ارائه روشی برای پیش‌بینی خطای محتوا درمحاسبات ابری با روش نزدیکترین همسایه

رضا ایمانی

چکیده

رایانش ابری یک تکنولوژی جدید برای مدیریت و ارائه خدمات بر روی اینترنت است. رایانش ابري، گروهي بزرگ از كامپيوترهاي متصل به هم است. اين كامپيوترها مي‌توانند رايانه‌هاي شخصي يا سرورهاي شبكه باشند. همچنين آن‌ها مي‌توانند عمومي يا خصوصي باشند. پیش‌بینی خطای محتوا یکی از چالش‌های اساسی در محاسبات ابری می‌باشد. در این پژوهش سعي بر اين است، با استفاده از الگوريتم نزديكترين همسايه دقت پيش-بيني خطای محتوا در محاسبات ابری افزایش داده شود. نتایج شبیه‌سازی در نرم‌افزار متلب نشان می‌دهد که روش براي مجموعه داده داراي كارايي بسيار بالايي مي‌باشد. نتایج را با عملکرد سه شبکه عصبی پرسپترون، بر اساس عملكرد شعاعي و نزديكترين همسايه مقايسه شده است و در نهایت، بهترین ترکیب در بین این سیستم‌ها، سیستم نزدیکترین همسایه شناسایی خطا را با دقت نزدیک ۷٤ درصد به دست داد.

کلید واژه : رایانش ابری، نزدیکترین همسایه، پیش‌بینی خطای محتوا، تابع عملکرد شعاعي

مقدمه

با رشد محبوبیت اینترنت و وب، به موازات دسترسی پردازش دستی قدرتمند، سیار و ابزارهای حسگر، مسیر فعل و انفعال،

اداره کردن زندگی، مدیریت تجارت و دسترسی یا دریافت خدمات را عوض کرده است. کاهش هزینه‌های محاسبات و ارتباطات توجه را از شخصی بودن به محاسبات مرکزداده[1] محوری جلب می کند. همچنین محاسبات توزیع‌شده و موازی حدود چندین سال است که وجود دارند که شکل‌های جدید آن رایانش ابری[2]، و چند هسته[3]تغییرات جامعی را در صنعت به همراه داشته است. این گرایشات کانون توجه صنعت را از توسعه نرم‌افزارها برای کامپیوترهای شخصی به مراکز داده ابری که میلیون‌ها کاربر را جهت استفاده از نرم افزار به طور همزمان، سوق می‌دهد[۱].

محاسبات ابری مدلی است برای دسترسی به یک شبکه یکپارچه و مناسب و برحسب تقاضا جهت به اشتراک‌گذاری حجم زیادی از منابع محاسباتی قابل تنظیم (مانند شبکه، سرورها، ذخیره‌گاه‌ها، برنامه‌ها و سرویس‌ها) که می‌تواند با کمترین نیاز به مدیریت و تنها به‌وسیله یک فراهم‌کننده سرویس ابر به‌سرعت قانونمند شده و انتشار یابد[۲]. رایانش ابری یک سبک ابتکاری و مهیج برای برنامه‌نویسی و استفاده از رایانه‌ها می‌باشد. رایانش ابری فرصت‌های فوق‌العاده‌ای را برای توسعه دهندگان نرم‌افزار ایجاد می‌کند[۳]. این معماری جدید از پهنای باند به صورت بهینه استفاده می‌کند و هزینه‌ی ارتباطی برای دسترسی به اطلاعات پرطرفدار را کاهش می‌دهد. این شبکه همچنین ازدحام در شبکه را کاهش داده و پیش‌بینی خطا، تحویل داده و تحمل قطعی بهتری را فراهم می‌کند[٤].

با پیشرفت محاسبات ابری، پیش‌بینی خطای داده‌ها به امری مهم در محاسبات ابری تبدیل شده است، بطوریکه پیش‌بینی خطاهای ابری مهمترین مانع سرعت و گسترش نرم‌افزارهای محاسبات ابری می‌باشد به همین دلیل پیش‌بینی خطا یکی از فاکتورهای کلیدی برای موفقیت ابر می‌باشد[٥]. رشد فناوری اطلاعات و ارتباطات، تغییرات پرشتاب محیط کسب و کار و همچنین گسترش روزافزون سرویس‌های ذخیره‌سازی ابری، شرکت‌ها و کاربران این عرصه را تشویق نموده تا نگهداری و مدیریت داده‌های خود را به ارائه دهندگان خدمات فضای ذخیره‌سازی ابری واگذار نمایند. اما نگرانی از پیش‌بینی خطا و حفظ حریم شخصی کاربران، همواره به عنوان چالشی مهم و بحث برانگیز در ابر مطرح است[٦].

