ارائه راهکاری نوین جهت بهبود شبکه های عصبی مصنوعی پرسپترون چند لایه مبتنی بر ترکیب با الگوریتم جهش قورباغه برای تشخیص ایمیل‌های اسپم

احمدحیدریان[۱]، فرهاد سلیمانیان قره‌چپق[۲]

۱- دانشجوی دکتری، دانشکده مهندسی برق وکامپیوتر،دانشگاه آزاد اسلامی واحد میانه، میانه ،ایران

۲- استادیار،دانشکده مهندسی برق وکامپیوتر،دانشگاه آزاد اسلامی واحدارومیه، ارومیه،ایران

چکیده

با پیشرفت تکنولوژی و استفاده چشمگیر افراد و گروه ها از اینترنت، تبلیغات در این عرصه گسترش انفجارگونه ایی یافته است به طوری که انواع روش ها برای پخش این تبلیغات به صورت انبوه برای کاربران اینترنت به صورت ایمیل مواجه شده اند. و این امر باعث ایجاد مشکلاتی برای کاربران اینترنت می شود. یکی از این مشکلات وجود ایمیل های اسپم می باشد که ایمیل اسپم یکی از متداول‌ترین و در عین‌حال منفی‌ترین جنبه‌های دارابودن یک آدرس ایمیل است. با این‌که در حال‌حاضر و باتوجه به تکنولوژی‌های موجود امکان حذف کامل این نوع از نامه‌های الکترونیکی ناخواسته وجود ندارد، ولی می‌توان بااستفاده از برخی روش‌های موجود تعداد آنان را کاهش داد. اسپم نسخه الکترونیکی از "نامه‌های غیرقابل استفاده" است. واژه اسپم به پیام‌های الکترونیکی ناخواسته، اطلاق می‌گردد. این نوع از نامه‌های الکترونیکی ارتباط مستقیمی با ویروس نداشته و حتی ممکن است پیام‌هایی که از منابع معتبر ارسال شده‌اند؛ نیز در زمره این گروه قرار گیرند. در این مقاله برای نیل به این هدف از روش جدیدی با بهبود شبکه عصبی مصنوعی پرسپترون چند لایه با الگوریتم جهش قورباغه استفاده شده‌است. بدین‌صورت که از الگوریتم جهش قورباغه برای یافتن اندیس بهترین ویژگی‌ها و از شبکه عصبی مصنوعی پرسپترون چندلایه برای تشخیص ایمیل‌های اسپم براساس بهترین ویژگی‌های یافته استفاده شده‌است. نتایج شبیه‌سازی نشان می‌دهد که الگوریتم جهش قورباغه توانسته است بهبود قابل توجهی در شبکه عصبی مصنوعی پرسپترون چندلایه را در مقایسه با شبکه عصبی مصنوعی تابع پایه شعاعی داشته باشد.

کلمات کلیدی: شبکه‌های عصبی مصنوعی، پرسپترون چندلایه، الگوریتم جهش قورباغه، تشخیص ایمیل‌های اسپم.

تاریخچه مقاله:

تاریخ ارسال: ۱۴۰۱/۰۹/۱۰
تاریخ اصلاحات: ۱۴۰۱/۱۱/۲۵
تاریخ پذیرش: ۱۴۰۱/۱۲/۲۰
تاریخ انتشار: ۱۴۰۱/۱۲/۲۹

Keywords:

Multilayer perceptron, Shuffled frog leaping algorithm Detection of spam emails

*** ایمیل نویسنده مسئول:**
farhad@iaurmia.ac.ir

Presenting a novel method to improve multi-layered perceptron artificial neural networks based on combination with frog leaping algorithm to detect spam emails

Ahmad Heydariyan[1] ,Farhad Soleymanian QareChopoq[2]
1.Ph.D. Student, Department of Electrical and Computer Engineering, Mianeh Branch Islamic Azad University, Mianeh, Iran
2.Assistant, Department of Electrical and Computer Engineering, Urmia Branch Islamic Azad University,Urmia, Iran

Abstract

With the advancement of technology and the significant use of the Internet by individuals and groups, advertisements in this field have expanded explosively, so that all kinds of methods for mass distribution of these advertisements to Internet users in the form of e-mail have been encountered. And this causes problems for Internet users. One of these problems is the presence of spam emails which Spam email is one of the most common negative features suffering the owner of an email address. Although existing technologies cannot eliminate unwanted spams, some existing methods can reduce their number. By definition, spam is an electronic version of "useless mails." Spam refers to unwanted and unsolicited e-mails. Such e-mails are not necessarily directly related to the virus, meaning that messages sent from valid sources may also be included in this class. This study proposes a new method based on improving the multilayer perceptron artificial neural network using the shuffled frog leaping algorithm to realize this objective. The shuffled frog leaping algorithm is used to find the best features, and the artificial multilayer perceptron neural network is used to detect spam emails based on the best features. The simulation results prove that the shuffled frog leaping algorithm has significantly improved the multilayer perceptron artificial neural network compared to the artificial neural network based on the radial base function.

۱ – مقدمه

با پیشرفت تکنولوژی و استفاده چشم‌گیر افراد و گروه‌ها از اینترنت، تبلیغات در این عرصه گسترش انفجارگونه‌ای یافته است. به‌طوری که انواع روش‌ها برای پخش این تبلیغات به‌طور انبوه برای افراد و گروه‌ها ایجاد شده و کاربران اینترنت با سیل عظیمی از این‌گونه تبلیغات به‌ویژه به‌صورت ایمیل مواجه شده‌اند. از آن‌جایی که این سیل عظیم از ایمیل‌های تبلیغاتی باعث ایجاد مشکلاتی برای کاربران اینترنت می‌شود، محققان در صدد رفع این مشکل برآمده‌اند و تاکنون روش‌ها و الگوریتم‌های زیادی نیز ارائه نموده‌اند[1].

ایمیل‌های اسپم علاوه‌بر تبلیغات برای مقاصدی چون سرقت اطلاعات، جاسوسی، سرگرمی، مزاحمت و ... برای کاربران اینترنت ارسال می‌شود[2]. ارسال ایمیل اسپم گاهی به‌شکل گروهی و توسط نرم‌افزارهای ارسال‌کننده به‌صورت انبوه به آدرس ایمیل‌های فراوانی ارسال می‌شود که از دو جنبه هدر رفتن پهنای باند اینترنت و هدر رفتن منابع پردازشی سرویس‌دهنده‌ها و دیتاسنترها مخرب می‌باشد[4,3]. بنابراین شناسایی و مسدود نمودن اسپم‌ها یکی از موارد مهم در بحث امنیت فضای مجازی است که می‌تواند تا حدود زیادی اثر این پدیده نامطلوب اینترنت و چالش امنیتی سرویس پست الکترونیک را کاهش دهد[5]. یکی از روش‌های مهم کشف و شناسایی اسپم‌ها استفاده از تکنیک‌های داده‌کاوی و یادگیری ماشین در شناسایی الگوهای پنهان آن می‌باشد[6].شناسایی الگوهای پنهان اسپم توسط تکنیک‌های داده‌کاوی و یادگیری ماشین باعث می‌شود که ایمیل‌های دریافتی با دقت مناسب به دو دسته نرمال و غیرنرمال طبقه‌بندی شوند. کاهش خطای طبقه‌بندی تشخیص اسپم‌ها توسط تکنیک‌های داده‌کاوی نظیر شبکه عصبی مصنوعی به‌عنوان یک مسئله بهینه‌سازی سخت و دشوار، یکی از چالش‌های مهم این روش-ها در تشخیص اسپم‌ها از ایمیل‌های عادی محسوب می‌شود[7].یافتن کمینه خطای طبقه‌بندی اسپم‌ها از ایمیل‌های عادی یک مسئله بهینه‌سازی است که توسط روش‌هایی نظیر الگوریتم‌های فراابتکاری قابل حل است. الگوریتم‌های فراابتکاری دارای زیرمجموعه‌های مختلفی نظیر الگوریتم‌های تکاملی و الگوریتم‌های هوش دسته جمعی است که نقش مهمی در حل مسائل بهینه‌سازی دارند[8].

