computer Engineering and Information Technology, Gilan University of Medical Sciences, Iran

^{***}Ph.D., Department of Medical Informatics, Faculty of Paramedicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran

^{****}BA, Department of Applied Mathematics, Ayendag Institute of Higher Education, Tonkabon, Mazandaran, Iran

Abstract

Cloud computing service models are growing rapidly, and inefficient use of resources in cloud data centers leads to high energy consumption and increased costs. Plans of resource allocation aiming to reduce energy consumption in cloud data centers has been conducted using live migration of Virtual Machines (VMs) and their consolidation into the small number of Physical Machines (PMs). However, the selection of the appropriate VM for migration is an important challenge. To solve this issue, VMs can be classified according to the pattern of user requests into Delay-sensitive (Interactive) or Delay-Insensitive classes, and thereafter suitable VMs can be selected for migration. This is possible by virtual machine workload prediction. In fact, workload predicting and predicting analysis is a pre-migration process of a virtual machine. In this paper, In order to classification of VMs in the Microsoft Azure cloud service, a hybrid model based on Convolution Neural Network (CNN) and Gated Recurrent Unit (GRU) is proposed. Microsoft Azure Dataset is a labeled dataset and the workload of virtual machines in this dataset are in two labeled Delay-sensitive (Interactive) or Delay-Insensitive. But the distribution of samples in this dataset is unbalanced. In fact, many samples are in the Delay-Insensitive class. Therefore, Random Over-Sampling (ROS) method is used in this paper to overcome this challenge. Based on the empirical results, the proposed model obtained an accuracy of 94.42 which clearly demonstrates the superiority of our proposed model compared to other existing models.

Keywords: Workload Prediction, Cloud data centers, Virtual machine selection, Convolution Neural Network, Gated Recurrent Unit

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

زینب خداوردیان*، حسین صدر**، مزده نظری سلیمان‌داری***، سید احمد عدالت پناه****

*دانشجوی دکتری تخصصی گروه کامپیوتر، دانشکده مکانیک، برق و کامپیوتر، دانشگاه آزاد اسلامی، واحد علوم و تحقیقات، تهران، ایران

**دکتر، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه علوم پزشکی گیلان، ایران

***دکتر، گروه انفورماتیک پزشکی، دانشکده پیراپزشکی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

****کارشناسی، گروه ریاضی کاربردی، موسسه آموزش عالی آیندگان، تنکابن، مازندران، ایران

تاریخ دریافت: ۱۴۰۰/۱۲/۱۰ تاریخ پذیرش: ۱۴۰۱/۰۸/۱۱

نوع مقاله: پژوهشی

چکیده

افزایش تقاضا برای کاربردهای مبتنی بر ابر و استفاده ناکارآمد از منابع، موجب مصرف بی‌رویه انرژی در مراکز داده ابری شده است. مدیریت پویای منابع در مراکز داده با هدف کاهش مصرف انرژی، از طریق پیش‌بینی بار کاری ماشین‌های مجازی امکان‌پذیر است. پیش‌بینی بار کاری ماشین‌های مجازی این امکان را می‌دهد که ماشین‌های مجازی متناسب با درخواست کاربران در زمان مناسب مهاجرت کنند و در مصرف انرژی موثر باشد و منابع را به کارآمدترین روش تخصیص دهد. پیش‌بینی بار کاری ماشین‌های مجازی می‌تواند بر اساس الگوی درخواست کاربران باشد برای این منظور می‌توان ماشین‌های مجازی را بر اساس پیش‌بینی مصرف منابع (به عنوان مثال میانگین مصرف پردازنده) در کلاس‌های حساس یا غیر حساس به تأخیر دسته‌بندی کرد و سپس، ماشین‌های مجازی متناسب با درخواست کاربران را به آن‌ها اختصاص داد. در واقع پیش‌بینی بار کاری و تحلیل پیش‌بینی به عنوان یک فرآیند اولیه برای مدیریت منابع (مانند کاهش تعداد مهاجرت در ادغام پویای ماشین‌های مجازی) باشد. از این رو در این مقاله از ترکیب شبکه عصبی پیچشی و واحد برگشتی دروازه‌دار به منظور پیش‌بینی بار کاری ماشین‌های مجازی میکروسافت آزور استفاده شده است. مجموعه داده آزور یک مجموعه داده دارای برچسب است و بار کاری ماشین‌های مجازی در این مجموعه داده در دو برچسب حساس یا غیر حساس به تأخیر قرار دارند. در این مجموعه داده اکثر ماشین‌های مجازی دارای برچسب غیر حساس به تأخیر می‌باشند؛ بنابراین بنابراین توزیع نمونه‌ها در این مجموعه داده به صورت نامتوازن است از این رو برای رفع این چالش از افزایش تصادفی نمونه‌های کلاس اقلیت استفاده شده است. طبق نتایج حاصل از آزمایش‌ها، روش پیشنهادی دارای دقت ۹۴/۴۲ است که نشان‌دهنده برتری مدل پیشنهادی نسبت به سایر مدل‌های پیشین است.

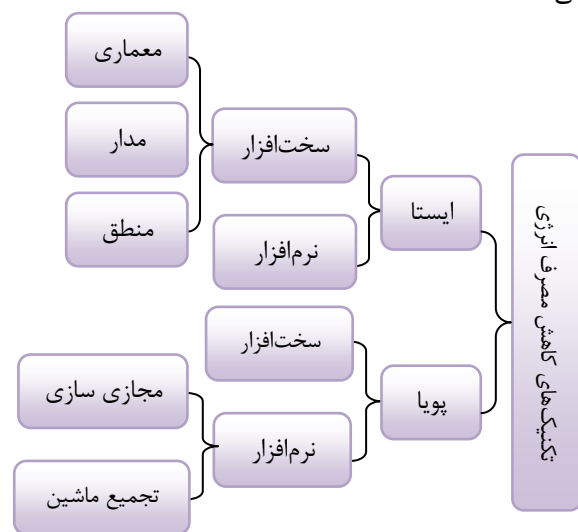
واژگان کلیدی: پیش‌بینی بار کاری، مراکز داده ابری، انتخاب ماشین‌های مجازی، شبکه عصبی پیچشی، واحد برگشتی دروازه‌دار

^۱ نویسنده مسئول: حسین صدر Sadr@qiau.ac.ir

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

۱. مقدمه

با پیشرفت‌های روزافزون فناوری اطلاعات، نیاز به انجام کارهای محاسباتی سنگین بدون داشتن سخت‌افزارها و نرم‌افزارهای گران به‌طور قابل توجهی افزایش یافته است. رایانش ابری روشی است که در آن منابع با استفاده از مجازی‌سازی، به‌صورت یک سرویس پردازشی و از طریق شبکه‌های ارتباطی عرضه می‌شود. محبوبیت استفاده از سرویس‌های ابری، موجب ساخت مراکز داده با مصرف انرژی بالا شده است [۱]؛ بنابراین کاهش مصرف انرژی جزو چالش‌های اصلی ارائه‌دهندگان سرویس‌های ابری است. به‌طور کلی تکنیک‌های کاهش مصرف انرژی در مراکز داده به دو دسته تکنیک‌های پویا و ایستا تقسیم می‌شوند. شکل (۱) روش‌های کاهش مصرف انرژی در مراکز داده ابری را نشان می‌دهد [۲، ۳].



شکل ۱. تکنیک‌های کاهش مصرف انرژی [۲، ۳]

کد، بلکه ترتیب دستورات نیز می‌تواند روی مصرف انرژی تأثیرگذار باشد. باوجود مدیریت انرژی در روش‌های ایستا، مصرف کلی انرژی همچنان به دلیل افزایش نیازمندی برای استفاده از منابع ابری، رو به رشد است و نیازمند مدیریت انرژی به‌صورت پویا می‌باشد. تکنیک‌های پویا شامل روش‌ها و استراتژی‌های زمان اجرا برای تصمیم‌گیری در چگونگی رفتار سیستم هستند. این تکنیک‌ها با توجه به میزان منابع موردنیاز و یا هر مشخصه پویای دیگری از حالت سیستم در یک مقطع زمانی مشخص، انجام می‌شود. یکی از دلایل و فرضیاتی که شیوه مدیریت پویا را ممکن می‌سازد این است که سیستم‌ها، بار کاری متغیری را در طول عملکرد خود تجربه می‌کنند. دلیل دیگر، توانایی در استنتاج و پیش‌بینی حالت آینده سیستم و تصمیم‌گیری مناسب با توجه به آن حالت می‌باشد. تکنیک‌های ایستا برای مدیریت انرژی در دو سطح سخت‌افزار و نرم‌افزار اجرا می‌شوند. در سطح سخت‌افزار، به‌عنوان نمونه با تغییر سطح ولتاژ و فرکانس پردازنده، مصرف کلی انرژی را به‌صورت پویا کاهش می‌دهد و در سطح نرم‌افزار با استفاده از مجازی‌سازی در سرور واحد و تجمیع ماشین‌های مجازی در چندین سرور، مصرف انرژی را مدیریت می‌کند [۴]. تجمیع ماشین‌های مجازی از طریق مهاجرت زنده انجام می‌شود و تعداد ماشین‌های مجازی بر روی تعداد کمتری از ماشین‌های فیزیکی قرار می‌گیرند و موجب کاهش مصرف انرژی می‌شوند [۵]. تجمیع ماشین‌های مجازی شامل تشخیص میزبان فرابار، انتخاب ماشین مجازی برای مهاجرت، تشخیص سرور فرورار و قرارگیری ماشین مجازی می‌باشد. یکی از مفروضات اصلی پژوهش انتخاب ماشین مجازی کاندیدای مهاجرت می‌باشد. انتخاب نامناسب ماشین مجازی برای مهاجرت می‌تواند مصرف انرژی را تا حد زیادی افزایش دهد. بنابراین انتخاب ماشین مجازی برای مهاجرت از اهمیت زیادی برخوردار است. انتخاب می‌تواند بر اساس الگوی درخواست کاربران و پیش‌بینی بار کاری ماشین مجازی صورت پذیرد. پیش‌بینی بار کاری و تحلیل پیش‌بینی به‌عنوان مفروض دوم و یک فرآیند اولیه قبل از انتخاب ماشین مجازی می‌باشد [۶]. این روش پیش‌بینی بار کاری، قادر به تطبیق سناریوی بار کاری فعلی با سناریوهای گذشته و برنامه‌ریزی منابع به ماشین‌های مجازی مختلف به روشی کارآمد است. نوع مشابهی از این روش برای تطبیق پذیری پویا در ابر با در نظر گرفتن

با توجه به شکل (۱) مدیریت ایستای انرژی در سطح سخت‌افزار شامل تمام روش‌های بهینه‌سازی در زمان طراحی است که با استفاده از قطعات سخت‌افزاری بسیار کارآمد، باعث کاهش مصرف انرژی می‌شوند. به‌عبارت بهینه‌سازی طراحی سیستم در سطح سخت‌افزار، پیاده‌سازی برنامه‌هایی که روی سیستم اجرا می‌شوند بسیار مهم است و طراحی ضعیف نرم‌افزار می‌تواند منجر به کاهش کارایی سیستم شود. تجزیه و تحلیل مصرف انرژی ناشی از برنامه‌های بزرگ، غیرعملی یا غیرممکن است زیرا نه تنها فرآیند تولید و کامپایل

به تأخیر می‌باشند؛ بنابراین بنابراین توزیع نمونه‌ها در این مجموعه داده به صورت نامتوازن است. با توجه به موارد بیان شده اهداف اصلی مقاله شامل موارد زیر است:

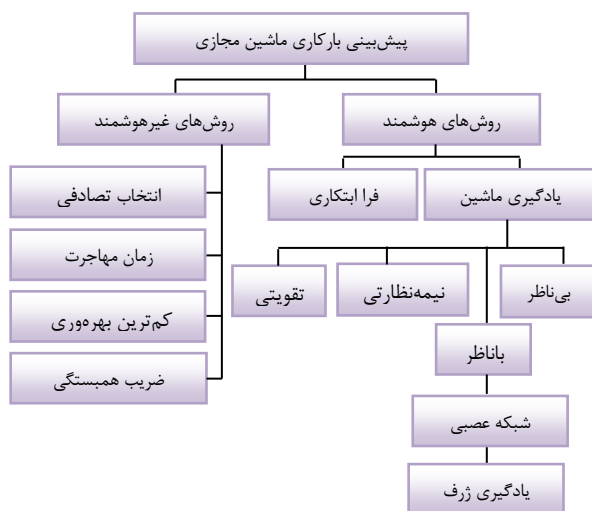
(۱) پیش‌بینی بار کاری ماشین‌های مجازی مجموعه داده آزور به منظور انتخاب ماشین مجازی کاندیدای مهاجرت، با استفاده از ترکیب شبکه عصبی پیچشی^۱ و واحد برگشتی دروازه‌دار^۲.

(۲) با توجه به اینکه تعداد ماشین‌های مجازی موجود در کلاس غیر حساس به تأخیر نسبت به کلاس دیگر بیشتر است و عدم توزیع یکنواخت نمونه‌ها در دو کلاس دقت طبقه بند را به سمت کلاس اکثریت منحرف کند و آموزش مدل را تحت تأثیر قرار بدهد؛ بنابراین لازم است تا برای مقابله با این مشکل مجموعه داده موردنظر متوازن شود و در این پژوهش از تکنیک افزایش تصادفی کلاس اقلیت^۳ برای این منظور استفاده می‌شود.

در ادامه، بخش دوم از این مقاله به کارهای پیشین اختصاص دارد. در بخش سوم روش پیشنهادی تشریح می‌گردد. در بخش چهارم آزمایش‌های مربوط به مدل پیشنهادی بر روی مجموعه داده میکروسافت آزور بررسی می‌شود و در بخش پنجم نتیجه‌گیری و کارهای آینده بیان می‌شود.

۲. کارهای پیشین

تکنیک‌های انتخاب ماشین مجازی را می‌توان در دو دسته هوشمند و غیرهوشمند دسته‌بندی کرد. شکل (۱) یک رده‌بندی از روش‌های موجود را نشان می‌دهد.