به طور کلی، پردازش ابری با چندین خطر پیش‌بینی خطای مجزا، اما مرتبط با هم مواجه است. علاوه بر این که ممکن است اطلاعات ذخیره شده، توسط مهاجمان به سرقت رفته یا در اثر خرابی سیستم از بین برود، ممکن است تأمین کننده خدمات ابری نیز از این اطلاعات سوء استفاده کند یا در اثر فشارهای دولت آن را افشا کند واضح است که این خطرات پیش‌بینی خطا پیامدهای وخیمی را در پی دارند[۷]. تحقیقاتی نیز در این زمینه انجام شده است. در[۸] برای محافظت از محیط محاسبات

---

[1] Data Center
[2] Cloud Computing
[3] Multicore

ابری یک قالب برای استفاده از سیستم تشخیص نفوذ[۴] طراحی گردید که در آن بتوان حملات از نوع DDoS را به‌صورت متمرکز مدیریت نمود. آن‌ها بر این عقیده بودند که اگر به ازای هر ماشین مجازی یک سیستم تشخیص نفوذ استفاده شود می‌توان مدیریت حملات را به‌خوبی انجام داد و با تبادل اطلاعات بین آن‌ها به حملات پاسخ دهند. در[۷] به طراحی یک معماری مدیریتی سیستم تشخیص نفوذ توسعه‌یافته پرداخته‌شده است که در آن سیستم تشخیص نفوذ در کل سیستم توزیع گردیده است که همگی زیر نظر یک مرکز کنترل و واحد مدیریت می باشند. در این تحقیق سنسورهایی به‌عنوان سیستم تشخیص نفوذ در سیستم فراهم‌کننده خدمات و کاربران کار گذاشته می‌شود که به‌محض بروز هر نوع ناهنجاری به مرکز کنترل و مدیریت گزارش داده می‌شوند و مرکز در مورد آن تصمیم‌گیری می‌کند. روش دیگری که برای محافظت از محیط ابری طراحی شد در[۹] می‌باشد. در این مقاله حملات با توجه به میزان خسارت و همچنین سطح حملات آن‌ها در محیط ابری دسته‌بندی شدند بر این اساس برای هر حمله یک ریسک خطر تهیه و به آن‌ها داده شد و بر اساس آن روشی برای مقابله با آن در نظر گرفته‌شده است.

در[۱۰] نیز همانند[۹] از روش تعیین میزان ریسک برای منابع استفاده‌شده است و در سیستم از یک سیستم تشخیص نفوذ مرکزی استفاده گردیده است. مقاله[۱۱] به طراحی یک سیستم امنیتی برای تبادل امن اطلاعات با استفاده از احراز هویت و رمزنگاری پرداخته است. در سیستم پیشنهادشده دو ابر(دوقلو) در نظر گرفته‌شده است. یک ابر معتبر[۵] و یک ابر ارتباطی[۶]. یکی از روش های سنجش موفقیت سیستم ابر روش K نزدیکترین همسایه است. روشK نزدیکترین همسایه[۷] یک گروه شامل Kرکورد از مجموعه رکوردهای آموزشی که نزدیکترین رکوردها به رکورد آزمایشی باشند را انتخاب کرده و بر اساس برتری رده یا برچسب مربوط به آن‌ها در مورد دسته رکورد آزمایشی مزبور تصمیم‌گیری می‌نماید. به عبارت ساده‌تر این روش رده‌ای را انتخاب می‌کند که در همسایگی انتخاب شده بیشترین تعداد رکورد منتسب به آن دسته باشند. بنابراین رده‌ای که از همه رده‌ها بیشتر در بینK نزدیکترین همسایه مشاهده شود، به عنوان رده رکورد جدید در نظر گرفته می‌شود[۱۲]. با توجه به مطالب فوق پیش‌بینی خطای محتوای داده و حفظ حریم خصوصی که از چالش‌های پیش‌بینی خطا و مسئله مهم در محیط محاسبات ابری هستند لذا در این پژوهش سعي داریم، با استفاده از الگوریتم نزدیکترین همسایه خطا ی محتوا در محاسبات ابری را پیش‌بینی کنیم.