شبکه‌های عصبی مصنوعی ابزاری تحلیلی و آموزش‌پذیر هستند که تلاش می‌کنند تا الگوهای پردازش اطلاعات در مغز بشر را تقلید کنند. این شبکه‌ها، سیستم‌هایی دینامیکی متشکل از واحدهای پردازش موازی یا همان نرون‌ها هستند که نوعی میل باطنی برای حفظ دانش تجربی و در دسترس قراردادن آن برای استفاده دارند[9].خاصیت یادگیری شبکه‌های عصبی از اهمیت ویژه‌ای برخوردار است. این شبکه‌ها به‌عنوان سیستم‌های یادگیری دارای این توانایی هستند که از گذشته، تجربه و محیط بیاموزند و رفتار خود را در حین هر یادگیری بهبود بخشند. بهبود یادگیری شبکه در طول زمان براساس معیاری سنجیده می‌شود که این معیار بهبود، هدف سیستم یادگیری را مدل می‌کند [۱۰]. یکی از دلایلی که شبکه‌های عصبی برای تشخیص ایمیل اسپم مناسب به‌نظر می‌رسد، توانایی آن‌ها در تعمیم‌پذیری، خودساماندهی و پردازش‌های موازی بااستفاده از سیستم‌های چندپردازنده می‌باشد [۱۱].

الگوریتم جهش قورباغه یکی از الگوریتم های الهام گرفته از طبیعت است. در این الگوریتم گروهی از قورباغه‌ها (مجموعه‌ای از جواب‌ها) به چندین زیرمجموعه تقسیم می‌شوند، که هر قورباغه فرهنگ مختص به‌خود را دارد و می‌تواند از فرهنگ‌ها یا ایده‌های قورباغه‌های دیگر در طول روند تکامل استفاده کند.

در بخش بعدی این مقاله، به مروری بر کارهای انجام‌شده در طی چند سال اخیر پرداخته می‌شود. در بخش ۳ راجع به تشخیص ایمیل اسپم و بیان روش کار، جمع‌آوری و آنالیز اطلاعات موردنیاز و بررسی نرم‌افزار و کدهای شبیه‌سازی توضیح داده شده است. در بخش ۴ نتایج تجربی پژوهش، شامل ارزیابی عملکرد و مقایسه دقت تشخیص ایمیل اسپم توسط روش‌های ذکر شده آورده شده و بخش ۵ شامل نتیجه‌گیری، جمع‌بندی و ارائه پیشنهاداتی برای کارهای آتی است.

۲-پیشینه تحقیق

به‌منظور آشنایی با روش‌های متداول تشخیص ایمیل اسپم، در این بخش چند روش استفاده شده برای تشخیص اسپم به اختصار شرح داده شده است.

Wang و همکاران از ماشین بردارپشتیبان جهت طبقه‌بندی و تشخیص اسپم و از الگوریتم ژنتیک جهت انتخاب ویژگی استفاده کرده‌اند. انتخاب ویژگی بااستفاده از حذف ویژگی‌های نامربوط و مقیاس‌بندی صحیح ویژگی‌های موجود صورت گرفته است و برای ارزیابی نتایج از مجموعه‌داده UCI استفاده کرده‌اند [۱۲].

Dinh و همکاران برای تشخیص مراکزی که اسپم‌ها را ارسال می-نمایند یک روش مبتنی‌بر درخت تکرار نمونه به‌عنوان یک روش داده‌کاوی ارائه نمودند[13]. در چارچوب موردنظر یک ایمیل از ورودی خوانده‌شده سپس ویژگی‌های مهم آن توسط فاز استخراج ویژگی تهیه و در پایگاه داده تعبیه‌شده این روش که از نوع SQL می‌باشد، ذخیره‌سازی می‌شود. ویژگی‌های ذخیره‌شده در پایگاه داده می‌تواند توسط الگوریتم درخت تکرار الگو مورد واکاوی قرارگرفته و میزان شباهت آن با الگوهای کشف‌شده اسپم‌ها مشخص شود. نتایج تحقیق آن‌ها نشان می‌دهد که روش پیشنهادی به‌خوبی دامنه و آدرس مراکز ارسال‌کننده اسپم و موضوعات موردعلاقه ارسال‌کنندگان اسپم را شناسایی‌نموده تا از کلمات کلیدی برای شناسایی دقیق‌تر اسپم‌ها استفاده شود و چالش اصلی این روش زمان و حافظه زیادی است که صرف ساختن درخت الگو می‌شود.

در [14] برای تشخیص هرزنامه‌های شبکه‌های اجتماعی یک روش مبتنی‌بر یادگیری ماشین و ماشین بردار پشتیبان ارائه نمودند. در روش پیشنهادی آن‌ها اسپم‌های شبکه اجتماعی Weibo موردبررسی قرار گرفته و از مجموعه‌داده‌ای متشکل از ۱۰۰ کاربر عادی و ۵۰ کاربر ارسال‌کننده اسپم جهت آموزش روش پیشنهادی خود استفاده نمودند. آن‌ها برای تشخیص نوع کاربر شبکه اجتماعی از دو تکنیک جستجوی دنبال‌کنندگی و پیام‌های ارسالی آن‌ها استفاده نمودند. نتایج پیاده‌سازی آن‌ها برروی مجموعه‌داده مرتبط با شبکه اجتماعی Weibo نشان می‌دهد که روش آن‌ها کارآیی بیشتری نسبت به روش‌هایی نظیر درخت تصمیم‌گیری و شبکه بیزین در تشخیص اسپم‌های شبکه اجتماعی Weibo دارد.

Alsmadi و Alhami یک روش جدید برای طبقه‌بندی انواع ایمیل در ۵ دسته شخصی، کاری، آموزشی، دوستی و غیره به کمک تکنیک خوشه‌بندی Ngram در ابزار وکا ارائه نمودند[15]. نتایج بررسی آن‌ها در ابزار وکا نشان می‌دهد که این روش جداسازی ایمیل‌ها برحسب محتوا، دارای دقت و صحتی به‌ترتیب برابر ۸۳٪ و ۸۸٪ می‌باشند.