شکل ۲. رده بندی روش‌های انتخاب ماشین مجازی [۵]

داده‌های تاریخی برای شناسایی الگوهای کاربرد مشابه با مجموعه رکورد فعلی استفاده می‌شود. با توجه به مطالب بیان شده پرسش‌های اصلی پژوهش شامل موارد زیر می‌باشد:

(۱) چگونه می‌توان بر اساس الگوی درخواست کاربران پیش‌بینی بار کاری ماشین مجازی را انجام داد؟

(۲) آیا مجموعه داده مناسبی به منظور پیش‌بینی بار کاری ماشین مجازی وجود دارد؟

(۳) با توجه به اهمیت کاهش مصرف انرژی و لزوم پیش‌بینی دقیق، چه مدلی می‌توان برای افزایش دقت پیش‌بینی ارائه کرد؟

برای پاسخ به سوالات فوق، می‌توان ماشین‌های مجازی را در کلاس‌های حساس به تأخیر یا غیر حساس به تأخیر دسته‌بندی کرد، سپس بر این اساس، ماشین‌های مجازی را برای مهاجرت انتخاب نمود [۶]. کلاس حساس به تأخیر مربوط ماشین‌های مجازی است برای اجرای برنامه‌هایی با زمان پاسخگویی کم و کارایی بالا در نظر گرفته می‌شوند و کلاس غیر حساس به تأخیر عموماً برای اجرای برنامه‌هایی می‌باشد که نسبت به تداخل حساس نیستند. در این پژوهش، به‌منظور پیش‌بینی بار کاری از ماشین‌های مجازی موجود در سرویس ابری میکروسافت آزور استفاده شده است. مجموعه داده آزور یک مجموعه داده دارای برچسب است و بار کاری ماشین‌های مجازی در این مجموعه داده در دو برچسب حساس یا غیر حساس به تأخیر قرار دارند.

اخیراً استفاده از تکنیک‌های یادگیری ژرف به‌عنوان راه‌حل‌های مؤثری در این زمینه معرفی شده‌اند. زیرا، برای انتخاب بهینه، پیش‌بینی و استخراج ویژگی‌های سطح بالا ماشین‌های مجازی بسیار حائز اهمیت است. یادگیری ژرف شاخه‌ای از یادگیری ماشین است و توانایی استخراج الگوهای غیرخطی از پیش‌بینی حجم کار ماشین مجازی را دارند. [۷، ۸] از این‌رو استفاده از تکنیک‌های یادگیری ژرف می‌تواند در طبقه‌بندی ماشین مجازی مؤثر باشد؛ یادگیری ژرف می‌تواند همبستگی چندگانه بین ماشین‌های مجازی را بر اساس بار کار قبلی خود استخراج می‌کند و بار کاری آینده آن‌ها را با دقت بالا پیش‌بینی کند. پیش‌بینی بار کاری ماشین مجازی به تصمیم‌گیری‌ها برای برنامه‌ریزی ظرفیت و استفاده از مکان مناسب و مهاجرت ماشین مجازی کمک می‌کند [۹]. در مجموعه داده آزور که مربوط به ماشین‌های مجازی سرویس ابری میکروسافت آزور می‌باشد، اکثر ماشین‌های مجازی دارای برچسب غیر حساس

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

داده، تعداد نمونه‌های کلاس اکثریت به صورت تصادفی کاهش پیدا کرد. نتایج نشان می‌دهند که استفاده از متد بگینگ برای الگوریتم ماشین بردار پشتیبان، می‌تواند به طور قابل توجهی عملکرد مدل را افزایش دهد. همچنین استفاده از تکنیک بوستینگ بر روی درخت تصمیم^۷ در آزمایش ورما و شارما [۱۵]، دارای دقت ۷۴ درصد برای طبقه‌بندی ماشین مجازی می‌باشد. در پژوهشی که توسط ویتانتو [۶] و همکاران انجام شد. آموزش شبکه عصبی مصنوعی بر روی مجموعه داده پلنت لب برای ۱۰۰ ایپوک انجام شد. دقت شبکه عصبی برابر با ۷۴/۹۲ درصد روی مجموعه داده آموزشی و ۷۵/۰۵ درصد برای داده آزمایش است. ژانگ و همکاران [۱۶] شبکه عصبی برگشتی^۸ را به عنوان مدل پیشنهادی در مقایسه با مدل میانگین حرکت یکپارچه رگرسیون خودکار^۹ بر روی مجموعه داده گوگل، مورد ارزیابی قرار دادند که رویکرد مبتنی بر شبکه عصبی برگشتی از دقت بالاتری برخوردار است. شبکه عصبی برگشتی برای پردازش داده‌های سری زمانی مناسب است. با این حال، از مشکل وابستگی بلند رنج می‌برد که مانع پردازش یک دنباله بلند می‌شود [۱۷]. برای غلبه بر چنین مشکلی ژو و همکاران [۱۸] شبکه عصبی حافظه طولانی کوتاه مدت^{۱۰} که قادر به یادگیری وابستگی‌های بلندمدت است، را پیشنهاد دادند. رویکرد پیشنهادی با استفاده از مکانیسم توجه^{۱۱} در مقایسه با شبکه عصبی برگشتی و واحد برگشتی دروازه دار برتری دارد. پاتل و میسرا [۹] شبکه عصبی برگشتی، شبکه عصبی حافظه‌ی کوتاه مدت طولانی، ماشین بولتزمن و شبکه عصبی پیچشی را بر روی مجموعه داده پلنت لب مورد آزمایش قرار دادند. در مقایسه بین روش‌های موجود، شبکه عصبی حافظه طولانی کوتاه مدت عملکرد پیش‌بینی را بهبود می‌بخشند. اوهامه و همکاران [۱۹] از ترکیب شبکه عصبی حافظه طولانی کوتاه مدت و اتورگرسیو^{۱۲} برای پیش‌بینی بار کاری ماشین‌های مجازی بر روی داده‌های واقعی از ۱۷۵۰ ماشین مجازی در یک مرکز داده توزیع شده استفاده کردند. در روش پیشنهادی ورودی الگوریتم، سری زمانی پردازنده و حافظه است که با استفاده از داده‌های تاریخی حجم کار تشکیل می‌شوند. ثابت بودن هر سری زمانی با استفاده از آزمون تقویت دیکی-فولر^{۱۳} بررسی می‌شود. از مدل اتورگرسیو برای فیلتر کردن وابستگی‌های خطی در بین سری‌های زمانی چند متغیره و مدل شبکه عصبی حافظه طولانی کوتاه مدت انباشته شده

برخی از روش‌های غیرهوشمند انتخاب ماشین مجازی شامل مواردی همچون انتخاب تصادفی، کمترین زمان مهاجرت، کمترین بهره‌وری و حداکثر ضریب همبستگی می‌باشند. در انتخاب تصادفی، یک ماشین مجازی به صورت تصادفی انتخاب می‌شود [۱۰]. کمترین زمان مهاجرت می‌تواند به عنوان روش دیگری برای مهاجرت ماشین مجازی در نظر گرفته شود و زمان مهاجرت یک ماشین مجازی برابر با مقدار حافظه مورد استفاده ماشین مجازی تقسیم بر پهناى باند میزبان (سرور فیزیکی که ماشین مجازی روی آن قرار دارد) می‌باشد. در روش دیگر ماشین مجازی که دارای کمترین بهره‌وری پردازنده است، می‌تواند برای مهاجرت انتخاب شود. در روش حداکثر ضریب همبستگی، انتخاب بر اساس بیش‌ترین همبستگی (از نظر بهره‌وری پردازنده) در مقایسه با دیگر ماشین‌های مجازی است [۱۱]. از آنجایی که انتخاب ماشین مجازی یک مسئله کلیدی در کاهش مصرف انرژی است، تکنیک‌های انتخاب هوشمند می‌تواند عملکرد بهتری در این زمینه داشته باشند و از طریق پیش‌بینی بار کاری ماشین مجازی انجام شود. پیش‌بینی عبارت است از تعیین مقدار یک متغیر وابسته برحسب مقادیر متغیر مستقل و مهم‌ترین روش پیش‌بینی عددی، رگرسیون است. رافائل مورنو [۱۲] مدل رگرسیون بردار پشتیبان را به منظور پیش‌بینی بار کاری ماشین‌های مجازی در مجموعه داده کمپلوتنسه^۴ باهدف کاهش مصرف انرژی پیشنهاد داد. مدل پیشنهادی او داده‌های ورودی را در فضای ویژگی با ابعاد بالا با نگاشت غیرخطی و با استفاده از تابع کرنل ترسیم می‌کند. سپس، یک مدل رگرسیون خطی برای برگشت به فضای ویژگی جدید مورد استفاده قرار می‌گیرد و نتایج بهتری در مقایسه با مدل‌های غیرهوشمند است. رگرسیون خطی، رگرسیون سه‌تیغی^۵، رگرسیون لاسو و رگرسیون بردار پشتیبان تکنیک‌هایی هستند که استریکی [۱۳] بر روی مجموعه داده آزور مورد آزمایش قرار داد. نتایج حاصل از آزمایش وی نشان‌دهنده برتری رگرسیون بردار پشتیبان نسبت به سایر روش‌های مبتنی بر رگرسیون است. شاو و همکاران [۱۴] تکنیک‌های تجمعی^۶ را بر روی ماشین بردار پشتیبان، شبکه عصبی مصنوعی و لجستیک رگرسیون باهدف طبقه‌بندی ماشین‌های مجازی آزور آزمایش کردند؛ در پژوهش آن‌ها به منظور مقابله با عدم توازن داده‌ها در مجموعه

تاریخی مورد استفاده در آزمایش‌ها به دست می‌آیند. در گام دوم، سه روش در مرحله پیش‌پردازش داده اتخاذ می‌شوند. با توجه به مقیاس بزرگ بار کار و سری‌های زمانی کاربرد منابع، لگاریتم طبیعی ابتدا برای کاهش مقیاس داده‌های اصلی استفاده می‌شود. علاوه بر این، یک فیلتر 1^8SG برای کاهش اطلاعات نویز موجود در داده‌های اصلی استفاده و بعد از آن، نرمال‌سازی برای هر ویژگی از داده‌ها استفاده می‌شود. پس از پیش‌پردازش اطلاعات، حافظه طولانی کوتاه‌مدت دو طرفه 1^9 و حافظه طولانی کوتاه‌مدت شبکه‌ای 2^0 برای آموزش و آزمایش داده‌های سری زمانی با هم ادغام می‌شوند. مدل پیشنهادی آن‌ها شامل یک لایه حافظه طولانی کوتاه‌مدت شبکه‌ای است که در وسط دو لایه حافظه طولانی کوتاه‌مدت دو طرفه قرار دارد. بعد از آن، خروجی آن از طریق یک لایه تمام متصل برای تولید خروجی نهایی عبور می‌کند.

شیشیرا و همکاران [۲۴] به منظور افزایش دقت طبقه‌بندی بر روی مجموعه داده آזור تکنیکی را برای استخراج ویژگی ارائه دادند. در پژوهش آن‌ها پیش‌پردازش به منظور حذف نویز از داده‌های اصلی انجام می‌شود. الگوریتم پردازش ویژگی سه فاکتور کلیدی را در محدوده تعریف شده مانند مصرف پردازنده، ظرفیت حافظه و دیسک تخمین می‌زند. بر اساس این الگوریتم متناسب با هر ورودی یک برچسب برای ۳ فاکتور نامبرده در نظر گرفته می‌شود و برچسب‌ها بر اساس یک آستانه از پیش تعریف شده اختصاص می‌یابند. به عنوان مثال اگر مصرف پردازنده کم‌تر از حداقل سطح انتظار باشد، به ورودی هیچ برچسب باری تخصیص داده نمی‌شود. اگر مقدار ویژگی بزرگ‌تر از حداکثر سطح باشد، برچسب خارج از حد مجاز اختصاص داده می‌شود. در مورد مقادیر ویژگی که در محدوده مجاز قرار دارند ۵ برچسب بار کاری که شامل بار سبک، بار ملایم، بار کم، بار زیاد و بار اضافه است، اختصاص داده می‌شود. برای اعتبار سنجی عملکرد مدل استخراج ویژگی پیشنهادی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان را انتخاب شدند. مدل پیشنهادی در مقایسه با جنگل تصادفی و شبکه عصبی برگشتی دارای دقت بالاتری است.

۱.۲. بررسی مزایا و معایب پژوهش‌های پیشین

روش‌های هوشمند بیان شده در بخش کارهای پیشین هر کدام دارای نقاط قوت و ضعفی هستند که بررسی مزایا و معایب آن‌ها می‌تواند در ارائه مدل پیشنهادی مؤثر باشد. جدول (۱)

برای گرفتن روندهای غیرخطی در باقیمانده‌های محاسبه شده از مدل اتورگرسیو استفاده می‌شود که خطای پیش‌بینی را کاهش می‌دهد و همچنین تعمیم‌پذیری پایینی دارد. ترکیب اسپارس اتوانکدر 1^4 و واحد برگشتی دروازه‌دار توسط چن و همکاران [۲۰] باهدف کاهش مصرف انرژی با پیش‌بینی بار کاری ماشین‌های مجازی روی مجموعه داده علی‌بابا و گوگل ارائه شد. مدل پیشنهادی دارای قابلیت تعمیم و پایداری بیشتری است و عملکرد بهتری در مقایسه با شبکه عصبی برگشتی، حافظه طولانی کوتاه‌مدت و واحد برگشتی دروازه‌دار دارد. یزدانیان و شریفیان [۲۱] مدلی ترکیبی مبتنی بر شبکه عصبی پیچشی و شبکه عصبی حافظه طولانی کوتاه‌مدت پیشنهاد کردند. ویژگی‌های استخراج شده از شبکه عصبی پیچشی به منظور پیش‌بینی بار کاری ماشین مجازی وارد شبکه عصبی حافظه طولانی کوتاه‌مدت می‌شود. مدل پیشنهادی آن‌ها بسیار مؤثر برای پردازش توالی‌های طولانی از داده‌های تاریخی با هزینه‌های محاسباتی کم است. همچنین یزدانیان و شریفیان [۲۲] یک معماری از شبکه عصبی مولد متخاصم 1^5 برای پیش‌بینی بار کاری ماشین‌های مجازی پیشنهاد دادند. معماری پیشنهادی آن‌ها شامل یک پشته از شبکه عصبی حافظه طولانی کوتاه‌مدت به عنوان مولد 1^6 است. دومین عنصر در ساختار شبکه عصبی مولد متخاصم که شبکه تشخیص دهنده 1^7 است، که موجب بهبود کیفیت پیش‌بینی شبکه مولد می‌شود در واقع، وظیفه اصلی شبکه تشخیص دهنده، آموزش خصوصت‌آمیز است که منجر به یادگیری پایدار و قوی شبکه تولید کننده می‌شود. یک شبکه پیچشی چند لایه و دو لایه کاملاً متصل نیز به عنوان تشخیص دهنده در معماری پیشنهادی آن‌ها استفاده می‌شود. معماری پیشنهادی آن‌ها قادر به شناسایی وابستگی‌های غیرخطی و پیچیده بین نمونه‌های پی در پی از بار کار سری‌های زمانی می‌باشد و در نتیجه بهبود قابل توجهی در دقت پیش‌بینی بار کار ماشین مجازی دارد که در تصمیم‌گیری مدیریت منابع آینده استفاده می‌شود.