روش پژوهش

الگوریتم نزدیک‌ترین همسایه

یکي از الگوریتم‌هاي معروف دسته‌بندي، الگوریتم نزدیک همسایه

[۴] IDS

[۵] Trust

[۶] Commodity

[۷] – K Nearest Neighbor

است؛ با اینکه از معرفی آن چندین دهه می‌گذرد. این روش برای آموزش، مدلی را بر روی نمونه‌ها فرض نمی‌کند. همچنین این روش نیازی به تنظیم پارامتر ورودی توسط فرد متخصص ندارد. در واقع این روش زمانی را برای آموزش صرف نکرده و تا آمدن درخواست برای برچسب نمونه‌ی بدون برچسب صبر می‌نماید. بنابراین این روش هزینه زمان برای فاز آموزش ندارد و در مقابل به زمان بیشتری در فاز تست نیاز دارد. همچنین برای تعیین مقدار شباهت بین نمونه‌ها، نیاز به تعیین یک معیار فاصله است. معیارهای فاصله مختلف در فضای ویژگی‌های متفاوت، کارایی‌های متفاوتی از خود نشان می‌دهند [۱۳].

اولین نکته قابل توجه در مورد الگوریتم نزدیک‌ترین همسایه، نحوه به دست آوردن فاصله بین نمونه‌ها است. معیارهای فاصله مختلفی وجود دارد که می‌توان در این روش استفاده کرد. همچنین در بعضی از کارهای گذشته، معیارهای خاص این روش نیز معرفی‌شده‌اند که با استفاده از آن‌ها می‌توان کارایی بهتری را به دست آورد. در ادامه، چند معیار فاصله پرکاربرد معرفی‌شده‌اند [۱۴].

فاصله اقلیدسی از گذشته، پرکاربردترین معیار فاصله در ضمینه‌های مختلف بوده است. همچنین در دسته‌بند نزدیک‌ترین همسایه سنتی نیز این معیار مورد استفاده شده است. معیار فاصله اقلیدسی بین دو نمونه‌ی Y و X به‌صورت زیر تعریف می‌شود:

(۱)
$$d(X,Y)=\sqrt{\sum_{i=1}^{m}(x_i-y_i)^2}$$

که در آن m تعداد ویژگی‌های نمونه‌ها و x_i و y_i مقدار ویژگی iام دو نمونه است.

واضح است که این معیار زمانی قابل‌استفاده و کاراست که در فضای ویژگی‌ها، تمام ویژگی‌ها عددی و مقادیر حقیقی بوده و دارای بازه تغییرات مشابه باشند. در مقابل در حالتی که ویژگی‌ها دارای مقادیر صحیح باشند، معیار فاصله منهتن[۸] کارایی بهتر را از خود نشان می‌دهد. این معیار فاصله به‌صورت زیر تعریف می‌شود:

(۲)
$$d(X,X)=\sum_{i=1}^{m}|x_i-y_i|$$

که همانند معیار فاصله اقلیدسی، m تعداد ویژگی‌های نمونه‌ها و x_i و y_i مقدار ویژگی iام دو نمونه است.

معیار فاصله ماهالانوبیس[۹] یکی دیگر از معیارهای فاصله‌ی پرکاربرد و معرف است. در این معیار، علاوه بر مقادیر ویژگی‌ها، چگونگی پخشش نمونه‌ها در فضای ویژگی در نظر گرفته می‌شود. این پخشش، توسط ماتریس کواریانس[۱۰] در تابع معیار فاصله تأثیر می‌گذارد. این معیار به‌صورت زیر تعریف می‌شود:

---

[۸] Manhattan

[۹] Mahalanobis

[۱۰] Covariance Marix

(۳)

$$d(X,X)=\sqrt{(\vec{x}-\vec{y})^T S^{-1}(\vec{x}-\vec{y})}$$

که در آن $\vec{x}$ و $\vec{y}$ دو بردار ویژگی و S ماتریس کواریانس ویژگی‌ها است.

در مسائل دنیای واقعی، مقادیر ویژگی‌ها ممکن است که همیشه مقدار حقیقی نداشته باشند. در بسیاری از مسائل، بعضی از ویژگی‌ها دارای مقدار نامی می‌باشند. در این‌گونه ویژگی‌ها، برای دو مقدار، رابطه بزرگتری و کوچکتری و اختلاف دو مقدار مفهومی ندارد و تعریف نمی‌شود. در این‌گونه مقادیر، تنها تفاوت و یا شباهت دو مقدار، تفاوت بین دو نمونه را نشان می‌دهد. برای چنین فضای ویژگی که هم دارای ویژگی‌های نامی هست و هم مقادیر حقیقی، *Wilson* و *Martinez* معیاری به نام معیار فاصله مقادیر ناهمگن[۱۱] (*HVDM*) را ارائه داده‌اند. این معیار، توسعه‌یافته معیاری با نام معیار فاصله مقادیر[۱۲] (*VDM*) است که قبل از آن ارائه‌شده بود.