محققان بااستفاده از یک سیستم فازی طبقه‌بندی و ماتریس عضویت هر ایمیل را در خوشه‌های مرتبط با موضوع آن قرار دادند[16]. در روش پیشنهادی آن‌ها ویژگی‌های ایمیل در قالب یک بردار ویژگی تحویل سیستم فازی پیشروشده و خروجی سیستم فازی در اختیار سه الگوریتم طبقه‌بندی سیستم نرو- فازی، شبکه عصبی بازگشتی شعاعی و سیستم فازی‌تطبیق شده قرار داده می‌شود تا هر ایمیل برحسب ویژگی‌هایش به کلاس مناسب تقسیم‌بندی شود. نتایج بررسی آن‌ها نشان می‌دهد روش پیشنهادی آن‌ها دقت مناسبی در شناسایی اسپم‌ها دارد و از چالش‌های مهم این روش نحوه انتخاب تعداد خوشه‌های به‌کار رفته می‌باشد.

Renuka و همکاران از یک طبقه‌بندی گروهی برای طبقه‌بندی ایمیل اسپم در محیط Hadoop استفاده کردند، که مجموعه‌ای از درخت تصمیم‌گیری ضعیف و رأی‌گیری اکثریت وزن با استفاده از درخت تصمیم‌گیری و بیزین ساده است[17].برای این منظور از مجموعه‌داده SPAMBASE استفاده کرده‌اند. نتایج پیشنهادی آن‌ها نشان می‌دهد که دقت طبقه‌بندی با استفاده از سیستم طبقه‌بندی ارتقای گرادیان (۹۴/۱٪) بیشتر از بیزین ساده (۸۰/۷۷٪) است.

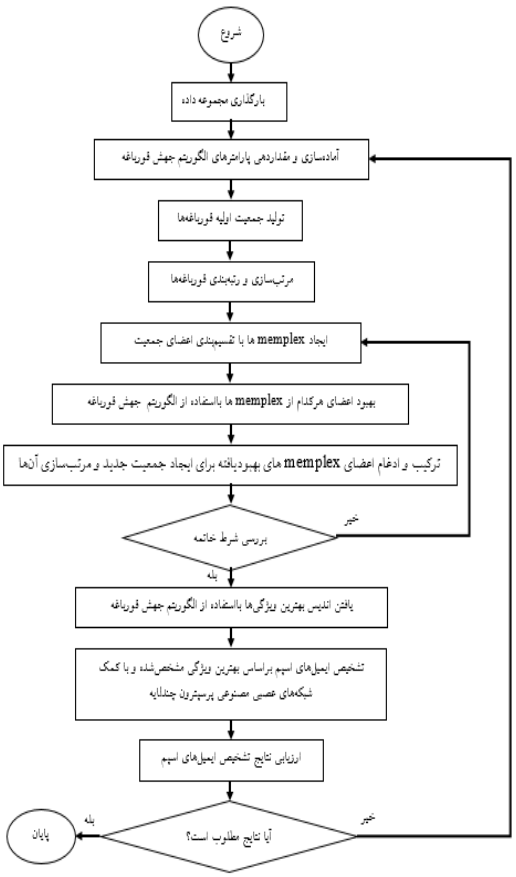
Rusland و همکاران با استفاده از آنالیز الگوریتم بیزین ساده به فیلترکردن ایمیل‌های اسپم پرداخته‌اند[18]. آن‌ها معتقدند که الگوریتم بیزین ساده نقش مهمی در کیفیت عملکرد طبقه‌بندی دارد. برای این منظور از دو مجموعه‌داده اسپم موجود در سایت UCI استفاده نموده‌اند. نتایج نشان می‌دهد که دقت طبقه‌بندی مجموعه‌داده اسپم برابر ۹۱/۱۳٪ و برای مجموعه‌داده SPAMBASE برابر ۸۲/۵۴٪ است.

ما در این مقاله بااستفاده از بهبود شبکه‌های عصبی مصنوعی پرسپترون چندلایه با الگوریتم جهش قورباغه سعی در افزایش دقت تشخیص ایمیل‌های اسپم داریم.

۳- راهکار پیشنهادی

در این بخش، به ارائه روشی جهت بهبود شبکه‌های عصبی مصنوعی پرسپترون چندلایه با الگوریتم جهش قورباغه برای تشخیص بهتر ایمیل‌های اسپم بااستفاده از مجموعه‌داده‌های سایت UCI می-پردازیم. برای این‌منظور از نرم‌افزار "MATLAB" به دلیل انعطاف پذیری، توابع آماده بسیار و کارآیی مناسب برای پیاده‌سازی الگوریتم-ها استفاده شده است. هدف یافتن بهترین ویژگی‌ها با کمک الگوریتم جهش قورباغه و تشخیص ایمیل‌های اسپم با کمک شبکه عصبی‌مصنوعی است. شبکه عصبی پرسپترون برمبنای یک واحد محاسباتی به‌نام پرسپترون ساخته می‌شود. یک پرسپترون برداری از ورودی‌ها با مقادیر حقیقی را گرفته و یک ترکیب خطی از ورودی ها را محاسبه می‌کند. اگر حاصل از یک مقدار آستانه بیشتر بود خروجی پرسپترون برابر با ۱ و درغیر این‌صورت معادل ۱- خواهد بود. شبکه‌های عصبی پرسپترون، به‌ویژه پرسپترون چندلایه در زمره کاربردی‌ترین شبکه‌های عصبی می‌باشند، این شبکه‌ها قادرند با انتخاب مناسب تعداد لایه‌ها و سلول‌های عصبی، که اغلب هم زیاد نیستند، یک نگاشت غیرخطی را با دقت دلخواه انجام دهند [19]. برای نیل به این هدف در شکل (۱) فلوچارت روش پیشنهادی به‌طور خلاصه نشان داده شده‌است.

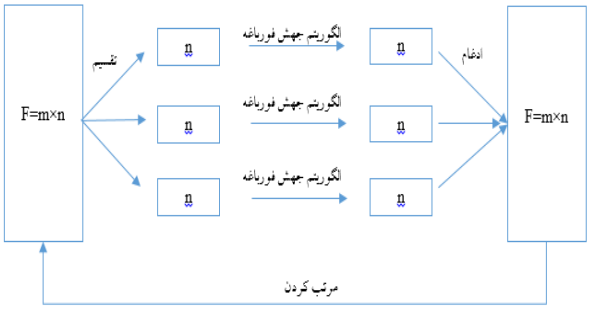
مجموعه‌داده ایمیل اسپم UCI شامل ۵۷ ویژگی است، که در تشخیص ایمیل‌های اسپم از آن‌ها می‌توان استفاده کرد. برای این‌منظور بهترین تعداد ویژگی را با بررسی نتایج الگوریتم جهش قورباغه به‌دست خواهیم آورد. پس تعداد بهترین ویژگی‌ها از ۲ تا ۵۷ ویژگی درنظر گرفته شده‌است. با اجرای این الگوریتم اندیس بهترین ویژگی‌ها به‌دست می‌آید.