در پژوهش جینگ بی و همکاران [۲۳] در ابتدا، داده‌های تاریخی موجود در ردیابی خوشه گوگل جمع‌آوری می‌شود. با تجزیه و تحلیل و سازماندهی اطلاعات کلیدی وظایف، تعداد وظایف و سوابق استفاده از منابع از جمله استفاده از پردازنده و رم، برای هر شیار زمانی شمارش می‌شوند. سپس، بار کار و سری‌های زمانی استفاده از منابع به عنوان سری‌های زمانی

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

برخی از مزایا و معایب پژوهش‌های پیشین را نشان می‌دهد. در بین روش‌های هوشمند بیان‌شده، استفاده از تکنیک‌های یادگیری ژرف با توجه به قابلیت استخراج ویژگی به صورت اتوماتیک، دارای نتایج بهتری نسبت به روش‌های یادگیری ماشین می‌باشد [۲۵]. از آنجایی که حجم کار ماشین‌های مجازی در یک بازه زمانی و با توالی منظم جمع‌آوری می‌شوند؛ بنابراین برای پیش‌بینی، مدل‌هایی همچون شبکه‌های عصبی برگشتی که قادر به ثبت وابستگی‌های بلندمدت هستند، دارای عملکرد بهتری می‌باشند. از طرفی شبکه‌های عصبی برگشتی دارای مشکلاتی همچون محوشدگی و انفجار گرادیان هستند و همچنین در استخراج ویژگی‌های محلی عملکرد ضعیفی دارند. شبکه‌های عصبی پیچشی می‌توانند در استخراج ویژگی‌های محلی به خوبی عمل کنند اما در استخراج وابستگی‌های طولانی مدت ضعیف هستند. همانطور که در مطالعات قبلی مشاهده شد، مدل‌های ترکیبی عملکرد خوبی داشته و موجب بهبود نتایج می‌شوند. دلیل استفاده از مدل‌های ترکیبی این است که مدل‌ها می‌توانند نقاط ضعف یکدیگر را بپوشانند و باعث هم‌افزایی شوند [۲۶]. اما از سوی دیگر، مدل‌های ترکیبی پیچیده‌تر از سایر مدل‌ها هستند. در این راستا، یک مدل ترکیبی مبتنی بر شبکه عصبی پیچشی و واحد برگشتی دروازه‌دار در این مقاله پیشنهاد شده است که هدف آن غلبه بر چالش‌های ذکر شده است.

جدول ۱. بررسی مزایا و معایب پژوهش‌های پیشین

روش	جزئیات مدل	داده	معیار ارزیابی	مزایا	معایب	مدل پیشنهادی
رگرسیون بردار پشتیبان [۱۱]	اندازه داده‌های آموزش، دوره تأخیر و پیش‌بینی حالت بعدی به عنوان پارامترهای مدل در نظر گرفته شده است، همچنین داده‌های یک دوره ۴ هفته‌ای به عنوان داده آموزش در نظر گرفته می‌شود	مجموعه داده کمپلوتن سه	خطای میانگین مربعات و خطای میانگین مطلق	امکان پیش‌بینی به صورت غیرخطی	طولانی بودن فاز آموزش داده‌ها (به دلیل بزرگ بودن مجموعه داده آموزش)	استفاده از یک مدل ترکیبی در کنار هم با تعداد پارامترهای کمتر علاوه بر افزایش دقت پیش‌بینی فاز آموزش را نیز کاهش می‌دهد.
تکنیک بگینگ روی ماشین بردار پشتیبان [۱۳]	ویژگی‌ها با استفاده از ضریب همبستگی پیرسون انتخاب شده است (ویژگی‌های انتخاب شده عبارتند از: میانگین و حداکثر استفاده از پردازنده، شناسه اشتراک و میزان حافظه ماشین مجازی) همچنین برای مقابله با عدم توازن مجموعه داده از حذف تصادفی نمونه‌های کلاس اکثریت استفاده شده است	ماشین‌های مجازی آزر	دقت، صحت و فراخوانی	افزایش دقت طبقه‌بندی نسبت به طبقه‌بند منفرد، کاهش بیش برآزش	همبستگی مثبت بین طبقه‌بندها، میزان واریانس و در نتیجه خطا را افزایش خواهد داد	حذف نمونه‌ها از کلاس اکثریت به منظور توازن در مجموعه داده موجب از بین رفتن اطلاعات مفید می‌شود. بنابراین در مدل پیشنهادی از افزایش نمونه‌ها در کلاس اقلیت استفاده شده است.
تکنیک بوستینگ روی درخت تصمیم [۱۴]	با همبستگی پیرسون، پارامترهای اصلی شامل (بار کاری در هر روز از هفته، بار کاری آخر هفته، تعداد شکست زمان بیکاری و زمان اوج کار سرور مجازی) انتخاب می‌شود.	یک مجموعه داده از مخزن UCI	دقت	افزایش دقت طبقه‌بندی	انتخاب تعداد مناسب از طبقه‌بند پایه	استفاده از دو تکنیک ژرف موجب افزایش دقت طبقه‌بندی می‌شود.

برای مقابله با عدم تعادل داده‌ها از آرایه‌های وزنی برای محاسبه توابع تلفات استفاده می‌شود که موجب پیچیدگی مدل خواهد شد. در این پژوهش توازن مجموعه داده در مرحله پیش پردازش انجام می‌شود.	افزایش پیچیدگی مدل با افزایش تعداد لایه‌ها	امکان دسته‌بندی به صورت غیرخطی	دقت	مجموعه داده پلنت لب	برای مقابله با عدم تعادل داده‌ها از آرایه‌های وزنی برای محاسبه توابع تلفات مورداستفاده قرار می‌گیرند. آموزش با استفاده از اندازه دسته‌ای ۱۰۰ برای ۱۰۰ ایپوک انجام می‌شود.	شبکه عصبی مصنوعی [۶]
استفاده از شبکه عصبی واحد برگشتی دروازه‌دار در معماری مدل پیشنهادی قادر به حل چالش محو شدگی گرادیان در مواجهه با داده‌های دارای توالی می‌باشد.	محوشدگی و انفجار گرادیان	مناسب برای پردازش داده‌های سری زمانی	خطای جذر میانگین	مجموعه داده خوشه گوگل	ترکیبی از ۳ گره ورودی، ۱۰ گره پنهان ۱۰ گام زمانی و نرخ یادگیری ۰,۰۱ برای پیش‌بینی مصرف پردازنده و ترکیبی از ۵ گره ورودی، ۱۰ گره پنهان ۱۰ گام زمانی و نرخ یادگیری ۰,۰۰۵ برای پیش‌بینی رم	شبکه عصبی برگشتی [۱۵]
استفاده از تعداد لایه‌های کم در معماری مدل پیشنهادی موجب کاهش پیچیدگی محاسباتی می‌شود.	افزایش پیچیدگی محاسباتی	بهبود دقت پیش‌بینی	خطای جذر میانگین مربعات	مجموعه داده علی‌بابا	مدل شامل یک شبکه رمزگذار و یک شبکه رمزگشایی مبتنی بر شبکه عصبی حافظه کوتاه‌مدت طولانی با ۶۴ واحد پنهان و یک بردار زمینه می‌باشد.	شبکه عصبی حافظه طولانی کوتاه‌مدت و مکانیسم توجه [۱۶]
شبکه عصبی پیچشی در معماری مدل پیشنهادی قادر به استخراج ویژگی‌های محلی خواهد بود.	ضعف در استخراج ویژگی محلی	حل مشکل محوشدگی و انفجار گرادیان	خطای جذر میانگین مطلق	مجموعه داده پلنت لب	استفاده پردازنده ماشین‌های مجازی به مدت ۷ روز به عنوان مجموعه آموزش. برای پیش‌بینی حجم کار تنها ۲۸۸۰ فواصل زمانی وجود دارد که ۷۰٪ برای آموزش و ۳۰٪ برای آزمایش استفاده می‌شود.	شبکه عصبی حافظه کوتاه‌مدت طولانی [۸]
مدل پیشنهادی با استفاده از تکنیک حذف تصادفی قابلیت تعمیم بالایی دارد.	قابلیت تعمیم پایین	فیلتر کردن وابستگی‌های خطی در بین سری‌های زمانی چند متغیره	خطای جذر میانگین	۱۷۵۰ ماشین مجازی در یک مرکز داده توزیع شده	از مدل اتورگرسیو برای فیلتر کردن وابستگی‌های خطی در بین سری‌های زمانی چند متغیره استفاده می‌شود بعد از آن از یک شبکه عصبی حافظه کوتاه‌مدت طولانی انباشته با ۱۰۰ بلوک (یا نورون) و تابع فعال‌سازی ریلو استفاده شده است	شبکه عصبی حافظه طولانی کوتاه‌مدت و اتورگرسیو [۱۷]
استفاده از تعداد لایه‌های کم در معماری مدل پیشنهادی موجب	افزایش پیچیدگی محاسباتی	قابلیت تعمیم و پایداری بالا	خطای جذر میانگین مربعات	مجموعه داده علی‌بابا و	اسپارس اتوانکدر برای فشرده‌سازی ابعاد داده‌های ورودی و واحد برگشتی دروازه‌دار برای پیش‌بینی بار کاری استفاده	اسپارس اتوانکدر و واحد برگشتی دروازه‌دار [۱۸]

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

کاهش پیچیدگی محاسباتی می‌شود.				خوشه گوگل	می‌شود. معیارهای اساسی مرتبط با پیش‌بینی، از جمله شناسه ماشین مجازی، زمان شروع، استفاده از پردازنده، میزان حافظه، استخراج می‌شود. ۵۰٪ از نمونه‌ها برای آموزش، ۲۵٪ برای اعتبار سنجی و ۲۵٪ برای تست هستند. همچنین تعداد دوره‌های آموزشی ۱۰۰ و نرخ یادگیری ۰,۰۳ می‌باشد.	
تعداد لایه‌های کمتر باعث کاهش تعداد پارامترها و در نتیجه کاهش زمان آموزش شده است.	افزایش تعداد پارامترها و افزایش زمان آموزش مدل	مناسب برای پردازش توالی‌های طولانی	خطای جذر میانگین	مجموعه داده خوشه گوگل	میزان استفاده از پردازنده و رم توسط شبکه عصبی پیچشی استخراج می‌شود و در ادامه از دو لایه شبکه عصبی حافظه طولانی کوتاه‌مدت برای پیش‌بینی حالت بعدی استفاده شده است. همچنین و برای مقابله با بیش‌برازش، از تکنیک دراپ‌اوت استفاده شد.	شبکه عصبی پیچشی و حافظه طولانی کوتاه مدت [۱۹]
تعداد لایه‌های کمتر باعث کاهش تعداد پارامترها و در نتیجه کاهش زمان آموزش شده است.	افزایش تعداد لایه ها و در نتیجه افزایش زمان آموزش	افزایش دقت پیش‌بینی برای داده‌های دارای توالی	خطای میانگین مربعات	مجموعه داده پلنت‌لب	یک پشته از شبکه عصبی حافظه طولانی کوتاه‌مدت به عنوان مولد و یک شبکه پیچشی چند لایه و دو لایه کاملاً متصل نیز به عنوان تشخیص دهنده در معماری پیشنهادی است.	شبکه عصبی مولد متخاصم [۲۰]
تعداد لایه‌های کمتر باعث کاهش تعداد پارامترها و در نتیجه کاهش زمان آموزش شده است.	زمان آموزش طولانی	مناسب برای پردازش توالی‌های طولانی	خطای میانگین مربعات	مجموعه داده خوشه گوگل	یک لایه حافظه طولانی کوتاه‌مدت شبکه‌ای که در وسط دو لایه حافظه طولانی کوتاه‌مدت دو طرفه قرار دارد. بعد از آن، خروجی آن از طریق یک لایه تمام متصل برای تولید خروجی نهایی عبور می‌کند.	حافظه طولانی کوتاه‌مدت [۲۱]
استفاده از دو تکنیک ژرف موجب انتخاب ویژگی به صورت اتوماتیک هستند.	انتخاب ویژگی با این روش مستلزم افزایش زمان و هزینه است	افزایش دقت طبقه‌بندی	دقت	ماشین‌های مجازی آزر	پردازش ویژگی سه فاکتور کلیدی را در محدوده تعریف شده مانند مصرف پردازنده، ظرفیت حافظه و دیسک تخمین می‌زند و متناسب با هر ورودی یک برچسب برای ۳ فاکتور نامبرده در نظر گرفته می‌شود و برچسب‌ها بر اساس یک	شبکه عصبی مصنوعی و ماشین بردار پشتیبان [۲۲]

					آستانه از پیش تعریف شده اختصاص می‌یابند.
--	--	--	--	--	--

۱.۳. پیش‌پردازش

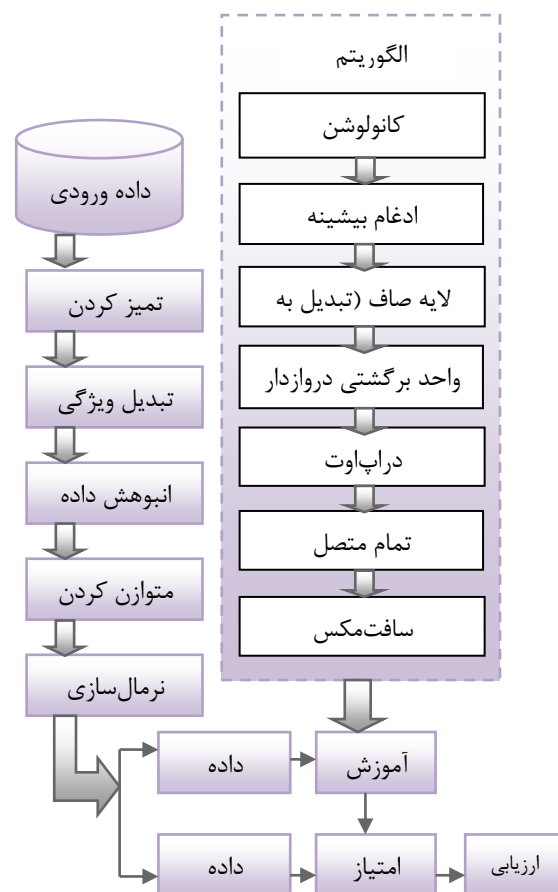
مراحل پیش‌پردازش در پژوهش حاضر شامل مراحل زیر است: **پاک‌سازی داده:** با توجه به این‌که اطلاعات مربوط به ماشین مجازی در یک بازه زمانی مشخص از یک یا چند مرکز داده جمع‌آوری می‌شوند بنابراین ریسک جمع‌آوری داده ناقص بالا می‌رود. همچنین ممکن است برخی از نمونه‌ها به‌درستی برچسب‌گذاری نشده باشند؛ بنابراین در این پژوهش به‌منظور پاک‌سازی داده مجموعه رکوردهایی که دارای برچسب‌های ناشناخته هستند از مجموعه داده حذف می‌شوند تا روند پیش‌بینی را دچار اختلال نکنند.