در *VDM* تفاوت بین دو مقدار a و b برای یک ویژگی f به‌صورت زیر تعریف می‌شود:

(٤)

$$vdm_f(X,Y)=\sum_{c=1}^{C}\left(\frac{N_{f,a,c}}{N_{f,a}}-\frac{N_{f,b,c}}{N_{f,b}}\right)^2$$

که در آن، $N_{f,a}$ برابر تعداد تکرار مقدار a در ویژگی f است و $N_{f,a,c}$ همین تکرار است در نمونه‌هایی که دارای برچسب c هستند.

بر اساس تعریف *VDM* معیار *HVDM* به‌صورت زیر تعریف می‌شود:

(٥)

$$d(X,Y)=\sqrt{\sum_{f=1}^{m} d_f^2(x_f,y_f)}$$

که در آن:

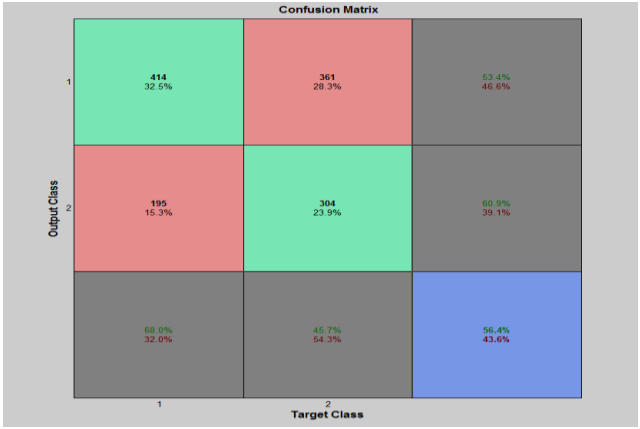
(٦)

$$d_f(x_f,y_f)=\begin{cases}1 & \text{if x or y is unknown}\\ vdm_f(x,y) & \text{if f is a nominal}\\ \dfrac{|x-y|}{4\sigma_f} & \text{if f is numeric}\end{cases}$$

یافته ها

۱. عملکرد شبکه عصبی پرسپترون

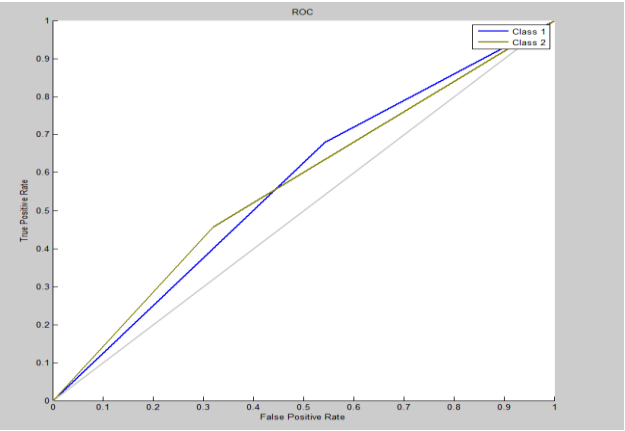
برای این سیستم در این تحقیق از تابع آماده متلب استفاده کرده ایم.



[۱۱] Heterogeneous Value Difference Metric (HVDM)

[۱۲] Value Difference Metric (VDM)

نمودار۱:کانفیوژن برای شبکه عصبی پرسپترون



نمودار۲: مشخصه عامل گیرنده برای شبکه عصبی پرسپترون

همچنین نتایج بدست آمده از مقایسات کلی در جدول زیر آورده شده است.