(شکل -۱): فلوچارت روش پیشنهادی

(Figure-1): Flowchart of proposal method

الگوریتم جهش قورباغه که با نام الگوریتم جهش قورباغه مخلوط‌شده نیز شناخته می‌شود، در دسته الگوریتم‌های رفتاری قرار می‌گیرد. این الگوریتم به الگوریتم تکامل مجتمع‌های مخلوط‌شده وابسته‌شده که ترکیبی از الگوریتم جستجوی تصادفی کنترل‌شده و الگوریتم ژنتیک است. عملکرد الگوریتم جهش قورباغه به‌طور خلاصه در شکل (۲) نشان داده شده‌است.



(شکل-2): عملکرد کلی الگوریتم جهش قورباغه

(Figure-2): The overall performance of the frog leaping algorithm

در[20] الگوریتم جهش قورباغه را نوعی الگوریتم Memetic معرفی کرده یعنی الگوریتمی که از رفتار و فرهنگ ایده گرفته است. در واقع الگوریتم جهش قورباغه از رفتار و حرکت و جهش قورباغه‌ها ایده گرفته است.

اندازه جمعیت قورباغه‌ها برابر F=m×n است که در آن m برابر تعداد MempIex ها و n برابر اندازه MempIex است. در مرحله اول جمعیت قورباغه‌ها به m تا MempIex با اندازه n تقسیم می‌شوند. تقسیم‌بندی قورباغه‌ها بدین‌صورت است که قورباغه اول به MempIex اول، قورباغه دوم به MempIex دوم و قورباغه mام به MempIex، mام و قورباغه m+۱ به MempIex اول تعلق دارند. این روند به‌صورت مشابه تا قورباغه آخر تکرار می‌شود. هر MempIex شامل n قورباغه خواهد بود. در مرحله دوم، هر MempIex توسط الگوریتم جهش قورباغه بهبود می‌یابند، سپس MempIexهای بهبودیافته با هم ادغام‌شده و جمعیت جدیدی با اندازه F=m×n حاصل می‌شود. جمعیت جدید از بهترین تا بدترین مرتب‌شده و دوباره این روال ادامه می‌یابد.

پس به‌طور خلاصه می‌توان مراحل الگوریتم جهش قورباغه را بدین شرح بیان نمود:

۱) آماده‌سازی و مقداردهی پارامترها؛

$$m = \text{تعداد mempIex} \geq ۱$$

$$n = \text{اندازه mempIex} \leq d_H$$

$$F = \text{اندازه جمعیت برابر } m \times n$$

تعداد mempIex ها برابر ۵، اندازه mempIex ها برابر ۱۰ درنظر گرفته شده‌است. پس اندازه جمعیت برابر ۵۰ خواهد بود.

۲) تولید جمعیت اولیه قورباغه‌ها: جمعیت اولیه قورباغه‌ها دارای سه ویژگی Position برای موقعیت، Cost برای محاسبه تابع هدف و Out برای اندیس ویژگی‌های مجموعه‌داده که از هر نسل از الگوریتم جهش قورباغه به‌دست می‌آید، است. برای محاسبه موقعیت قورباغه‌ها از روش تصادفی استفاده شده‌است؛ بدین‌صورت که موقعیت هر قورباغه در بازه بین حد بالا و پایین مقادیر مجموعه‌داده (در اینجا ۱- تا ۱) تعیین می‌شود. مقدار Cost از تابع هدف به‌دست می‌آید. مقدار Cost برابر میانگین مربع خطا خواهد بود.

۳) مرتب‌سازی و رتبه‌بندی قورباغه‌ها: مرتب‌سازی و رتبه‌بندی قورباغه‌ها براساس میانگین مربع خطای به‌صورت صعودی انجام می‌گیرد.

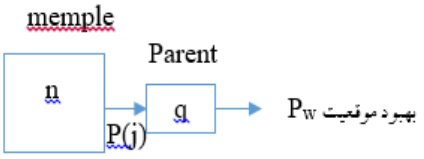
۴) ایجاد mempIex ها با تقسیم‌بندی اعضای جمعیت: جمعیت قورباغه‌ها به m تا MempIex با اندازه n تقسیم می‌شوند.

۵) بهبود اعضای هرکدام از mempIex ها بااستفاده از الگوریتم جهش قورباغه: نحوه کار الگوریتم جهش قورباغه در ادامه به طور کامل بحث خواهد شد.

۶) ترکیب و ادغام اعضای mempIex های بهبودیافته برای ایجاد جمعیت جدید و مرتب‌سازی آن‌ها؛

۷) بررسی شرایط خاتمه و تکرار مرحله (۴): در الگوریتم جهش قورباغه نیاز به متغیری است که شرط خاتمه را مشخص نماید. پارامتر MaxIteration حداکثر تکرار حلقه اصلی الگوریتم را مشخص می‌کند. در واقع این روش شرط خاتمه، پس از این تکرار مشخص نسل، اجرای الگوریتم متوقف می‌شود. مقدار پارامتر MaxIteration برابر ۱۰۰ درنظر گرفته شده‌است.

۸) نحوه عملکرد الگوریتم جهش قورباغه نیز به‌طور خلاصه در شکل (۳) نشان داده شده‌است.



(شکل -۳): نحوه عملکرد الگوریتم جهش قورباغه

(Figure-3): How the leapfrog algorithm works

بخشی از MempIex که هرکدام دارای احتمال انتخاب p(j) هستند، به اندازه q، انتخاب و از بهترین تا بدترین مرتب می‌شوند. هدف این الگوریتم بهبود موقعیت بدترین والد است. برای درک بهتر الگوریتم جهش قورباغه در ادامه مراحل الگوریتم بیان شده‌است:

۱) آماده‌سازی و مقداردهی پارامترها؛

q= تعداد والدها برابر ۳۰ درصد تعداد mempIex ها؛

α= تعداد فرزندان برابر ۳۰؛

β= ماکزیمم تکرار حلقه اصلی الگوریتم برابر ۵؛

σ= گام تغییر موقعیت قورباغه ها برابر ۲ درنظر گرفته شده‌است.

۲) تولید یک زیرجمعیت با نمونه‌برداری تصادفی وزن‌دار و بدون جایگذاری از جمعیت [احتمالات به‌صورت تصادفی]: برای تولید این زیرجمعیت از رابطه (۱) استفاده شده که nPop اندازه جمعیت است.

$$Pop = 2\times(nPop+1-(1:nPop))/(nPop\times(nPop+1)) \quad (۱)$$

برای انتخاب بدون جایگذاری نیز از تابع RandsampIe موجود در محیط نرم‌افزار متلب استفاده شده است.

۳) مرتب‌سازی اعضای زیرجمعیت به‌نحوی که اولین عضو، بهترین باشد.