متوازن‌سازی داده: در این پژوهش به‌منظور مقابله با عدم توازن نمونه‌ها در کلاس‌های مختلف، از تکنیک افزایش تصادفی نمونه‌های کلاس اقلیت استفاده می‌شود. مزیت این روش نسبت به روش کاهش تصادفی نمونه‌های کلاس اکثریت^{۲۱} این است که در این روش نمونه‌های ارزشمند حذف نمی‌شوند. همچنین نسبت به روش‌های دیگر مانند حساس به هزینه^{۲۲} سریع‌تر اجرا می‌شوند [۲۷، ۲۸]. با توجه به این‌که در استفاده از تکنیک‌های یادگیری ژرف، تعداد نمونه‌های بیشتر تأثیر مثبت در آموزش مدل دارد، افزایش نمونه‌ها در کلاس اقلیت می‌تواند عملکرد آموزش مدل را بهبود دهد [۲۹].

تبدیل ویژگی: ویژگی‌های در نظر گرفته شده برای ماشین‌های مجازی می‌توانند از نوع اسمی باشند. بنابراین در مقاله حاضر به‌منظور تبدیل ویژگی‌های با نوع اسمی به ویژگی‌های عددی از روش کدبندی برچسب^{۲۳} استفاده شده است. روش کدبندی برچسب به تبدیل ویژگی‌ها به فرم قابل‌فهم برای الگوریتم‌های یادگیری ماشین و شبکه عصبی اطلاق می‌شود. این روش برای هر مقداری عددی با شروع از صفر تا $n-1$ در نظر می‌گیرد [۳۰]. علت انتخاب این روش در این پژوهش این است که کدبندی برچسب ابعاد مجموعه داده پس از اعمال تغییر نمی‌کند.

۳. روش پیشنهادی

در این بخش از مقاله روش پیشنهادی تشریح خواهد شد که شامل مراحل پیش‌پردازش، ساخت مدل دسته‌بندی پیشنهادی و آموزش مدل خواهد بود. شکل (۳) دیاگرام کلی روش پیشنهادی را نشان می‌دهد.



شکل ۳. دیاگرام کلی روش پیشنهادی

مراحل نشان داده شده در دیاگرام فوق در زیر بخش‌های بعدی تشریح خواهند شد.

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

نقشه‌های ویژگی^{۲۸} را می‌سازند. عناصر فیلترها در واقع وزن‌های شبکه عصبی می‌باشد واضح است که خروجی حاصل از شبکه پیچشی ۶۴ نگاشت ویژگی می‌باشد. مقادیر خروجی از لایه پیچشی باید از تابع فعال‌سازی غیرخطی ریلو عبور کنند. در غیر این صورت شبکه قادر به یادگیری الگوهای غیرخطی موجود در داده‌ها نخواهد بود. استراتژی ادغام بیشینه پس از لایه پیچشی و باهدف کاهش اندازه نقشه‌های ویژگی به کار می‌رود و می‌تواند باعث انتخاب ویژگی‌های نامتغیر و با ارزش بیشتر شود. نتایج به‌دست‌آمده از لایه پیچشی و لایه ادغام توسط لایه صاف^{۲۹} تبدیل به بردار یک‌بعدی خواهند شد. با توجه به این‌که داده‌های مربوط به اطلاعات ماشین مجازی در یک بازه زمانی مشخص و با یک توالی منظم جمع‌آوری شده‌اند؛ بنابراین برای استخراج ویژگی‌های بلندمدت به واحد برگشتی دروازه‌دار ارسال می‌شوند. جزئیات مدل پیشنهادی در شکل (۴) آمده است. همان‌طور که در شکل (۴) نشان داده شد. خروجی شبکه عصبی پیچشی به‌عنوان ورودی واحد برگشتی دروازه‌دار در نظر گرفته می‌شود زیرا شبکه عصبی پیچشی در استخراج وابستگی‌های بلندمدت عملکرد ضعیفی دارد. البته می‌توان شبکه عصبی پیچشی را با افزایش تعداد لایه‌ها برای داده‌های سری زمانی به کار برد اما این کار موجب افزایش هزینه محاسباتی خواهد شد [۸، ۳۲]. از طرفی واحد برگشتی دروازه‌دار قادر به حل این مشکل می‌باشد. همچنین وجود دروازه‌های بازنشانی^{۳۰} و به‌روزرسانی^{۳۱} در معماری این شبکه مشکلات مربوط به شبکه عصبی برگشتی مانند محوشدگی و انفجار گرادیان را حل می‌کند و نسبت به شبکه عصبی حافظه طولانی کوتاه‌مدت دارای پارامترهای کمتری برای آموزش است [۳۲].

انبوهش داده^{۲۴}: انبوهش داده نیز یک مرحله از پیش‌پردازش داده است که شامل عملیاتی است که بتوان در پی آن از ترکیب دو یا چند ویژگی، ویژگی جدیدی ایجاد کرد. این ویژگی جدید باید بتواند در کنار سایر ویژگی‌های پیشین اطلاعات موجود در یک مجموعه داده را مؤثرتر و کامل‌تر از ویژگی‌های اولیه نشان دهد [۳۱]. یکی از اهداف انبوهش داده در این پژوهش کاهش تعداد ویژگی‌ها است. علت انتخاب مرحله انبوهش داده در این پژوهش علاوه بر کاهش بعد این است که در واقع با انبوهش داده ارزش ویژگی‌ها بیشتر خواهد شد. همچنین با انبوهش داده، داده‌های پایدارتری وجود خواهد داشت.

نرمال‌سازی: در این مقاله از روش نرمال‌سازی حداقل-حداکثر^{۲۵} برای نرمال‌سازی داده‌های ورودی استفاده می‌شود. این روش یک تبدیل خطی بر روی مجموعه داده‌های اصلی انجام می‌دهد و همچنین رابطه بین مقادیر داده‌های اصلی را حفظ می‌کند. با توجه به این‌که همبستگی بین داده‌ها و روابط بین متغیرهای مستقل در پیش‌بینی مؤثر است، بنابراین از این روش نرمال‌سازی در این پژوهش استفاده شده است. رابطه (۱) فرمول این نرمال‌سازی را نشان می‌دهد:

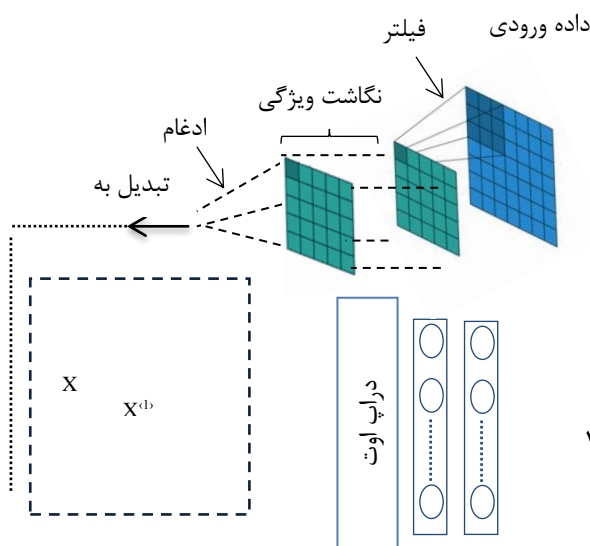
$$x' = \frac{x - \text{Min}(A)}{\text{Max}(A) - \text{Min}(A)} \quad (1)$$

در رابطه فوق $\text{Min}(A)$ و $\text{Max}(A)$ به ترتیب حداقل و حداکثر مقدار یک ویژگی و X برابر با مقدار ویژگی فعلی می‌باشند.

خروجی مرحله پیش‌پردازش داده، شامل داده‌ای است که رکوردهایی با برچسب ناشناخته از آن حذف شده است، ویژگی‌های موجود به منظور ایجاد ویژگی جدید با هم ترکیب شده‌اند، ویژگی‌هایی که به صورت رشته بودند به فرم عددی تبدیل شدند. همچنین تعداد نمونه‌های دو کلاس با تکنیک متوازن‌سازی بیان شده برابر است و در نهایت مقادیر متغیرها بین صفر و یک نرمال شده‌اند.

۲.۳. مدل دسته‌بند^{۲۶} پیشنهادی

مدل ارائه‌شده در این مقاله شامل شبکه عصبی پیچشی و واحد برگشتی دروازه‌دار است. همان‌طور که در شکل (۳) نشان داده شد، بعد از مراحل پیش‌پردازش، در فرایند یادگیری ابتدا از شبکه عصبی پیچشی استفاده می‌شود. این شبکه متشکل لایه پیچشی و ادغام بیشینه^{۲۷} می‌باشد. لایه پیچشی از ۶۴ فیلتر با اندازه ۳ تشکیل شده است که با ضرب داخلی در ورودی،



با توجه به شکل (۵) ورودی مرحله فعلی با $c\langle t-1 \rangle$ نشان داده شده است که خروجی واحد قبلی نیز می باشد همچنین خروجی واحد فعلی که ورودی واحد بعدی نیز هست با $c\langle t \rangle$ نمایش داده شده است. در این شکل $x\langle t \rangle$ ورودی و $y\langle t \rangle$ به عنوان خروجی برای این واحد است. Γr و Γu به ترتیب دروازه به روزرسانی و دروازه بازنشانی هستند و مقدار آن‌ها توسط رابطه (۲) و (۳) محاسبه می شود. $c\langle t \rangle$ کاندید وضعیت پنهان است که مشابه شبکه عصبی برگشتی سنتی در رابطه (۴) محاسبه می شود و $c\langle t \rangle$ نیز از رابطه (۵) به دست می آید.

$$\Gamma u = \sigma(Wu[c\langle t-1 \rangle, x\langle t \rangle] + bu) \quad (2)$$

$$\Gamma r = \sigma(Wr[c\langle t-1 \rangle, x\langle t \rangle] + br) \quad (3)$$

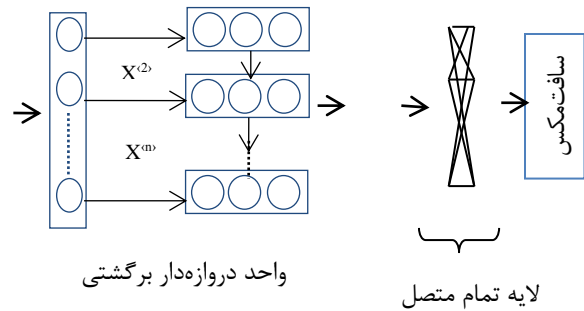
$$c\langle t \rangle = \tanh(Wc[\Gamma r * c\langle t-1 \rangle, x\langle t \rangle] + bc) \quad (4)$$

$$c\langle t \rangle = (1 - \Gamma u) * c\langle t-1 \rangle + \Gamma u * c\langle t \rangle \quad (5)$$

در روابط فوق (bu, br, bc) برابر بایاس، σ و \tanh به ترتیب توابع فعال سازی سیگموئید و تانژانت هایپربولیک هستند و (Wc, Wr, Wu) ماتریس وزن دهی را نشان می دهند. در ادامه به منظور جلوگیری از بیش برآش^{۳۲} از تکنیک حذف تصادفی (دراپ‌اوت)^{۳۳} استفاده می شود [۷, ۳۵]. نتایج به دست آمده از لایه های قبل را می توان با محاسبه مقدار تمام نورون‌ها در لایه های به هم متصل به طور کامل به دست آورد. در قسمت آخر از تابع دسته بند سافت مکس برای احتمال تعلق به کلاس غیرخطی در لایه ی خروجی شبکه عصبی و برای مشکلات مربوط به دسته بندی استفاده می شوند [۳۶]. خروجی های این تابع به گونه ای نرمال سازی شده اند که مجموع آن ها عدد یک باشد و مطابق رابطه (۶) محاسبه می شود:

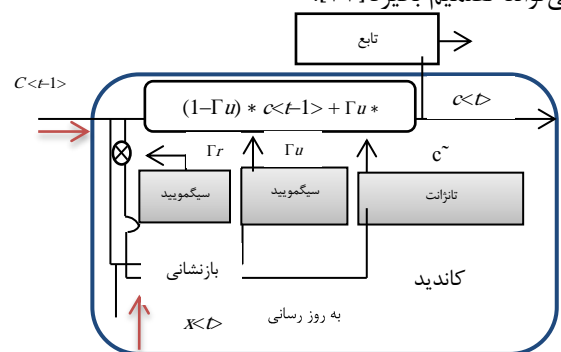
$$P(Y=k|X=x_i) = e^{s^k} / \sum_j e^{s^j} \quad (6)$$