جدول۱:نتایج بدست آمده از ترکیب شبکه عصبی پرسپترون

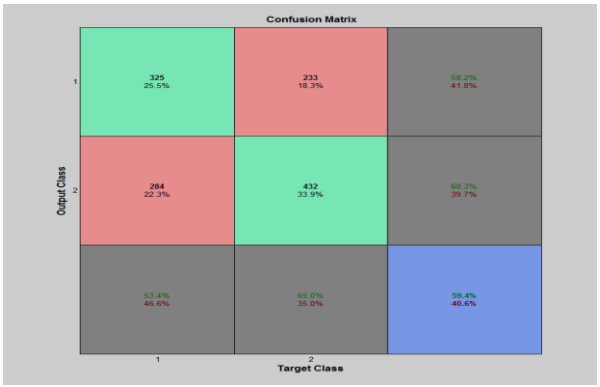
MEAN		k¹	k²	k³	k٤
۰,۲۰٤٦۹۳	نرخ کلاس بندی اشتباه	۰,۲٥	۰,۱۳۱٥۷۸۹	۰,۲۱۰٥۲٦	۰,۲۲٦٦٦۷
۰,۲۸۷۱۲	میزان هزینه کلاس بندی اشتباه	۰,۳٥۹۱۳۹۷۸	۰,۲۱٦۸٤٥۹	۰,۳٤۲٥۰۹	۰,۲۲۹۹۸٤
۰,۱٤۳٥٦	نرمال شده میزان هزینه کلاس بندی اشتباه	۰,۱۷۹٥٦۹۸۹	۰,۱۰۸٤۲۲۹	۰,۱۷۱۲٥٤	۰,۱۱٤۹۹۲
۰,۷۸٦۳٤۲	حساسیت	۰,۷۳۳۳۳۳۳۳	۰,۸۸۸۸۸۸۹	۰,۸۲۹۲٦۸	۰,٦۹۳۸۷۸
۰,۸۱۹۷۰۹	نرخ اختصاصی	۰,۷۷٤۱۹۳٥٥	۰,۸۳۸۷۰۹۷	۰,۷٤۲۸٥۷	۰,۹۲۳۰۷۷
۰,۷۰۹۱۲٥	دقت	۰,۷۱	۰,٦۹	۰,۷۲	۰,۷۱٦٥
۰,۸٦۲۲٥۸	صحت	۰,۸۲٥	۰,۸۸۸۸۸۸۹	۰,۷۹۰٦۹۸	۰,۹٤٤٤٤٤
۰,۷۸٦۳٤۲	فراخوانی مجدد	۰,۷۳۳۳۳۳۳۳	۰,۸۸۸۸۸۸۹	۰,۸۲۹۲٦۸	۰,٦۹۳۸۷۸
۰,۸۱۸۷۲۱	معیار ترکیبی F	۰,۷۷٦٤۷۰٥۹	۰,۸۸۸۸۸۸۹	۰,۸۰۹٥۲٤	۰,۸
۰,٤٥٥۰۱	سازگاری	۰,۳٤٦۲۳٦٥	۰,۷۲۷٥۹۸	۰,٦۲۹۲٦	۰,۱۱٦۹٥

٤		٦	٦	٨	٤
٠,٨٠٣٠٢٦	AUC	٠,٧٥٣٧٦٣٤٤	٠,٨٦٣٧٩٩٣	٠,٧٨٦٠٦٣	٠,٨٠٨٤٧٧
٠,٧٩٣٢٤٤	معیار توازن	٠,٧٥٢٩١٧٣٦	٠,٨٦١٥٠٧٧	٠,٧٨١٧٤٤	٠,٧٧٦٨٠٩
٤,٠٨٠٣٣١	زمان اجرا	٤,٢٠٢٦٩٧٨٦	٤,٣١٨٤٤٢٨	٣,٥٩١٤٤١	٤,٢٠٨٧٤٢

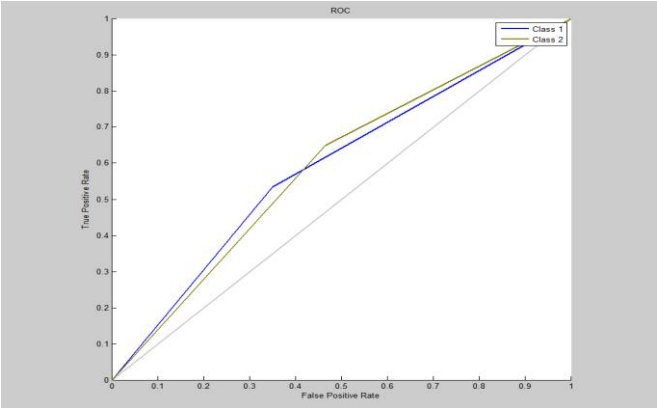
همان طور که از نتایج پیدا است در هر ٤ اجرا نتایج ثبات شبکه عصبی پرسپترون را نمایش می دهد و در این اجرای سوم بهترین نتیجه را به همراه داشته است. بنابراین تا این مرحله بهترین دقت بدست آمده ٠,٧٠٩١ است.

٢. عملکرد سیستم RBF[١٣]

برای سیستم RBF در این تحقیق از تابع آماده متلب استفاده کرده ایم. در این تابع شعاع پوشش برابر ٢ قرار داده ایم. نتایج بدست آمده به شرح زیر است.