(۲)

۴) بهبود موقعیت بدترین والد با P_B

$$\begin{cases} U_q^{new} = P_W \times S \\ S = rand \times (P_B - P_W) \end{cases} \quad (۲)$$

- اگر u_q^{new} در فضای جستجو باشد، مقدار تابع هدف برای آن محاسبه می‌شود. درغیر این‌صورت اجرای مرحله (۵).
- اگر u_q^{new} بهتر از $P_w=u_q$ باشد، جایگزین آن شده و اجرای مرحله (۷). درغیر این‌صورت اجرای مرحله (۵).

۵) اگر مرحله (۴) موفقیت‌آمیز نباشد، بهبود موقعیت بدترین والد با P_x انجام می‌شود. که در آن P_x بهترین پاسخ کل جمعیت است.

$$\begin{cases} U_q^{new} = P_W \times S \\ S = rand \times (P_X - P_W) \end{cases} \quad (۳)$$

- اگر u_q^{new} در بازه جستجو باشد، ارزیابی می‌شود، درغیر این‌صورت اجرای مرحله (۶).
- اگر u_q^{new} بهتر از $P_w=u_q$ باشد، جایگزین آن شده و اجرای مرحله (۷).

۶) مرحله سانسور: اگر بهبود P_w (به‌عنوان یک ایده یا Meme) موفقیت‌آمیز نباشد، از انتشار آن پیشگیری می‌شود. r یک پاسخ تصادفی است.

$$u_q = r \quad (۴)$$

۷) اعضای MempIax با اعضای تغییریافته جایگزین می‌شود.

۸) مراحل (۱) تا (۷) به تعداد مشخصی تکرار می‌شوند.

برای بررسی جواب‌های mempIex ها از تابع هدف استفاده می‌شود. که شبکه‌های عصبی‌مصنوعی پس‌انتشار به‌عنوان تابع هدف درنظر گرفته شده‌اند.

با بهبود شبکه‌های عصبی مصنوعی پرسپترون چندلایه با الگوریتم جهش قورباغه بااستفاده از روش ارائه شده، بهترین ویژگی‌های مجموعه‌داده انتخاب و سپس با حداکثر دقت ایمیل‌های اسپم تشخیص داده می‌شوند.

۴-نتایج شبیه سازی و ارزیابی نتایج

پس از اجرای الگوریتم جهش قورباغه برای ۵۰ جمعیت و۱۰۰ تکرار، اندیس بهترین ویژگی‌های مجموعه داده برابر ۸-۱۶-۱۸-۴۳-۳-۱۳-۳۴-۹-۲۱-۳۵-۳۶-۴۷-۴۱-۴۹-۲۲-۲۵-۲۹-۵۱-۱۱-۵۲-۲۴-۱۰-۱۲-۴۵-۵۷-۱-۳۳-۶-۵۳-۳۹-۴۲-۲-۳۱-۳۰-۲۸-۳۲-۱۴-۴۸-۴۴-۵۰-۴-۴۶-۷-۲۳-۵۴-۵-۳۷-۳۸-۱۷-۲۰-۲۷-۵۶-۵۵ بوده است.

(جدول-۱): نتایج انتخاب بهترین ویژگی ها برای ۱۰۰تکرار

(Table-1): The results of selecting the best features for 100 repetitions

تعداد ویژگی‌ها	اندیس ویژگی‌ها	میانگین مربع خطا
۳	۲۳-۴۲-۳۳	۰/۲۱۱
۴	۲۷-۳۶-۵-۴۶	۰/۱۶۷
۵	۳۸-۱۶-۳۱-۸-۳۴	۰/۱۵۸