در رابطه فوق S به عنوان ورودی k بعدی از اعداد حقیقی است. e^{s^k} تابع نمایی استاندارد برای هر عنصر از بردار ورودی اعمال می شود. مخرج کسر در رابطه فوق تضمین می کند که تمام مقادیر خروجی تابع به ۱ برسد و هر یک در محدوده صفر و یک باشند. برای ادامه باید امتیاز کلاس درست بیشینه شود. برای این کار می توان لگاریتم آن را بیشینه کرد و برای کمینه



شکل ۴. جزئیات مدل پیشنهادی

معماری واحد بررسی دروازه دار در سس (۳) آمده است. دروازه های بازنشانی و به روزرسانی در اصل دو بردارند که با استفاده از آن ها تصمیم گرفته می شود چه اطلاعاتی به خروجی منتقل شده و چه اطلاعاتی منتقل نشود [۳۳]. نکته خاص درباره این دروازه ها این است که این دروازه ها را می توان آموزش داد تا اطلاعات مربوط به گام های زمانی قبل را بدون آن که در طی گام های زمانی مختلف دستخوش تغییر شوند، حفظ کند. دروازه به روزرسانی وظیفه کنترل جریان اطلاعات جدید را بر عهده دارد. این دروازه مشخص می کند آیا در گام زمانی فعلی باید از اطلاعات جدید مورد استفاده قرار گیرد یا خیر و اگر بلی به چه میزان. دروازه بازنشانی نیز مشخص می کند چه میزان از اطلاعات گام زمانی قبل با اطلاعات گام زمانی فعلی به گام زمانی بعد منتقل شود. به طور دقیق تر با صفر بودن این سویچ این دروازه در عمل شبکه را وادار می کند به گونه ای عمل کند که گویا در حال خواندن اولین بخش از دنباله ورودی است و این طور شبکه را قادر به فراموشی حالت محاسبه شده قبلی می کند و به همین صورت می تواند با فاصله گرفتن از صفر، حالت مابینی را فراهم آورد وجود این دروازه ها به این شکل است که مکانیسم کنترلی بسیار دقیقی را ایجاد می کند. این مکانیسم کنترلی توسط تابع فعال ساز سیگموئید است که مقدار ۰ یا ۱ برمی گرداند و تصمیم می گیرد حافظه قبلی چقدر به ورودی جاری مربوط می شود و بر اساس ورودی می تواند تصمیم بگیرد [۳۴].



شکل ۵. معماری کلی واحد برگشتی دروازه دار

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

۴. آزمایش‌ها

در این بخش ابتدا مجموعه داده، معیار ارزیابی، نیازهای سخت‌افزاری و نرم‌افزاری موردنیاز معرفی می‌شود و بعداز آن به توصیف آزمایش‌ها پرداخته خواهد شد.

۱.۴. مجموعه داده

مجموعه داده برچسب دار شامل اطلاعات مربوط به تمامی ماشین‌های مجازی میکروسافت آزور است^{۳۷} که به مدت ۳ ماه از تاریخ ۱۶ نوامبر ۲۰۱۶ تا ۱۶ فوریه ۲۰۱۷ جمع‌آوری شده است [۳۷]. توصیفی مجموعه داده آزور در جدول (۲) آمده است.

جدول ۲. توصیف مجموعه داده عمومی آزور

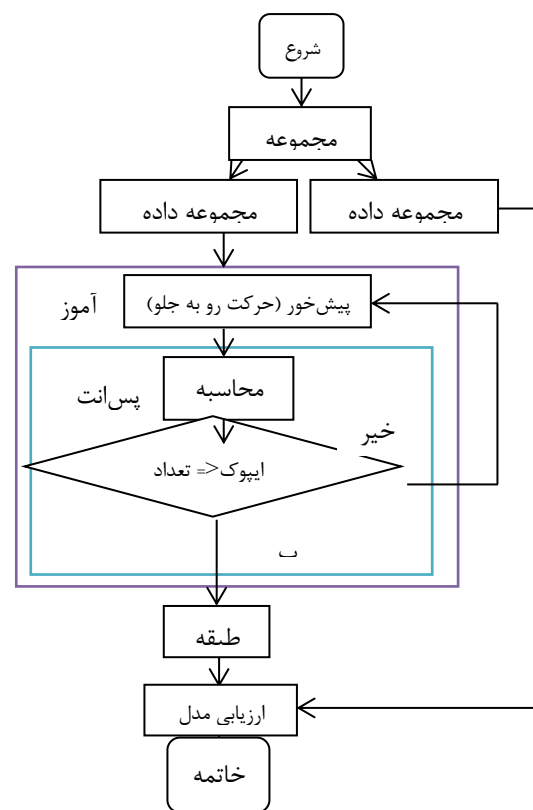
ردیف	مشخصات مجموعه داده	تشریح
۱	نوع ذخیره‌سازی مجموعه داده	رکوردهای (ماتریسی)
۲	ابعاد مجموعه داده	دوبعدی
۳	تعداد سطر (تعداد نمونه‌ها)	۱۰۴۸۵۷۸
۴	تعداد ستون (تعداد ویژگی‌ها)	۱۰ ستون ویژگی و یک ستون برچسب
۵	نوع ویژگی	اسمی (به صورت رشته) و عددی
۶	داده مفقود	ندارد
۷	برچسب ناشناخته	دارد
۸	تعداد برچسب‌ها	۲
۹	توزیع نمونه‌ها در کلاس	نامتوازن
۱۰	فرمت ذخیره‌سازی	csv

ماشین‌های مجازی آزور به منظور ارائه خدمات به دو گروه زیرساخت به‌عنوان سرویس^{۳۸} و پلت فرم به‌عنوان سرویس^{۳۹} تقسیم‌بندی می‌شوند و حجم کار بین این دو گروه ماشین مجازی تعریف می‌شود. بر این اساس حجم کار ماشین‌های مجازی آزور در گروه اول ۵۲ درصد و حجم کار گروه دوم ۴۸ درصد بر اساس شمارش تعداد ماشین‌های مجازی می‌باشد [۳۷]. تعداد ماشین‌های مجازی گروه اول کمی بیشتر از گروه دوم است، درحالی‌که ماشین‌های مجازی گروه دوم مصرف منابع را با تقریباً ۶۱ درصد از کل ساعات اصلی کنترل می‌کند. به‌منظور استفاده از خدمات و سرویس‌های آزور

کردن تابع هزینه می‌توان لگاریتم را در یک منفی ضرب کرد. مطابق رابطه (۷):

$$L_i = -\log(e^{s_i} y_i / \sum_j e^{s_j}) \quad (7)$$

فرآیند آموزش مدل روی مجموعه داده آموزشی در شکل (۶) نشان داده شده است. ابتدا داده آموزشی به‌عنوان ورودی وارد می‌شوند. وزن‌های اولیه با مقادیر تصادفی مقداردهی می‌شوند. در فرآیند پیش‌خور^{۴۰}، وزن‌ها توسط هر نورون در داده ورودی ضرب شده و با مقادیر بایاس جمع می‌شوند. در مراحل آموزش، مقدار پارامترها که شامل وزن و بایاس است توسط شبکه در عملیات پس انتشار^{۴۱} تنظیم خواهد شد تا وزن بهینه به دست بیاید. بدین‌صورت که بعد از هر ایپوک در فرآیند آموزش اختلاف مقادیر واقعی و مقادیر پیش‌بینی شده به‌عنوان تابع خطا^{۴۲} در نظر گرفته می‌شود، گرادینان تابع خطا نسبت به وزن محاسبه می‌شود و در ایپوک بعدی وزن‌ها به‌گونه‌ای به‌روزرسانی می‌شوند که مقدار تابع خطا کمینه شود. به تعداد ایپوک‌های در نظر گرفته شده مراحل فوق تکرار خواهند شد و زمانی که در نتایج تغییر قابل‌توجهی حاصل نشود مدل به همگرایی رسیده است.



شکل ۵. فلوچارت آموزش مدل

به تأخیر (تعاملی) و کلاس غیر حساس به تأخیر. کلاس تعاملی در مورد تخصیص ماشین مجازی به اجرای برنامه‌هایی است که حساس به تأخیر هستند و به زمان پاسخگویی کم و کارایی بالا نیاز دارند. نمونه‌هایی از این نوع برنامه‌ها شامل برنامه‌های بازی آنلاین است که از داده‌های زمان واقعی، پخش فیلم و برنامه‌های وب استفاده می‌کنند. کلاس غیر حساس به تأخیر عموماً برای اجرای برنامه‌هایی که نسبت به تداخل حساس نیستند در نظر گرفته می‌شود. نمونه‌هایی از این نوع بارهای کاری شامل بارهای کاری دسته‌ای مانند پردازش داده‌های مالی برای اتوماسیون کردن یک شرکت است. تعدادی از نمونه‌های این مجموعه داده نیز دارای برچسب ناشناخته هستند. دو کلاس حجم کار بر اساس تجزیه و تحلیل سری‌های زمانی متوسط استفاده از پردازنده در یک دوره ۳ روزه و با استفاده از الگوریتم تبدیل سریع فوریه به منظور پیش‌بینی حجم کار ماشین‌های مجازی برچسب زده می‌شوند. الگوریتم تبدیل سریع فوریه با توجه به اینکه می‌تواند تناوبی را در مقیاس‌های زمانی مختلف تشخیص دهد به عنوان یک الگوریتم ایده آل برای مشخص کردن حجم کار ماشین‌های مجازی در نظر گرفته می‌شود.

در این مجموعه داده میزان استفاده از پردازنده در فواصل زمانی ۵ دقیقه یک‌بار اندازه‌گیری شده است با توجه به فواصل زمانی تعریف شده و بر اساس توزیع تجمعی ۴۰ برای متوسط و حداکثر استفاده از پردازنده، ۶۰ درصد از ماشین‌های مجازی دارای متوسط استفاده از پردازنده و ۴۰ درصد از ماشین‌های مجازی دارای حداکثر استفاده از پردازنده می‌باشند. بر این اساس میزان استفاده از پردازنده در درصد زیادی از ماشین‌های مجازی کم است، به خصوص برای بار کاری گروه اول یعنی ماشین‌های مجازی که برای خدمات زیرساخت به عنوان سرویس هستند. یکی از دلایلی که استفاده واقعی از پردازنده گروه اول کم است، مربوط به ماشین‌های مجازی می‌باشد که برای آزمایش عملکرد و مقیاس‌پذیری مورد استفاده قرار می‌گیرند. به طور مثال ماشین‌های مجازی که ایجاد می‌شوند اما بدون انجام هیچ کار واقعی به سرعت از بین می‌روند. در این مجموعه داده ۱۵ درصد از ماشین‌های مجازی برای آزمایش عملکرد و مقیاس‌پذیری مورد استفاده قرار گرفتند که این ماشین‌ها متعلق به گروه اول می‌باشند. با توجه به موارد فوق تجزیه و تحلیل تناوبی، ماشین‌های مجازی را که به اندازه کافی طولانی کار می‌کنند و حداکثر استفاده از پردازنده را دارند،

کاربران می‌توانند یک یا چند اشتراک ایجاد کنند و ماشین‌های مجازی را در یک ناحیه انتخابی مستقر کنند. تمام ماشین‌های مجازی در یک استقرار در یک خوشه انتخابی در یک ناحیه (می‌تواند شامل یک یا چند مرکز داده باشد) یعنی مجموعه بزرگی از سرورهایی که در آن استقرار برای جابه‌جایی نیاز دارند اجرا می‌شوند. استقرار ممکن است قبل از پایان یافتن آن بر اساس درخواست کاربران بزرگ یا کوچک شود [۳۷]. در این مجموعه داده برای هر ماشین مجازی ویژگی‌هایی در نظر گرفته شده است. این مجموعه داده شامل ۱۰ ستون ویژگی و یک ستون برچسب داده است. معرفی و تشریح ویژگی‌های مجموعه داده آزور در جدول (۳) آمده است.

جدول ۳. ویژگی‌های ماشین مجازی در مجموعه داده آزور

ردیف	ویژگی	نوع	تشریح
۱	شناسه ماشین مجازی	رشته	مرجع شناسایی منحصربه‌فرد مربوط به هر ماشین مجازی موجود در آزور می‌باشد
۲	شناسه اشتراک	رشته	مرجع شناسایی برای هر اشتراک آزور است و هر اشتراک می‌تواند تعدادی ماشین مجازی با توجه به اهداف خود در اختیار داشته باشد
۳	شناسه استقرار	رشته	یک مرجع شناسایی منحصربه‌فرد برای ماشین مجازی در یک خوشه انتخابی در یک ناحیه
۴	زمان سنج ایجاد ماشین مجازی	عددی	زمانی که ماشین مجازی برای اولین بار شروع به کار می‌کند
۵	زمان سنج حذف ماشین مجازی	عددی	زمانی که کار ماشین مجازی خاتمه می‌یابد
۶	حداقل استفاده از پردازنده	عددی	اندازه‌گیری‌های مصرفی مربوط به فواصل ۵ دقیقه‌ای هستند
۷	میانگین استفاده از پردازنده	عددی	اندازه‌گیری‌های مصرفی مربوط به فواصل ۵ دقیقه‌ای هستند
۸	حداکثر استفاده از پردازنده	عددی	اندازه‌گیری‌های مصرفی مربوط به فواصل ۵ دقیقه‌ای هستند
۹	شمارنده هسته مجازی ماشین مجازی	عددی	شامل تعداد هسته‌های ماشین مجازی می‌باشد
۱۰	حافظه ماشین مجازی	عددی	حافظه ماشین مجازی برحسب گیگابایت محاسبه می‌شود

ماشین‌های مجازی در این مجموعه داده می‌توانند در دو کلاس حجم کار قرار بگیرند. این دو کلاس عبارت‌اند از کلاس حساس

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

۲.۴. معیار ارزیابی

به منظور ارزیابی مدل پیشنهادی از معیار ارزیابی دقت^{۴۱}، صحت^{۴۲} و فراخوانی^{۴۳} استفاده می‌شود و به ترتیب مطابق رابطه (۸)، (۹) و (۱۰) محاسبه می‌شوند:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (8)$$

$$precision = TP / (TP + FP) \quad (9)$$

$$Recall = TP / (TP + FN) \quad (10)$$

در روابط فوق TP و TN به ترتیب نمونه‌های مثبت و منفی هستند که درست طبقه‌بندی شده‌اند. همچنین FP و FN به ترتیب نمونه‌های مثبت و منفی هستند که نادرست طبقه‌بندی شده‌اند. همچنین به منظور پیش‌بینی میزان مصرف پردازنده از معیار خطای جذر میانگین مربعات استفاده شده است که از رابطه (۱۱) محاسبه می‌شود:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \|(y_i) - (\hat{y}_i)\|^2}{N}} \quad (11)$$

که در رابطه فوق N تعداد نمونه‌ها، y_i نشان‌دهنده مقدار واقعی و \hat{y}_i نشان‌دهنده مقدار پیش‌بینی شده می‌باشد.