نمودار٣:کانفیوژن برای سیستم RBF



نمودار٣:مشخصه عامل گیرنده برای سیستم RBF

١- تابع عملکرد شعاعی

جدول۲: نتایج بدست آمده از ترکیب سیستم RBF

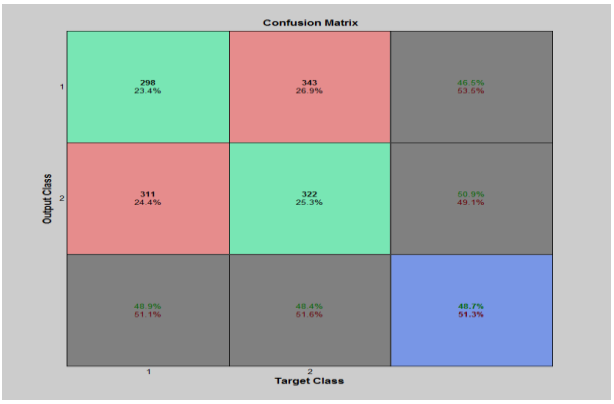
MEAN		k۱	k۲	k۳	k٤
۰,٤۳۲٥٤٤	نرخ کلاس بندی اشتباه	۰,٤۳٤۲۱۰٥۳	۰,۳۹٤۷۳٦۸	۰,٤۰۷۸۹٥	۰,٤۹۳۳۳۳
۰,٥۹٥۰۸٦	میزان هزینه کلاس بندی اشتباه	۰,٦۲۱۳۱۱٤۸	۰,٦٤٦۸٥۳۱	۰,٤٥۸۳۳۳	۰,٦٥۳۸٤٦
۰,۲۹۷٥٤۳	نرمال شده میزان هزینه کلاس بندی اشتباه	۰,۳۱۰٦٥٥۷٤	۰,۳۲۳٤۲٦٦	۰,۲۲۹۱٦۷	۰,۳۲٦۹۲۳
۰,٥٦۲۱۰۱	حساسیت	۰,٥٥۷۳۷۷۰٥	۰,٦۱٥۳۸٤٦	۰,٥۸۳۳۳۳	۰,٤۹۲۳۰۸
۰,٦۲۳۸٦٤	نرخ اختصاصی	۰,٦	۰,٥٤٥٤٥٤٥	۰,۷٥	۰,٦
۰,٥٦۷٤٥٦	دقت	۰,٥٦٥۷۸۹٤۷	۰,٦۰٥۲٦۳۲	۰,٥۹۲۱۰٥	۰,٥۰٦٦٦۷
۰,۹۰۱۱۳	صحت	۰,۸٥	۰,۸۸۸۸۸۹	۰,۹۷٦۷٤٤	۰,۸۸۸۸۸۹
۰,٥٦۲۱۰۱	فراخوانی مجدد	۰,٥٥۷۳۷۷۰٥	۰,٦۱٥۳۸٤٦	۰,٥۸۳۳۳۳	۰,٤۹۲۳۰۸
۰,٦۹۱۱٦	معیار ترکیبی F	۰,٦۷۳۲٦۷۳۳	۰,۷۲۷۲۷۲۷	۰,۷۳۰٤۳٥	۰,٦۳۳٦٦۳
-۳,۱٥٦۰۸	سازگاری	-۱,۲٤۲٦۲۳	-۱,٦٥۷۳٤۲۷	-٦,۹۱٦٦۷	-۲,۸۰۷٦۹
۰,٥۹۲۹۸۲	AUC	۰,٥۷۸٦۸۸٥۲	۰,٥۸۰٤۱۹٦	۰,٦٦٦٦٦۷	۰,٥٤٦۱٥٤
۰,٥۸۹۱۲۳	معیار توازن	۰,٥۷۸۱٤۹۸٦	۰,٥۷۸۹٦٥۲	۰,٦٥٦٤۰۸	۰,٥٤۲۹۷۱
۲۲,۲۱۱۹۸	زمان اجرا	۲۷,۸٥۲۸۰٦۸	۲۳,٦۳۰٤۳۹	۲٤,۰٥۷۲۷	۱۳,۳۰۷۳۹

بنابراین تا این مرحله بهترین دقت به دست آمده ۰,٦۰٥ است که از روش قبلی نتیجه بهتری است.

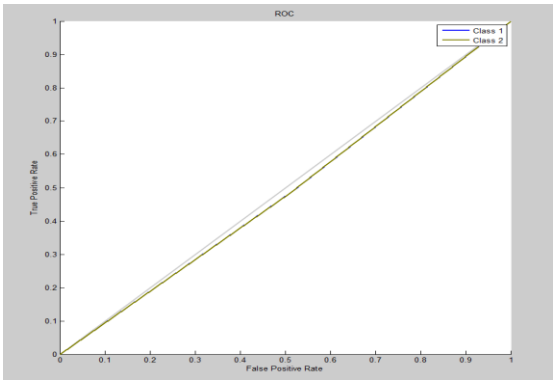
۳.عملکرد سیستم K نزدیکترین همسایه

سیستم K نزدیکترین همسایه در این تحقیق بر مبنای خوشه‌بندی K نزدیکترین همسایه عمل می‌کند.