۶	۴۰-۲۵-۲-۳۰-۱۰-۳۵	۰/۱۵۷
۷	۲۲-۲۳-۵۵-۵-۱۴-۲۵-۴۶	۰/۱۰۶
۸	۵۶-۱۵-۳۱-۱۲-۳۲-۲۷-۴-۱۱	۰/۱۴۲
۹	۱۷-۲۳-۱-۲۸-۵۱-۱۳-۵۳-۲۴-۴۳	۰/۱۲۱
۱۰	۱۵-۱۳-۲۸-۱-۲۱-۴۹-۱۸-۴۲-۱۹-۴۶	۰/۱۲۳
۱۱	-۱۷-۱۲-۴۰-۲۸-۷-۴-۲۱-۵۱-۲۳-۲۴ ۲۵	۰/۱۰۰
۱۲	-۲۲-۵۳-۱۱-۳۴-۵۵-۴۶-۴۵-۱-۳۷-۱۷ ۱۶-۳	۰/۱۱۶
۱۳	-۱۵-۶-۴۳-۵۲-۲۸-۱۸-۱۱-۵-۵۵-۳۷ ۹-۴۲-۵۳	۰/۰۸۲
۱۴	-۳۲-۴۴-۵۶-۲۱-۴۹-۲۲-۶-۵۰-۲۵-۴۷ ۵۴-۱۳-۱۰-۳۰	۰/۱۲۱
۱۵	-۴۸-۵-۵۳-۱۸-۵۱-۵۵-۱۹-۱۷-۳۸-۲۸ ۲۵-۴۳-۷-۲۲-۳۷	۰/۰۸۱
۱۶	-۸-۴۲-۳۲-۹-۳۰-۴-۳۳-۱۵-۵۱-۵۶-۱۰ ۱۸-۳۴-۱۱-۱۳-۲۶	۰/۱۴۵
۱۷	-۴-۴۶-۱۹-۴۰-۳۸-۵۷-۴۹-۱۰-۲۹-۳۲ ۵۲-۲۳-۲۱-۴۷-۳۵-۱۴-۴۲	۰/۱۰۵
۱۸	-۲۷-۳۴-۲۲-۱۷-۴۴-۶-۳۵-۵۴-۲۵-۴ ۲۸-۲۰-۴۱-۳۱-۱۱-۱۸-۲۱-۵۰	۰/۱۰۶
۱۹	-۴۰-۴۲-۱۳-۴۹-۱-۵۶-۴۵-۸-۵-۲۸-۶ ۱۷-۳۲-۷-۹-۳-۱۴-۵۱-۳۳	۰/۱۰۶
۲۰	-۴۸-۳۲-۲-۳۰-۱۳-۱۲-۲۹-۳۵-۴۹-۵۰ ۵-۳-۳۱-۴۰-۲۴-۱-۱۸-۴-۵۲-۲۱	۰/۱۰۳
۲۱	-۸-۹-۲۶-۱۸-۲۵-۳۸-۲۷-۴۳-۲۸-۱۰ -۵۴-۲۹-۶-۴۷-۴۱-۱۲-۵۶-۲۴-۳۵-۴۰ ۳۰	۰/۱۰۲
۲۲	-۲۲-۴۵-۱۹-۷-۲-۲۱-۱۱-۴۲-۲۷-۳ -۳۸-۳۱-۵۱-۱۸-۴۹-۳۷-۹-۲۸-۶-۱۷ ۱۴-۵۷	۰/۰۹۱
۲۳	-۲۱-۵۷-۱۵-۴۴-۴۳-۷-۲۷-۱۹-۴۰-۲۶ -۲۰-۳۰-۳۸-۳۷-۵۶-۲۳-۴-۱۳-۱۸-۵۴ ۱۲-۴۲-۵۲	۰/۷۴
۲۴	-۱۹-۲۴-۳۲-۴۴-۴۸-۳۱-۱۴-۳۰-۴۵ -۲۹-۳۹-۲-۵۳-۴۱-۱۸-۳-۹-۳۸-۲۳-۲۶ ۷-۴۷-۵۲-۱۳	۰/۰۸۵
۲۵	-۴۰-۱۵-۲-۳۹-۲۴-۵۵-۲۸-۲۱-۳۸-۴۲ -۱۷-۸-۴۹-۳۱-۳۴-۱۰-۴۷-۱۶-۵-۹-۴۳ ۳۲-۵۷-۴۸-۱۸	۰/۰۹۶
۲۶	-۳۸-۳۷-۵۴-۳۳-۴۱-۴۵-۲۳-۲۶-۱۶	۰/۰۷۵
	-۵۲-۴۷-۳۹-۲۵-۴۸-۱-۹-۲۱-۸-۴۴ ۱۳-۴۰-۴-۴۳-۷-۳۱-۱۸	
۲۷	-۴۸-۱۱-۲۵-۷-۱۵-۳۱-۳۶-۱۴-۵۵-۲۸ -۵۳-۴۳-۵۷-۴۲-۲۴-۲۹-۹-۱۷-۵۰-۲۱ ۱-۱۶-۲۲-۱۰-۱۸-۲-۴۵	۰/۰۶۷
۲۸	-۲۶-۲۴-۶-۱۱-۲۲-۳۶-۲۰-۲۵-۱۹-۴-۷ -۲۷-۴۱-۱۴-۳۱-۱-۱۶-۵۱-۱۸-۳۲-۵۰ ۱۳-۱۲-۵۲-۴۰-۳۵-۳۷-۹	۰/۰۸۲
۲۹	-۱۶-۱۷-۱۵-۵-۵۴-۲۱-۵۰-۵۳-۲-۳۹ -۳۱-۱۰-۵۲-۱۹-۲۳-۴۴-۲۹-۱۳-۳۸ ۲۵-۴۶-۴۱-۵۶-۱۸-۲۴-۳۳-۴۵-۲۲-۴۹	۰/۰۷۳
۳۰	-۲۹-۲۲-۴۱-۱۳-۳۱-۴۶-۱۸-۲-۲۶-۲۴ -۲۰-۳۳-۵۴-۱۰-۳۲-۵۵-۳۸-۳۷-۵۲-۸ ۶-۳۰-۹-۱۱-۴۸-۴۲-۳۶-۵۶-۳۹	۰/۰۹۷
۳۱	-۵۷-۳۰-۵۰-۳۳-۴۴-۲۶-۲۲-۴۷-۴۹ -۳-۳۱-۸-۳۵-۲۵-۳۹-۱۹-۱۸-۱۵-۵۵ -۳۲-۳۴-۳۷-۴۳-۴۵-۵-۴۱-۴۰-۱۶-۱ ۷-۴۸	۰/۰۸۶
۳۲	-۱۴-۲۴-۲۸-۸-۱۲-۳۹-۱-۲-۴۰-۵۳ -۵۷-۲۳-۵۵-۳۲-۱۱-۵۲-۱۳-۴-۱۷-۴۷ -۳۵-۳۶-۱۹-۴۴-۴۲-۴۱-۱۰-۳۱-۵۱ ۶-۲۷-۴۶	۰/۰۸۵
۳۳	-۳۲-۷-۳۵-۴۹-۴۴-۲۶-۴۰-۸-۴۶-۱۱ -۲۵-۱۸-۲۹-۵۶-۱۶-۵۳-۲۰-۹-۳۶-۴۲ -۱۴-۱۳-۱۷-۴۷-۳۴-۲۷-۵۵-۲۳-۴۵-۱ ۳-۳۰-۳۷	۰/۰۶۵
۳۴	-۴۹-۱۳-۴۴-۱۴-۱۸-۲۴-۴۷-۱۱-۵۳-۵ -۲۷-۳۰-۳۳-۶-۴۱-۳۶-۲۹-۱۲-۱۵-۲۸ -۵۲-۵۵-۴۶-۲-۱۷-۳۹-۱۰-۴۸-۱۹-۵۱ ۴۵-۴-۲۵-۳۱	۰/۰۶۸
۳۵	-۳۰-۴۸-۵۲-۱۹-۵۶-۴۴-۱۴-۵۱-۲۹ -۲۴-۴-۵-۱۷-۵۰-۲۶-۳۲-۱۸-۴۵-۲۱ -۵۴-۵۵-۳۶-۱۰-۵۳-۳۵-۱۶-۲۳-۳۱ ۹-۱۱-۴۰-۴۳-۳-۲۵-۳۷	۰/۰۷۰
۳۶	-۴۹-۲۳-۵-۲۵-۵۳-۵۰-۶-۱۷-۴۸-۸-۳۱ -۱۹-۵۵-۴۶-۴۲-۲۲-۱۳-۹-۱۸-۳۸-۵۷ -۳۴-۳۳-۲۶-۲۰-۴-۵۱-۱۶-۵۲-۲۱-۱ ۵۶-۳۹-۳۰-۴۱-۱۲	۰/۰۶۰
۳۷	-۴۰-۵۱-۵۶-۱۶-۱۷-۲-۲۲-۴۲-۴۸-۱۳ -۴۶-۱۱-۱۸-۱۴-۳۸-۹-۳۳-۵۳-۳۲-۴۷ -۳۴-۵۲-۳۰-۳۱-۲۱-۳۶-۵۰-۳۹-۱۰	۰/۰۷۶