۴.۳. نیازهای سیستمی

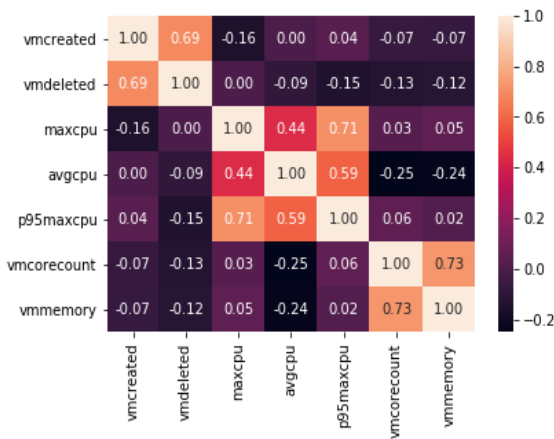
سخت‌افزار مورد استفاده برای آزمایش در این پژوهش شامل پردازنده با معماری Intel، تعداد هسته Core™ i5-4200U، سرعت ۱٫۶۰ GHz می‌باشد. سیستم از نوع ۶۴ بیتی دارای رم ۴ و سیستم عامل ویندوز ۱۰ می‌باشد. همچنین نرم‌افزار مورد استفاده در این پژوهش توزیع آناکوندا است که شامل زبان برنامه‌نویسی پایتون نیز می‌باشد و از محیط ژوپیتر نوت‌بوک برای اجرای کد و آنالیز مجموعه داده استفاده شده است. پکیج‌های مورد استفاده شامل پکیج نامپای، پانداس برای حذف رکورد و تغییر ویژگی‌ها، مت‌پلات و سی‌بورن به منظور رسم نمودارها، سایکیت‌لرن برای انجام مراحل پیش‌پردازش و کراس و تنسورفلو به منظور ساخت مدل می‌باشد.

هدف قرار می‌دهد تا یک الگوی قابل اعتماد در مدت زمان مورد نظر (حداقل ۳ روز) را شناسایی کند. یک نمونه از ماشین مجازی در کلاس حساس به تاخیر و یک نمونه از ماشین مجازی متعلق به کلاس غیر حساس به تاخیر در جدول (۴) آمده است.

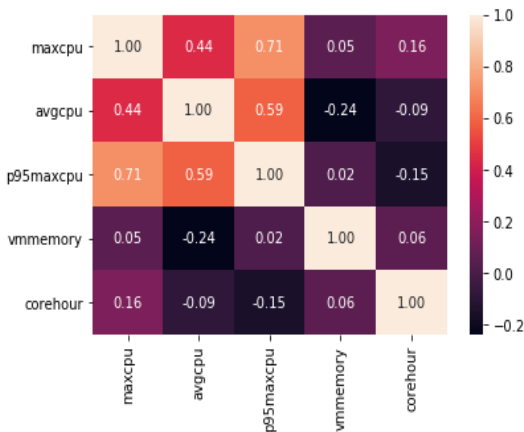
جدول ۴. مشخصات نمونه‌ها برای هر دو کلاس موجود

ردیف	ویژگی	کلاس حساس به تاخیر	کلاس غیر حساس به تاخیر
۱	شناسه ماشین مجازی	H5CxMmMoVc ZSpjgGbohn VA3R+7uCT e/hM2ht2uIYi 3t7KwX	x/XsOfHO4 ocsV99i4N1 uqKDuxctW 2MMVmwq OPAlg4wp8 m
۲	شناسه اشتراک	BSXOcywx8 pUU0DueDo 6UMo11YzR6 tn47KLEKao Xp0a	VDU4C8cq dr+ORcqqu wMRcsBA2 l0SC6lCPys 0wdg
۳	شناسه استقرار	3J17LcV4gXj Fat62qhVFRf oiWArHnY76 3HVqqI6orJC fV8	Pc2VLB8a DxK2DCC9 6itq4vW/zV Dp4wioAUi B3Ho
۴	زمان سنج ایجاد ماشین مجازی	.	.
۵	زمان سنج حذف ماشین مجازی (بر حسب ثانیه)	۱۵۳۹۳۰۰	۲۵۹۱۷۰۰
۶	حداقل استفاده از پردازنده	۳۳/۹۸۱۳۶۰	۱۰/۱۹۴۳۰۹
۷	میانگین استفاده از پردازنده	۶/۱۸۱۷۸۴	۳/۴۲۴۰۹۴
۸	حداکثر استفاده از پردازنده	۱۰۰/۰۰۰۰۰۰	۹۹/۳۶۹۸۶۹
۹	شمارنده هسته مجازی ماشین مجازی	۱	۱
۱۰	حافظه ماشین مجازی (بر حسب گیگابایت)	.۷۵	۱/۷۵

خواهد شد. در گام انبوهش داده می‌توان تعداد ویژگی‌ها را از ۱۰ به ۸ کاهش داد. به‌عنوان مثال طول عمر ماشین مجازی ۴۴ برحسب ساعت با کم کردن مقدار ویژگی حذف ماشین مجازی از مقدار ویژگی ایجاد ماشین مجازی و تقسیم آن‌ها به ۳۶۰۰ محاسبه می‌شود. از ضرب ویژگی طول عمر ماشین مجازی در ویژگی تعداد هسته ماشین مجازی ویژگی جدیدی به نام ساعت هسته ۴۵ به دست می‌آید. ویژگی‌های جدید به مجموعه داده اضافه می‌شوند و ویژگی‌های قبلی حذف خواهند شد. این کار موجب همبستگی بیشتر بین ویژگی‌ها می‌شود. نمودار همبستگی قبل از انبوهش داده در شکل (۹) و بعد از انبوهش داده در شکل (۱۰) نشان داده شده است.



شکل ۸. همبستگی بین متغیرها قبل از گام انبوهش داده

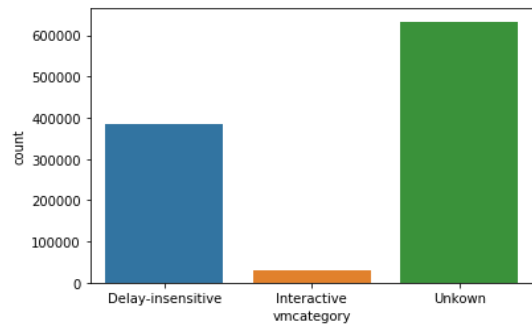


شکل ۹. همبستگی بین متغیرها بعد از گام انبوهش داده

در ادامه بعد از نرمال‌سازی، نمونه‌های موجود به مجموعه داده آموزشی، مجموعه داده اعتبارسنجی و مجموعه آزمایش تقسیم شوند. در این پژوهش ۷۰ درصد از نمونه‌های مجموعه داده در هر کلاس به‌عنوان مجموعه داده آموزشی برای مدل ۱۰ درصد مجموعه داده اعتبارسنجی و ۲۰ درصد از نمونه‌ها به‌عنوان

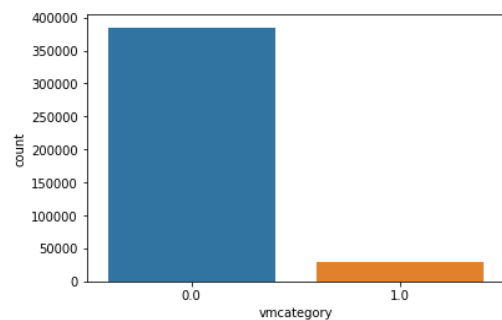
۴.۴. توصیف آزمایش‌ها

آزمایش‌ها با پیش‌پردازش داده انجام می‌شود. در گام اول نمونه‌هایی که دارای برچسب ناشناخته هستند از مجموعه داده حذف می‌شوند. شکل (۷) توزیع نمونه‌ها در کلاس‌های مختلف را نشان می‌دهد. تعداد کل نمونه‌ها ۱۰۴۸۵۷۶ بود که بعد از حذف نمونه‌های دارای برچسب ناشناخته، به ۴۱۵۲۳۸ نمونه کاهش می‌یابد.



شکل ۶. توزیع نمونه‌ها در کلاس‌های مختلف

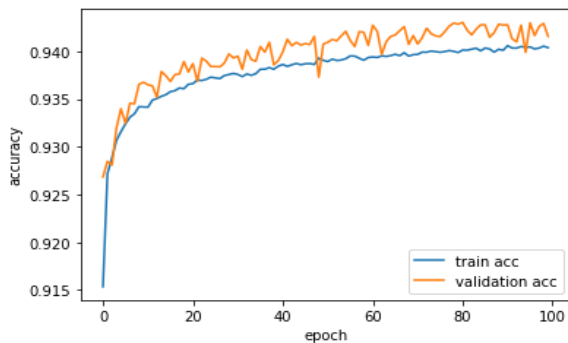
با توجه به اینکه در این مقاله از روش کدبندی برچسب به‌منظور تبدیل ویژگی‌ها استفاده شده است، بنابراین ستون حاوی برچسب مجموعه داده نیز با مقادیر صفر و یک جایگزین می‌شود. مقدار صفر مربوط به کلاس غیر حساس به تأخیر و کلاس ۱ متعلق به نمونه‌های دارای برچسب تعاملی (حساس به تأخیر) هستند. شکل (۸) توزیع نمونه‌ها بعد از حذف برچسب ناشناخته را نشان می‌دهد.



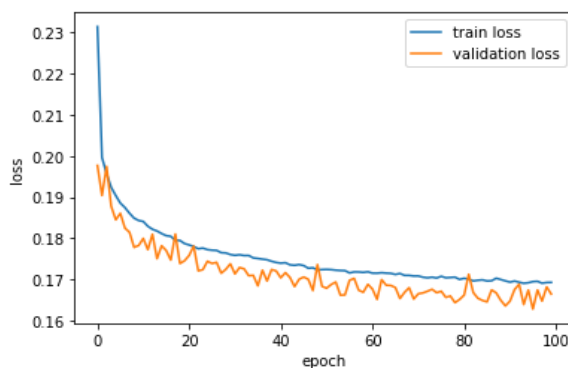
شکل ۷. توزیع نمونه‌ها بعد از حذف نمونه‌های برچسب ناشناخته

با توجه به شکل (۸) تعداد ۳۸۵۰۸۶ نمونه در کلاس غیر حساس به تأخیر و ۳۰۱۵۲ در کلاس تعاملی (حساس به تأخیر) قرار دارند. به‌منظور متوازن‌سازی مجموعه داده در این پژوهش از افزایش تصادفی نمونه‌ها در کلاس اقلیت استفاده می‌شود؛ بنابراین با توجه به افزایش نمونه‌ها در کلاس تعاملی، مجموع نمونه‌ها در مجموعه داده برابر با ۷۷۰۱۷۲ نمونه

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

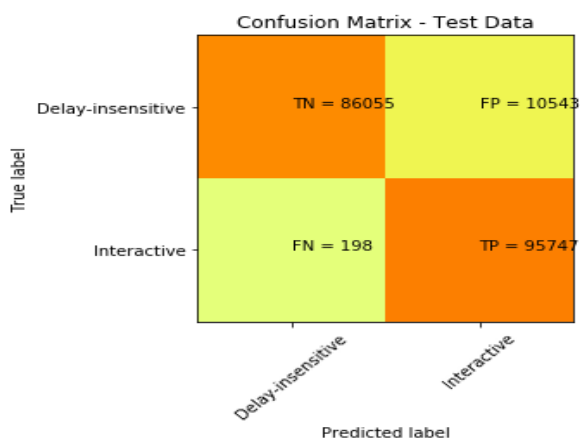


شکل ۱۰. نمودار همگرایی مدل بر اساس دقت



شکل ۱۱. نمودار همگرایی مدل بر اساس خطا

ماتریس آشفتگی برای دو کلاس تعاملی (حساس به تأخیر) و غیر حساس به تأخیر در شکل (۱۳) آمده است.



شکل ۱۲. ماتریس آشفتگی برای طبقه‌بندی دو کلاس

تعداد ۱۹۲۵۴۳ نمونه به عنوان مجموعه آزمایش در نظر گرفته شده است. با توجه به ماتریس آشفتگی، تعداد ۸۶۰۵۵ نمونه در کلاس غیر حساس به تأخیر و همچنین تعداد ۹۵۷۴۷ نمونه در کلاس تعاملی (حساس به تأخیر)، توسط مدل درست تشخیص داده شده‌اند. تعداد ۱۰۵۴۳ نمونه موجود در کلاس غیر حساس به تأخیر توسط مدل در کلاس تعاملی و به صورت

مجموعه آموزش در نظر گرفته شده‌اند. ابر پارامترها برای آموزش مدل در جدول (۵) آمده است.

جدول (۵): ابر پارامترها برای آموزش مدل

مقدار	ابر پارامتر
شبکه عصبی پیچشی و واحد برگشتی دروازه‌دار	نوع شبکه عصبی
سافت‌مکس	تابع دسته‌بند
ریلو	تابع فعال‌سازی
۶۴	تعداد فیلتر
۳	اندازه کرنل
۲	اندازه پولینگ
۶۴	تعداد واحد برگشتی دروازه‌دار
۱۲۸	تعداد نورون‌های لایه تمام متصل
۶۴	اندازه دسته‌ها
۰/۳	نرخ دراپ‌اوت
۱۰۰	تعداد دوره (ایپوک)
آدام ^{۴۶}	بهینه‌سازی وزن‌ها
۰/۰۱	نرخ یادگیری
آنترابی متقابل ^{۴۷}	تابع خطا

۵.۴. نتایج پیاده‌سازی

اجرای فرآیند آموزش مدل بر روی مجموعه داده آموزشی طی ۱۰۰ گام انجام می‌شود. تعداد نمونه‌های آموزشی ۵۱۹۸۶۶ و تعداد نمونه‌های اعتبار سنجی ۵۷۷۶۳ می‌باشد. نمودار همگرایی مدل روی داده آموزش و اعتبارسنجی بر اساس دقت در شکل (۱۱) و بر اساس خطا در شکل (۱۲) نشان داده شده است. استفاده از تکنیک حذف تصادفی از بیش‌برازش مدل جلوگیری کرده است و بیشترین میزان دقت بر روی داده‌های اعتبارسنجی در ایپوک ۸۴ اتفاق افتاده است.