نتایج بدست آمده به شرح زیر است.



نمودار۴:کانفیوژن برای سیستم K نزدیکترین همسایه



نمودار۵:مشخصه عامل گیرنده برای سیستم K نزدیکترین همسایه

جدول۳: نتایج بدست آمده از ترکیب سیستم K نزدیکترین همسایه

MEAN		k۱	k۲	k۳	k٤
۰,۲۵٤۱ ٦۷	نرخ کلاس بندی اشتباه	۰,۲۲۳٦۸ ٤۲۱	۰,۱۱۸٤۲ ۱۱	۰,٤۰۷۸ ۹۵	۰,۲٦٦٦ ٦۷
۰,۳۱٤۷ ۳۸	میزان هزینه کلاس بندی اشتباه	۰,۳۲۸٤۰ ۰۲۸	۰,۲۱۲۱۸ ٤۹	۰,٤۵۸۳ ۳۳	۰,۲٦۰۰ ۳۳
۰,۱۵۷۳ ٦۹	نرمال شده میزان هزینه کلاس بندی اشتباه	۰,۱٦٤۲۰ ۰۱٤	۰,۱۰٦۰۹ ۲٤	۰,۲۲۹۱ ٦۷	۰,۱۳۰۰ ۱۷
۰,۷۳۳۲ ۹۸	حساسیت	۰,۷٦۷٤٤ ۱۸٦	۰,۹۲۸۵۷ ۱٤	۰,۵۸۳۳ ۳۳	۰,٦۵۳۸ ٤٦
۰,۸۱۸٦ ۱۳	نرخ اختصاصی	۰,۷۸۷۸۷ ۸۷۹	۰,۸۲۳۵۲ ۹٤	۰,۷۵	۰,۹۱۳۰ ٤۳
۰,۷٤۵۸ ۳۳	دقت	۰,۷۷٦۳۱ ۵۷۹	۰,۸۸۱۵۷ ۸۹	۰,۵۹۲۱ ۰۵	۰,۷۳۳۳ ۳۳
۰,۹۰۳۲ ۱٤	صحت	۰,۸۲۵	۰,۸٦٦٦٦ ٦۷	۰,۹۷٦۷ ٤٤	۰,۹٤٤٤ ٤٤
۰,۷۳۳۲ ۹۸	فراخوانی مجدد	۰,۷٦۷٤٤ ۱۸٦	۰,۹۲۸۵۷ ۱٤	۰,۵۸۳۳ ۳۳	۰,٦۵۳۸ ٤٦
۰,۷۹۸۷ ۲٤	معیار ترکیبی F	۰,۷۹۵۱۸ ۰۷۲	۰,۸۹٦۵۵ ۱۷	۰,۷۳۰٤ ۳۵	۰,۷۷۲۷ ۲۷
– ۱,٤۳۵۱	سازگاری	۰,٤٦٤٤۱	۰,۸٤۰۳۳	– ٦,۹۱٦٦	– ۰,۱۲۸۷

۷		۱۵٦	٦۱	۷	٦
۰,۷۷۵۹۵٦	AUC	۰,۷۷۷٦٦۰۳۲	۰,۸۷٦۰۵۰٤	۰,٦٦٦٦٦۷	۰,۷۸۳٤٤۵
۰,۷٦۱۷۱۱	معیار توازن	۰,۷۷۷٤۲۵٦۳	۰,۸٦۵۳۸۲۲	۰,٦۵٦٤۰۸	۰,۷٤۷٦۲۷
۹,۷۷۳٦۹٦	زمان اجرا	٤,٤٤٦۳۷۸۵۳	٤,٤۱۳۵۷۹٦	۲٦,۰۳۷۳۷	٤,۱۹۷٤۵٦

همان طور که از نتایج پیدا است هر ٤ اجرا باعث بهبود سیستم K نزدیکترین همسایه نسبت به حالت قبلی شده‌اند و از بین این سه روش، روش K نزدیکترین همسایه بهترین نتیجه را به همراه داشته است.