	۵-۴-۲۰-۵۵-۴۴-۲۸-۲۳-۴۳	
۳۸	-۳۸-۷-۴۲-۴۴-۱۲-۲۹-۳۳-۳۴-۲۳-۲۰ -۱۹-۳۶-۵۱-۵۴-۴۸-۱۷-۲۵-۴۰-۳۲ -۱۵-۴۷-۳۹-۱۳-۸-۲۲-۵۰-۱۸-۴۱-۵۵ ۱۶-۴۵-۱۰-۴۶-۳۱-۱۴-۲۴-۴۳-۶	۰/۰۶۵
۳۹	-۱۱-۲۸-۲۹-۵۶-۲۰-۲۵-۳۶-۷-۵۷-۳۲ -۵۲-۱۶-۱۹-۳۴-۲۶-۵۱-۹-۴۰-۲۳-۳۷ -۴۷-۵۰-۴۸-۵-۳۹-۱۸-۲۱-۴۳-۱۰-۴۵ ۱۳-۴۶-۵۴-۴-۱۷-۳۳-۳۸-۱-۳۵	۰/۰۶۲
۴۰	-۲۱-۲۶-۵۲-۲۸-۱۲-۱۶-۳۲-۲۹-۳۳ -۳-۳۹-۲۲-۵۴-۸-۹-۳۶-۴-۳۰-۲-۵-۵۶ -۱۱-۱-۱۵-۴۳-۷-۱۸-۲۳-۳۵-۴۲-۴۶ ۱۰-۱۳-۴۷-۶-۲۷-۳۸-۳۴-۴۸-۴۴	۰/۰۶۷
۴۱	-۱۱-۱۶-۴۷-۲۰-۵۳-۱۰-۱-۲۸-۴۳-۲۶ -۵۲-۸-۴۱-۵۴-۱۳-۶-۲۷-۴۲-۲۴-۴-۳۲ -۷-۱۸-۱۲-۵۷-۲-۳۷-۴۵-۵-۳۳-۱۹ ۳۵-۵۱-۱۷-۳-۱۵-۴۰-۲۹-۵۰-۴۹-۵۵	۰/۰۷۱
۴۲	-۲۲-۴۴-۱۵-۵۰-۵-۲-۴۱-۲۶-۳۶-۵۳ -۲۰-۳۱-۳۵-۱۲-۴۵-۵۷-۷-۳۳-۸-۱۹ -۵۴-۴۲-۴۶-۲۴-۱۶-۱-۴-۱۱-۳۴-۲۳ -۴۰-۴۷-۱۳-۶-۲۱-۳۷-۴۹-۵۱-۲۹-۲۵ ۳۸-۳۲	۰/۰۵۷
۴۳	-۲۴-۳۳-۲۰-۵۳-۳۶-۱۰-۵-۱۲-۴۷-۱ -۴۲-۴۹-۲۷-۳۹-۳-۱۹-۴۶-۵۲-۳۲-۱۴ -۳۰-۱۷-۲۳-۴۰-۲۶-۱۱-۶-۳۴-۲۵-۱۳ -۴۴-۵۷-۱۸-۳۸-۵۴-۳۵-۱۵-۵۵-۳۱ ۵۶-۸-۱۶-۴۸	۰/۰۵۶
۴۴	-۲۳-۳۲-۱۳-۵۲-۴۹-۱۹-۳۴-۹-۴۳-۴ -۵۳-۴۷-۳۳-۳۰-۱۱-۴۰-۵-۱۶-۴۱-۱۲ -۲۰-۴۴-۲۹-۴۵-۳۸-۲۶-۲۱-۳۶-۵۷-۸ -۳-۵۱-۱۴-۳۵-۱-۴۸-۶-۴۶-۲۲-۲۷-۳۷ ۲۵-۱۰-۳۹	۰/۰۷۱
۴۵	-۵۴-۴۹-۲۸-۳۰-۳۷-۲-۲۹-۳۸-۵۵-۳۱ -۱۵-۵۱-۴۶-۵۶-۱۴-۱۲-۵-۵۰-۲۷-۴۳ -۵۲-۱-۷-۳۹-۲۰-۳۴-۱۷-۲۴-۲۳-۶-۴۷ -۹-۳۲-۴۰-۳-۴۴-۵۷-۴۲-۴۱-۱۶-۳۳ ۸-۵۳-۱۱-۱۰	۰/۰۶۶
۴۶	-۵۶-۲-۲۶-۹-۳۶-۱۷-۱۸-۵۳-۳۰-۵۵ -۶-۴۳-۵۴-۲۸-۱۱-۲۳-۴-۴۱-۳۳-۷-۱۹ -۲۹-۱۴-۵۱-۱۳-۵-۴۲-۱۶-۵۷-۳۵-۳۹ -۲۰-۳۴-۲۲-۳۱-۵۲-۱-۱۲-۱۵-۵۰-۳۲	۰/۰۶۸
	۴۴-۴۹-۴۸-۳-۳۷	
۴۷	-۵۲-۳۲-۴-۲۹-۸-۳۹-۴۱-۳۰-۴۷-۱۴ -۹-۲۴-۲۱-۵۴-۴۸-۲۵-۲۶-۴۵-۴۹-۳۱ -۱۳-۱۶-۲-۴۰-۵۷-۲۲-۴۳-۱۹-۳۶-۵ -۵۵-۱۲-۱۵-۱۱-۲۷-۱۸-۳۴-۳۳-۴۲ ۲۸-۲۳-۴۴-۵۱-۳-۲۰-۴۶-۳۷	۰/۰۵۴
۴۸	-۳۸-۴۹-۲۸-۴۸-۶-۴۴-۱۶-۲۵-۳۶-۱-۳ -۲۶-۱۲-۳۴-۱۱-۴۳-۸-۱۵-۲۰-۳۲-۱۴ -۲۴-۲۷-۵۴-۴-۳۷-۵۱-۳۳-۲۲-۴۲-۱۸ -۵۳-۴۵-۵۲-۲۹-۲-۵۰-۲۱-۵۷-۱۰-۴۶ ۴۰-۵-۱۹-۱۷-۳۱-۷-۵۵	۰/۰۶۳
۴۹	-۲۹-۲۳-۱۷-۴۱-۵۷-۴۶-۴-۱۳-۵-۶-۹ -۴۵-۱۲-۳۸-۳۷-۴۲-۵۲-۵۶-۲۴-۱۰ -۲۵-۲۷-۵۳-۱۴-۱۸-۳۶-۲-۷-۳۵-۱-۲۱ -۱۵-۱۱-۳۳-۲۲-۳۴-۱۶-۲۸-۴۸-۵۴ ۴۴-۵۰-۴۷-۲۶-۵۱-۳۲-۴۰-۵۵-۲۰	۰/۰۵۴
۵۰	-۱۱-۳۷-۴۲-۴-۲۱-۶-۲۲-۳۳-۲۸-۳۵ -۳۰-۱۴-۱-۱۶-۳۹-۲۰-۵۰-۳۸-۵۲-۱۳ -۹-۴۰-۱۷-۲۴-۱۹-۱۰-۲۵-۳۴-۲-۵۴ -۴۷-۳۱-۵-۲۳-۴۳-۲۹-۴۵-۱۵-۲۷-۵۵ ۵۱-۱۲-۸-۴۹-۵۳-۲۶-۱۸-۳-۴۱-۴۴	۰/۰۵۹
۵۱	-۳۹-۵۱-۵۴-۲۴-۱۲-۵۶-۱۷-۱۰-۱-۴۹ -۵۵-۲۶-۳۱-۹-۳۸-۲۰-۴۰-۱۹-۴۵-۴۸ -۲۷-۴۲-۳-۶-۴۳-۱۳-۱۵-۴-۲۹-۲۱-۳۶ -۱۸-۳۲-۱۶-۵۳-۵۲-۲۵-۵۷-۲۳-۱۴ -۲-۳۷-۴۱-۳۳-۵۰-۱۱-۲۸-۴۶-۳۰-۲۲ ۸	۰/۰۶۲
۵۲	-۲۴-۵۰-۲۰-۲۷-۴۲-۵۲-۵۵-۴۷-۳۷ -۴۵-۵-۶-۱۵-۲۳-۱۲-۴۶-۴۰-۲۵-۳۸ -۴۱-۲۱-۲۶-۱۱-۳۹-۱-۳۶-۳۰-۵۴-۵۷ -۴۸-۲۲-۸-۳۳-۳۲-۲۹-۴-۹-۵۳-۱۶-۵۶ -۱۴-۲-۴۳-۵۱-۱۳-۳-۳۵-۷-۳۱-۴۴ ۳۴-۱۹	۰/۰۵۷
۵۳	-۳۶-۳۵-۲۱-۹-۳۴-۱۳-۳-۴۳-۱۸-۱۶-۸ -۵۲-۱۱-۵۱-۲۹-۲۵-۲۲-۴۹-۴۱-۴۷ -۳۹-۵۳-۶-۳۳-۱-۵۷-۴۵-۱۲-۱۰-۲۴ -۵۰-۴۴-۴۸-۱۴-۳۲-۲۸-۳۰-۳۱-۲-۴۲ -۲۷-۲۰-۱۷-۳۸-۳۷-۵-۵۴-۲۳-۷-۴۶-۴ ۵۵-۵۶	۰/۰۵۲
۵۴	-۸-۴۹-۳۵-۴۴-۲۴-۴-۳۳-۶-۳۸-۵۱-۹ -۵۳-۱۷-۱۳-۵۴-۳۱-۳۶-۲۵-۵۰-۲۷	۰/۰۶۱