۰/۹۶	۰/۸۷	مدل پیشنهادی+ کاهش نمونه‌های کلاس اکثریت	۶
۰/۹۵	۰/۶۴	مدل پیشنهادی+ بدون متوازن سازی	
۰/۹۷	۰/۸۸	مدل پیشنهادی+ افزایش نمونه‌های کلاس اقلیت	

مشابه تعداد ۱۹۸ نمونه متعلق به کلاس تعاملی در کلاس غیر حساس به تأخیر بر اساس پیش‌بینی مدل دسته‌بندی شده است. میزان دقت بر اساس مدل پیشنهادی و با استفاده از تکنیک افزایش تصادفی نمونه‌ها در کلاس اقلیت، برابر ۹۴/۷۹ می‌باشد. همچنین این مقدار با استفاده از تکنیک کاهش تصادفی نمونه‌ها در کلاس اکثریت و حالتی که نمونه‌ها به صورت نامتوازن در مجموعه داده وجود داشتند مقایسه شده است. جدول (۶) میزان دقت برحسب درصد را برای مدل پیشنهادی و سایر مدل‌ها نشان می‌دهد.

جدول ۸. مقایسه میزان فراخوانی برای هر دو کلاس

ردیف	نام مدل	میزان فراخوانی	
		تعاملی	غیر حساس به تأخیر
۱	لجستیک رگرسیون [۱۳]	۰/۸۵	۰/۵۹
۲	ماشین بردار پشتیبان [۱۴]	۰/۹۴	۰/۴۶
۳	شبکه عصبی مصنوعی [۳۸]	۰/۸۳	۰/۶۴
۴	شبکه عصبی پیچشی [۳۹]	۰/۸۴	۰/۸۴
۵	شبکه عصبی برگشتی [۱۶]	۰/۹۳	۰/۸۶
۶	مدل پیشنهادی+ کاهش نمونه‌های کلاس اکثریت	۰/۹۶	۰/۸۶
	مدل پیشنهادی+ بدون متوازن سازی	۰/۶۴	۰/۹۵
	مدل پیشنهادی+ افزایش نمونه‌های کلاس اقلیت	۰/۹۸	۰/۹۶

جدول ۶. مقایسه میزان دقت برحسب درصد

ردیف	نام مدل	میزان دقت برحسب درصد
۱	لجستیک رگرسیون [۱۳]	۷۰/۴۴
۲	ماشین بردار پشتیبان [۱۴]	۷۲/۳۴
۳	شبکه عصبی مصنوعی [۳۸]	۸۴/۰۶
۴	شبکه عصبی پیچشی [۳۹]	۹۱/۸۹
۵	شبکه عصبی برگشتی [۱۶]	۹۲/۸۸
۶	مدل پیشنهادی+ کاهش نمونه‌های کلاس اکثریت	۹۲/۷۶
	مدل پیشنهادی+ بدون متوازن سازی	۹۳/۱۶
	مدل پیشنهادی+ افزایش نمونه‌های کلاس اقلیت	۹۴/۴۲

همچنین جدول (۷) میزان صحت و جدول (۸) میزان فراخوانی را برای مدل پیشنهادی و سایر مدل‌ها نشان می‌دهد.

جدول ۷. مقایسه میزان صحت برای هر دو کلاس

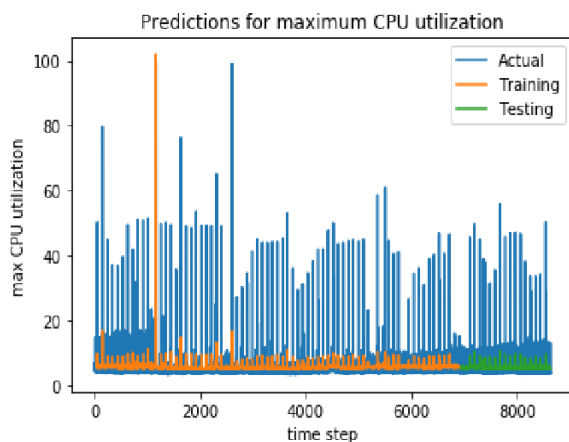
ردیف	نام مدل	میزان صحت	
		تعاملی	غیر حساس به تأخیر
۱	لجستیک رگرسیون [۱۳]	۰/۶۸	۰/۸۱
۲	ماشین بردار پشتیبان [۱۴]	۰/۶۴	۰/۸۹
۳	شبکه عصبی مصنوعی [۳۸]	۰/۷۱	۰/۷۹
۴	شبکه عصبی پیچشی [۳۹]	۰/۸۰	۰/۹۲
۵	شبکه عصبی برگشتی [۱۶]	۰/۸۴	۰/۹۳

به منظور ارائه تحلیل بهتری از پیچیدگی مدل پیشنهادی، تعداد دوره‌ها، تعداد پارامترهای قابل آموزش و زمان آموزش برای مدل پیشنهادی و سایر مدل‌ها در جدول (۹) نشان داده شده است.

جدول ۹. مقایسه دوره آموزش، مدت زمان اجرا و تعداد پارامترهای قابل آموزش

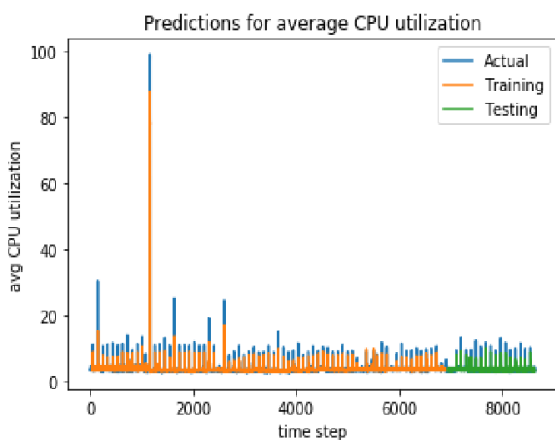
ردیف	نام مدل	ایپوک	پارامتر مدل	زمان آموزش (ثانیه)
۱	شبکه عصبی مصنوعی	۱۰۰	۱۶۸۹۶	۵۸۸/۳۱

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف



شکل ۱۵. پیش‌بینی بیشترین میزان مصرف پردازنده مرکزی با استفاده از مدل پیشنهادی

همچنین شکل (۱۶) پیش‌بینی میانگین مصرف پردازنده مرکزی را با استفاده از مدل پیشنهادی در گام‌های زمانی متفاوت بر روی مجموعه داده آموزش و آزمایش نشان می‌دهد.

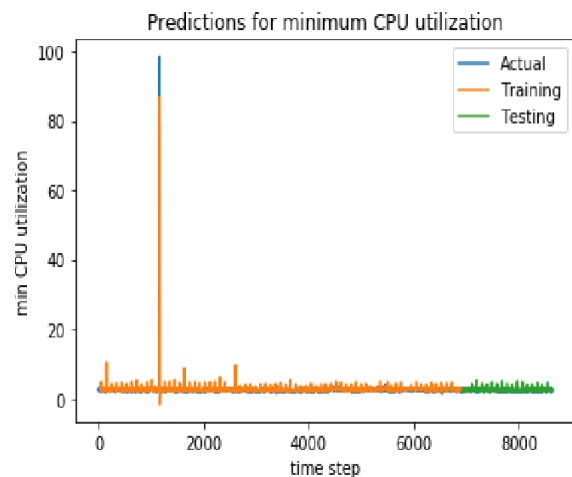


شکل ۱۶. پیش‌بینی میانگین میزان مصرف پردازنده مرکزی با استفاده از مدل پیشنهادی

با توجه به اینکه در مطالعات پیشین پیش‌بینی میانگین میزان مصرف پردازنده مرکزی به عنوان معیاری برای ارزیابی عملکرد مدل‌های پیشنهاد شده مورد بررسی قرار گرفته است [۱۳]، بنابراین در این مقاله برای مقایسه دقیق‌تر بین مدل‌های مبتنی بر رگرسیون و مدل پیشنهادی از میانگین میزان مصرف پردازنده مرکزی استفاده می‌شود. مقایسه میزان خطای جذر میانگین مربعات به منظور پیش‌بینی میانگین میزان مصرف

۹۴۶/۵۲	۱۷۱۵۴	۱۰۰	شبکه عصبی پیش‌بینی	۲
۹۶۷/۱۴	۲۰۹۶۰	۱۰۰	شبکه عصبی برگشتی	۳
۱۷۰۲/۴۳	۳۳۴۱۰	۱۰۰	(شبکه عصبی پیش‌بینی + حافظه کوتاه مدت طولانی)	۴
۱۵۷۷/۸۷	۲۵۳۴۶	۱۰۰	مدل پیشنهادی	۵

همچنین به منظور مقایسه دقیق‌تر بار کاری ماشین‌های مجازی در مجموعه داده میکروسافت آزور، پیش‌بینی میزان مصرف پردازنده مرکزی با استفاده از مدل پیشنهادی و سایر مدل‌های مبتنی بر رگرسیون مورد بررسی قرار گرفت. شکل (۱۴) پیش‌بینی کمترین میزان مصرف پردازنده مرکزی را نشان می‌دهد. همانطور که در شکل (۱۴) نشان داده شده است کمترین میزان مصرف پردازنده با استفاده از مدل پیشنهادی به خوبی پیش‌بینی شده است و دارای خطای کمی در داده آزمایش است. (۱۵) نیز نشان‌دهنده پیش‌بینی بیشترین میزان مصرف پردازنده مرکزی است. مدل پیشنهادی در پیش‌بینی بیشترین میزان مصرف پردازنده مرکزی عملکرد ضعیفی دارد.



شکل ۱۴. پیش‌بینی کمترین میزان مصرف پردازنده مرکزی با استفاده از مدل پیشنهادی

از این دو مدل در کنار هم این است که نقاط قوت و ضعف هر دو مدل بیان شده به منظور دسترسی به دقت بیشتر باهم همپوشانی دارند و استفاده از مزایای هر دو روش موجب هم‌افزایی^{۴۸} شده است.

(۴) شبکه های ژرف دارای چندین لایه ی پنهان جدا از لایه های ورودی و خروجی هستند و با استفاده از تعداد لایه های بیشتر موجب افزایش عملکرد در طبقه بندی می شوند، در حالی که شبکه های عصبی سطحی^{۴۹} از لایه های مخفی کمتری بین لایه های ورودی و خروجی استفاده می کنند [۴۰، ۴۱]. مدل پیشنهادی با استفاده از تعداد لایه های کمتر به دقت بالاتری دست یافته است و با توجه به این که از تعداد لایه های کمتری در معماری مدل پیشنهادی استفاده شده است. تعداد لایه های کمتر باعث کاهش قابل توجهی در پارامترهای قابل یادگیری شده است. این امر علاوه بر کاهش زمان آموزش، موجب کاهش پیچیدگی محاسباتی مدل نیز خواهد شد.

(۵) همان طور که قبلا بیان شد در این پژوهش از تکنیک افزایش تصادفی نمونه ها در کلاس اقلیت برای متوازن کردن مجموعه داده استفاده شده است. یکی از معایب این روش این است که موجب بیش برآزش در مدل می شود. زیرا در این روش نمونه های کلاس اقلیت کپی می شوند تا به تعداد نمونه ها در کلاس اکثریت برسند. برای غلبه بر این چالش از تکنیک دراپ اوت استفاده شده است. استفاده از تکنیک دراپ اوت در معماری مدل پیشنهادی علاوه کاهش بیش برآزش، موجب بهبود قابلیت تعمیم پذیری در مدل شده است که در نهایت منجر به افزایش پایداری مدل خواهد شد.

۵. نتیجه گیری

پیش بینی دقیق در انتخاب ماشین مجازی کاندید مهاجرت موجب کاهش مصرف انرژی در مراکز داده ابری خواهد شد. این کار از طریق طبقه بندی ماشین مجازی در کلاس حجم کار امکان پذیر خواهد بود. در این مقاله طبقه بندی ماشین های مجازی آزور با استفاده از ترکیب دو شبکه عصبی پیچشی و واحد برگشتی دروازه دار انجام شد. مدل پیشنهاد شده در این پژوهش با استفاده از مزایای هر کدام از مدل های بیان شده با دقت ۹۴/۴۲ موجب بهبود عملکرد در طبقه بندی شده است که نسبت به سایر مدل ها برتری دارد.

پردازنده برای مدل های مبتنی بر رگرسیون و مدل پیشنهادی در جدول (۱۰) نشان داده شده است.

جدول ۱۰. میزان خطای جذر میانگین مربعات برای مدل های مبتنی بر رگرسیون و مدل پیشنهادی

ردیف	نام مدل	میزان خطای جذر میانگین مربعات
۱	رگرسیون خطی	۱۴/۳۸
۲	رگرسیون لاسو	۱۴/۳۷
۳	رگرسیون سه تایی	۱۴/۶۴
۴	رگرسیون بردار پشتیبان	۱۵/۳۲
۵	مدل پیشنهادی	۳/۷۳

۶.۴. بحث

معیارهای اندازه گیری نشان دهنده برتری مدل پیشنهادی نسبت به مدل های دیگر می باشد. در بیان برتری مدل پیشنهادی می توان به موارد زیر اشاره کرد:

(۱) استفاده از تکنیک افزایش تصادفی نمونه ها در کلاس اقلیت به منظور مقابله با عدم توازن در مجموعه داده، موجب عملکرد بهتر در طبقه بندی شده است. زیرا در داده متوازن دقت مدل به سمت کلاس اکثریت منحرف نمی شود. در روش حذف تصادفی نمونه ها در کلاس اکثریت، نمونه های با ارزش از بین می روند. مزیت روش حذف تصادفی نمونه ها این است که زمان آموزش را کاهش می دهد، اما با توجه به اینکه مدل پیشنهادی دارای تعداد پارامترهای زیادی است، از این رو کاهش نمونه ها بر روی آموزش مدل تاثیر خوبی نخواهند داشت، بنابراین در این پژوهش از افزایش تصادفی نمونه ها در کلاس اقلیت استفاده شده است. یکی از مزیت های این روش پیاده سازی آسان می باشد.