نتیجه‌گیری

با توجه به موضوع مورد پژوهش، در ابتدا شناسایی خطا در محاسبات ابری و انواع شاخص‌های ارزیابی شرح داده شد و در ادامه به مرور عملکرد سه شبکه عصبی پرسپترون، براساس عملکرد شعاعی و K نزدیکترین همسایه پرداختیم در نهایت بهترین ترکیب در این بین سیستم K نزدیکترین همسایه با دقت نزدیک ۷٤ درصد است. با توجه به نتایج به دست آمده کاملاً مشهود است که در بین سیستم‌های پیش‌بینی مورد استفاده بهترین عملکرد را K نزدیکترین همسایه دارا است.

منابع

۱. Buyya, R., Vecchiola, C., & Selvi, S. T. (۲۰۱۳). Mastering cloud computing : foundations and applications programming.

۲. H. Brian, T. Brunschwiler, H. Dill, H. Christ, B. Falsafi, M. Fischer, et al., "Cloud computing," Communications of the ACM, vol. ۵۱, pp. ۹-۱۱, ۲۰۰۸.

۳. Magoulès, F., Pan, J., & Teng, F. (۲۰۱۲). Cloud Computing : Data-Intensive Computing and Scheduling. Boca Raton: CRC Press.

٤. A. Soule, K. Salamatian, and N. Taft, "Combining filtering and statistical methods for anomaly detection," in Proceedings of the ۵th ACM SIGCOMM conference on Internet Measurement, pp۸. ۳۱-۳۱, ۲۰۰۵

۵. P. Barford, J. Kline, D. Plonka, and A. Ron, "A signal analysis of network traffic anomalies," in Proceedings of the ۲nd ACM SIGCOMM Workshop on Internet measurment, pp۸. ۷۱-۸۲, ۲۰۰۲.

٦. A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies ",in ACM SIGCOMM Computer Communication Review, pp۸. ۲۱۹-۲۳۰, ۲۰۰٤.

۷. S. Roschke, F .Cheng, and C. Meinel, "Intrusion detection in the

cloud," in Dependable, Autonomic and Secure Computing, ۲۰۰۹. DASC'۰۹. Eighth IEEE International Conference on, pp^. ۷۲۹-۷۳٤, ۲۰۰۹.

۸. C.-C. Lo, C.-C. Huang, and J. Ku, "A cooperative intrusion detection system framework for cloud computing networks," in Parallel Processing Workshops (ICPPW), ۲۰۱۰ ۳۹th International Conference on, pp^. ۲۸۰-۲۸٤, ۲۰۱۰.

۹. J.-H. Lee, M.-W. Park, J.-H. Eom, and T.-M. Chung, "Multi-level Intrusion detection system and log management in cloud computing," in Advanced Communication Technology (ICACT), ۲۰۱۱ ۱۳th International Conference on, pp^. ٥٥۲-٥٥٥, ۲۰۱۱.

۱۰. C. Mazzariello, R. Bifulco, and R. Canonico, "Integrating a network ids into an open source cloud computing environment," in Information Assurance and Security (IAS), ۲۰۱۰ Sixth International Conference on, pp^. ۲٦٥-۲۷۰, ۲۰۱۰.

۱۱. S. Bugiel, S. Nürnberger, A.-R. Sadeghi, and T. Schneider, "Twin clouds: An architecture for secure cloud computing," in Workshop on Cryptography and Security in Clouds (WCSC ۲۰۱۱), ۲۰۱۱.

۱۲. Manvi, S. S., & Shyam, G. K. (۲۰۱٤). Resource management for Infrastructure as a Service (IaaS) in cloud computing: A survey. Journal of Network and Computer Applications, ٤۱, ٤۲٤-٤٤۰.

۱۳. Muja, M., & Lowe, D. G. (۲۰۱٤). Scalable nearest neighbor algorithms for high dimensional data. IEEE Transactions on Pattern Analysis and Machine Intelligence, ۳٦(۱۱), ۲۲۲۷-۲۲٤۰.

۱٤. R. O. Duda, et al., Pattern classification: John Wiley & Sons, ۲۰۱۲.

Abstract

Cloud computing is a new technology for managing and providing services over the Internet. Cloud computing is a large group of connected computers. These computers can be personal computers or network servers. They can also be public or private. The prediction of content error is one of the fundamental challenges in cloud computing. In this research, we try to increase the accuracy of prediction error of content in cloud computing using the nearest neighbor algorithm. The simulation results in MATLAB software show that the method for the data set has a very high performance. The results are compared with the performance of three perceptron neural networks based on radial performance and nearest neighbor. Finally, the best combination between these systems gave the nearest neighbor the fault detection with close accuracy of ۷٤٪.

Keyword: cloud computing, closest neighbor, content error prediction, radial function function