(۲) استفاده از تکنیک های یادگیری ژرف در مقایسه با روش های کلاسیک یادگیری ماشین از عملکرد بهتری در پیش بینی بار کاری ماشین مجازی برخوردارند زیرا با استخراج ویژگی ها به صورت اتوماتیک، می تواند همبستگی چندگانه بین ماشین های مجازی را بر اساس حجم کار قبلی استخراج می کند و بار کاری آینده آن ها را با دقت بالا پیش بینی کند.

(۳) استفاده از یک مدل ترکیبی ژرف در این پژوهش می تواند دقت طبقه بندی را افزایش دهد. در واقع مزیت استفاده

پیش بینی بار کاری ماشین‌های مجازی به منظور کاهش مصرف انرژی در مراکز داده ابری با استفاده از ترکیب مدل‌های یادگیری ژرف

مراجع

- [1] A. Yousafzai *et al.*, "Cloud resource allocation schemes: review, taxonomy, and opportunities," *Knowledge and Information Systems*, vol. 50, no. 2, pp. 347-381, 2017.
- [2] I. Hamzaoui, B. Duthil, V. Courboulay, and H. Medromi, "A Survey on the Current Challenges of Energy-Efficient Cloud Resources Management," *SN Computer Science*, vol. 1, no. 2, pp. 1-28, 2020.
- [3] A. Beloglazov, "Energy-efficient management of virtual machines in data centers for cloud computing," 2013.
- [4] S. Singh and I. Chana, "A survey on resource scheduling in cloud computing: Issues and challenges," *Journal of grid computing*, vol. 14, no. 2, pp. 217-264, 2016.
- [5] M. H. Sayadnavard, A. T. Haghghat, and A. M. Rahmani, "A reliable energy-aware approach for dynamic virtual machine consolidation in cloud data centers," *The Journal of Supercomputing*, vol. 75, no. 4, pp. 2126-2147, 2019.
- [6] J. N. Witanto, H. Lim, and M. Atiquzzaman, "Adaptive selection of dynamic VM consolidation algorithm using neural network for cloud resource management," *Future generation computer systems*, vol. 87, pp. 35-42, 2018.
- [7] H. Sadr, M. M. Pedram, and M. Teshnehlalab, "Multi-View Deep Network: A Deep Model Based on Learning Features From Heterogeneous Neural Networks for Sentiment Analysis," *IEEE Access*, vol. 8, pp. 86984-86997, 2020.
- [8] H. Sadr and M. Teshnehlalab, "Efficient Method Based on Combination of Deep Learning Models for Sentiment Analysis of Text," *Signal and Data*

جمع‌آوری داده‌های واقعی از حجم کار ماشین مجازی امکان عدم توازن در داده را در پی خواهد داشت. کار با داده‌های نامتوازن مشکلاتی مانند مدل جانب‌دارانه و رده‌بندی اشتباه را به وجود می‌آورد. به همین دلیل در این پژوهش از تکنیک افزایش تصادفی نمونه‌ها در کلاس اقلیت به منظور مقابله با عدم توازن نمونه‌ها در دو کلاس استفاده شده است. این تکنیک می‌تواند با افزایش نمونه‌های آموزشی با توجه به افزایش تعداد پارامترهای مدل موجب بهبود عملکرد طبقه‌بندی شود؛ اما از طرفی افزایش تعداد نمونه‌ها زمان آموزش را افزایش خواهد داد.

پژوهش حاضر بر پیش‌بینی بار کاری ماشین مجازی تاکید دارد و پیش‌بینی بار کاری ماشین مجازی یک گام اولیه در مدیریت منابع به خصوص ادغام پویای ماشین مجازی می‌باشد. با توجه به اینکه ادغام پویای ماشین مجازی شامل ۴ مرحله اساسی است که عبارتند از تشخیص سرور فروار، شناسایی سرور فرابار، انتخاب ماشین مجازی و جایابی ماشین مجازی. از طرفی هر کدام از مراحل می‌توانند به عنوان پژوهشی مستقل بررسی شود و از آنجایی که در این پژوهش تنها پیش‌بینی بار کاری به منظور انتخاب مناسب ماشین مجازی مورد بررسی قرار گرفته است، بنابراین در آینده با ارائه الگوریتم مناسب برای مراحل دیگر، تعداد مهاجرت‌ها و میزان مصرف انرژی در یک مرکز داده مجازی مورد آزمایش قرار خواهد گرفت.

استفاده از مدل پیشنهادی در طبقه‌بندی ماشین مجازی می‌تواند در کاربردهای دیگری همچون امنیت و مدیریت درخواست کاربر نیز استفاده شود. همچنین مدل پیشنهادی را می‌توان برای شناسایی سرورهای فرابار و فروار در مراکز داده ابری به کار برد.

علاوه بر موارد ذکر شده، می‌توان از سایر مجموعه‌های داده برای ارزیابی مدل پیشنهادی استفاده کرد. همچنین می‌توان از ترکیبی از سایر مدل‌های شبکه عصبی برای طبقه‌بندی استفاده نمود. همانطور که قبلاً ذکر شد، مجموعه داده‌های مورد استفاده در این مقاله نامتعادل است، بنابراین برای غلبه بر این چالش، می‌توان از شبکه‌های مولد متخاصم^{۵۰} (GAN) برای افزایش نمونه‌های کلاس اقلیت استفاده کرد.

- [16] W. Zhang, B. Li, D. Zhao, F. Gong, and Q. Lu, "Workload prediction for cloud cluster using a recurrent neural network," in *2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI)*, 2016, pp. 104-109: IEEE.
- [17] H. Sadr and M. Nazari Soleimandarabi, "ACNN-TL: attention-based convolutional neural network coupling with transfer learning and contextualized word representation for enhancing the performance of sentiment classification," *The Journal of Supercomputing*, vol. 78, no. 7, pp. 10149-10175, 2022.
- [18] Y. Zhu, W. Zhang, Y. Chen, and H. Gao, "A novel approach to workload prediction using attention-based LSTM encoder-decoder network in cloud environment," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 274, 2019.
- [19] S. Ouhamme and Y. Hadi, "Multivariate workload prediction using Vector Autoregressive and Stacked LSTM models," in *Proceedings of the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society*, 2019, pp. 1-7.
- [20] Z. Chen, J. Hu, G. Min, A. Y. Zomaya, and T. El-Ghazawi, "Towards accurate prediction for high-dimensional and highly-variable cloud workloads with deep learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 4, pp. 923-934, 2019.
- [21] P. Yazdani and S. Sharifian, "Cloud Workload Prediction Using ConvNet And Stacked LSTM," in *2018 4th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS)*, 2018, pp. 83-87: IEEE.
- [9] Y. S. Patel and R. Misra, "Performance comparison of deep VM workload prediction approaches for cloud," in *Progress in Computing, Analytics and Networking: Springer*, 2018, pp. 149-160.
- [10] H. Khani and H. Khanmirza, "Randomized routing of virtual machines in IaaS data centers," *PeerJ Computer Science*, vol. 5, p. e211, 2019.
- [11] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," *Future generation computer systems*, vol. 28, no. 5, pp. 755-768, 2012.
- [12] R. Moreno-Vozmediano, R. S. Montero, E. Huedo, and I. M. Llorente, "Efficient resource provisioning for elastic Cloud services based on machine learning techniques," *Journal of Cloud Computing*, vol. 8, no. 1, p. 5, 2019.
- [13] M. Hariharasubramanian, "Improving application infrastructure provisioning using resource usage predictions from cloud metric data analysis," *Rutgers University-School of Graduate Studies*, 2018.
- [14] R. Shaw, E. Howley, and E. Barrett, "An intelligent ensemble learning approach for energy efficient and interference aware dynamic virtual machine consolidation," *Simulation Modelling Practice and Theory*, vol. 102, p. 101992, 2020.
- [15] N. Verma and A. Sharma, "Workload prediction model based on supervised learning for energy efficiency in cloud," in *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*, 2017, pp. 66-71: IEEE.

- mechanism and transfer learning for enhancing performance of sentiment analysis," *Journal of AI and data mining*, vol. 9, no. 2, pp. 141-151, 2021.
- [30] E. Jackson and R. Agrawal, "Performance Evaluation of Different Feature Encoding Schemes on Cybersecurity Logs," in *2019 SoutheastCon*, 2019, pp. 1-9: IEEE.
- [31] M. P. Kalashami, M. M. Pedram, and H. Sadr, "EEG Feature Extraction and Data Augmentation in Emotion Recognition," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [32] H. Sadr, M. M. Pedram, and M. Teshnehlab, "A robust sentiment analysis method based on sequential combination of convolutional and recursive neural networks," *Neural Processing Letters*, vol. 50, no. 3, pp. 2745-2761, 2019.
- [33] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.
- [34] Y. Wang, M. Liu, Z. Bao, and S. Zhang, "Short-term load forecasting with multi-source data using gated recurrent unit neural networks," *Energies*, vol. 11, no. 5, p. 1138, 2018.
- [35] S. Wager, S. Wang, and P. S. Liang, "Dropout training as adaptive regularization," *Advances in neural information processing systems*, vol. 26, 2013.
- [36] H. Sadr, M. M. Pedram, and M. Teshnelab, "Improving the performance of text sentiment analysis using deep convolutional neural network integrated with hierarchical attention layer," *International journal of information and communication technology research*, vol. 11, no. 3, pp. 57-67, 2019.
- [22] P. Yazdanian and S. Sharifian, "E2LG: a multiscale ensemble of LSTM/GAN deep learning architecture for multistep-ahead cloud workload prediction," *The Journal of Supercomputing*, pp. 1-31, 2021.
- [23] J. Bi, S. Li, H. Yuan, and M. Zhou, "Integrated deep learning method for workload and resource prediction in cloud systems," *Neurocomputing*, vol. 424, pp. 35-48, 2021.
- [24] S. Shishira and A. Kandasamy, "A Novel Feature Extraction Model for Large-Scale Workload Prediction in Cloud Environment," *SN Computer Science*, vol. 2, no. 5, pp. 1-7, 2021.
- [25] F. Mohades Deilami, H. Sadr, and M. Tarkhan, "Contextualized Multidimensional Personality Recognition using Combination of Deep Neural Network and Ensemble Learning," *Neural Processing Letters*, pp. 1-18, 2022.
- [26] Z. Khodaverdian, H. Sadr, and S. A. Edalatpanah, "A shallow deep neural network for selection of migration candidate virtual machines to reduce energy consumption," in *2021 7th International Conference on Web Research (ICWR)*, 2021, pp. 191-196: IEEE.
- [27] B. Santoso, H. Wijayanto, K. Notodiputro, and B. Sartono, "Synthetic over sampling methods for handling class imbalanced problems: a review," in *IOP conference series: earth and environmental science*, 2017, vol. 58, no. 1, p. 012031.
- [28] S. Soleymanpour, H. Sadr, and M. Nazari Soleimandarabi, "CSCNN: cost-sensitive convolutional neural network for encrypted traffic classification," *Neural Processing Letters*, vol. 53, no. 5, pp. 3497-3523, 2021.
- [29] H. Sadr, M. M. Pedram, and M. Teshnehlab, "Convolutional neural network equipped with attention

- Sciences: Springer, 2019, pp. 139-154.
- [40] M. A. Wani, F. A. Bhat, S. Afzal, and A. I. Khan, *Advances in deep learning*. Springer, 2020.
- [41] A. Plebe and G. Grasso, "The unbearable shallow understanding of deep learning," *Minds and Machines*, vol. 29, no. 4, pp. 515-553, 2019.
- [37] E. Cortez, A. Bonde, A. Muzio, M. Russinovich, M. Fontoura, and R. Bianchini, "Resource central: Understanding and predicting workloads for improved resource management in large cloud platforms," in *Proceedings of the 26th Symposium on Operating Systems Principles*, 2017, pp. 153-167.
- [38] E. Patel, A. Mohan, and D. S. Kushwaha, "Neural network based classification of virtual machines in IaaS," in *2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, 2018, pp. 1-8: IEEE.
- [39] A. M. Aslam and M. Kalra, "Using Artificial Neural Network for VM Consolidation Approach to Enhance Energy Efficiency in Green Cloud," in *Advances in Data and Information*

پانویس ها

- 27 Max Pooling
- 28 Feature map
- 29 Flatten
- 30 Reset Gate
- 31 Update Gate
- 32 Overfitting
- 33 Dropout
- 34 Feedforward
- 35 Backpropagation
- 36 Loss Function
- 37 <https://github.com/Azure/AzurePublicDataset>
- 38 Infrastructure as a Service(IaaS)
- 39 Platform as a Service (PaaS)
- 40 Cumulative Distribution Function (CDF)
- 41 Accuracy
- 42 Precision
- 43 Recall
- 44 VM Lifetime
- 45 VM Core Hour
- 46 Adam
- 47 Cross-Entropy
- 48 Synergy
- 49 Shallow
- 50 Generative Adversarial Networks (GANs)
- 1 Convolution Neural Network (CNN)
- 2 Gated Recurrent Unit(GRU)
- 3 Random Over-Sampling(ROS)
- 4 <https://goo.gl/Jez9Kg>
- 5 Ridge Regression
- 6 Ensemble learning
- 7 Boosted Tree
- 8 Recurrent Neural Network (RNN)
- 9 Autoregressive Integrated Moving Average (ARIMA)
- 10 Long Short-Term Memory (LSTM)
- 11 Attention
- 12 Vector Autoregressive (VAR)
- 13 Augmented Dickey-Fuller (ADF)
- 14 Sparse Autoencoder
- 15 Generative Adversarial Networks (GAN)
- 16 Generator
- 17 Discriminator
- 18 Savitzky-Golay filter
- 19 Bidirectional LSTM
- 20 Grid LSTM
- 21 Random Under-Sampling (RUS)
- 22 Cost-Sensitive
- 23 Label Encoding
- 24 Data Aggregation
- 25 Min-Max normalization
- 26 Classifier