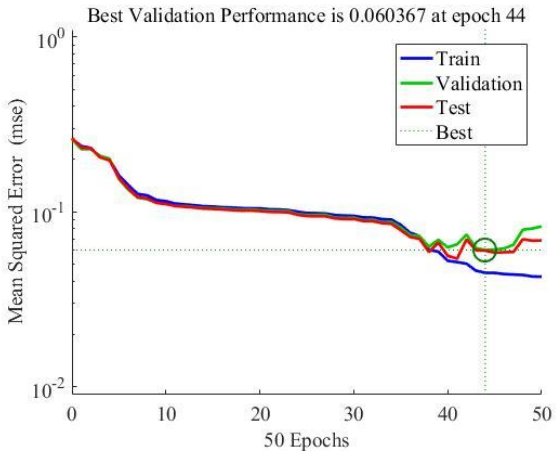
	۴۲-۱-۴۳-۲۰-۱۱-۵۶-۳۴-۲۶-۴۶-۲۱- ۲۳-۵۷-۱۴-۳۰-۵۵-۵۲-۴۸-۴۵-۳-۲۲- ۴۰-۲۹-۵-۱۸-۲۸-۱۵-۱۶-۳۷-۴۱-۲-۷- ۴۷-۳۹-۳۲	
۵۵	۲۲-۱۲-۳۸-۱۰-۱۵-۶-۷-۲-۲۷-۳۵-۱۴- ۳۲-۵۱-۲۶-۲۴-۴۷-۲۵-۹-۵-۲۸-۲۳- ۳۳-۱-۵۳-۴۵-۴۱-۳۰-۲۹-۴۴-۲۱-۴۳- ۸-۱۹-۱۱-۵۷-۵۰-۵۲-۱۳-۱۷-۱۶-۴۲- ۴۸-۲۰-۳۱-۳۴-۵۴-۵۶-۳۹-۵۵-۳۶- ۳۷-۴۹-۴۶-۳-۴	۰/۰۵۷
۵۶	۵۴-۲۲-۱۷-۳۲-۱۱-۵۲-۲۵-۵۱-۱۲- ۲۶-۳۵-۶-۴۳-۲۸-۵۰-۲-۱۳-۵۶-۸-۴۴- ۱۴-۲۹-۵۳-۳۸-۲۷-۵-۳-۴۰-۱۶-۴۱- ۱۰-۴۹-۱-۱۵-۹-۳۰-۵۵-۱۹-۳۹-۲۱- ۳۶-۳۱-۳۳-۴۶-۴۵-۵۷-۴۷-۳۷-۳۴- ۴۲-۱۸-۴۸-۷-۲۴-۲۳-۴	۰/۰۹۳
۵۷	۴۳-۴۷-۱۰-۵۴-۳۹-۱۴-۴۰-۴۲-۳۸- ۲۵-۱۲-۱۸-۹-۴۸-۷-۳۳-۵۵-۲۴-۵۳- ۳۵-۴۵-۵۲-۴-۱-۶-۱۵-۵-۵۶-۲۲-۲- ۱۷-۳۷-۲۸-۳۴-۸-۴۴-۲۱-۵۰-۱۳-۴۱- ۳۰-۵۷-۲۰-۴۹-۵۱-۳۱-۱۶-۳۶-۲۳- ۲۶-۴۶-۲۷-۳۲-۱۹-۲۹-۱۱-۳	۰/۰۶۸

اندیس‌های ۱ الی ۱۸، ۲۰ الی ۲۵، ۲۷ الی ۲۸ مربوط به درصد کلمات موجود در ایمیل، اندیس‌های ۴۹ الی ۵۴ مربوط به درصد کاراکترهای موجود در ایمیل و اندیس‌های ۵۵، ۵۶ و ۵۷ به‌ترتیب مربوط به متوسط، طولانی‌ترین و مجموع طول دنباله‌های متوالی از حروف بزرگ است.

پس برای تشخیص ایمیل‌های اسپم از این ۵۳ ویژگی استفاده خواهد شد. برای این‌منظور ۸۰ درصد مجموعه‌داده (۳۶۸۱ نمونه) به‌عنوان داده آموزش، ۱۰ درصد مجموعه‌داده (۴۶۰ نمونه) به‌عنوان داده آزمایش و ۱۰ درصد باقی‌مانده (۴۶۰ نمونه) به‌عنوان داده وارسی درنظر گرفته شده‌است.

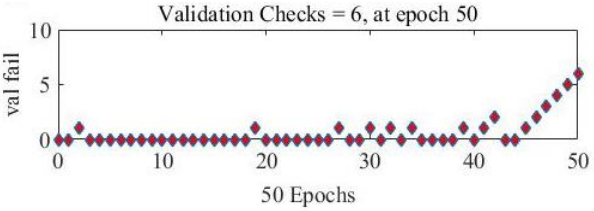
شکل (۴) نمودار عملکرد شبکه عصبی‌مصنوعی پرسپترون چند لایه را نشان می‌دهد. نمودار عملکرد شبکه از طریق میانگین مربع‌خطا قابل بررسی می‌باشد. این نمودار اعتبارسنجی و همچنین بهترین عملکرد شبکه را نشان می‌دهد.

روال آموزش در صورتی‌که خطای مجموعه‌داده وارسی در ۶ سیکل متوالی افزایش یابد، متوقف می‌شود. باتوجه به شکل نمودار عملکرد، این توقف در تکرار ۵۰ام رخ داده است. برای تأیید این موضوع شکل (۵) مقدار خطاها را نشان می‌دهد که در ۶ سیکل متوالی (۴۵ الی ۵۰) بهبودی حاصل نگردیده و همچنین صعود خطاها به‌سمت بالا است، پس باعث توقف شبیه‌سازی شده‌است. همان‌طور که در شکل قابل مشاهده است، خطای میانگین مربعات شبکه از یک مقدار بزرگ آغاز شده و رفته‌رفته کاهش می‌یابد. این امر بدان معناست که روال یادگیری شبکه دارای پیشرفت می‌باشد



(شکل -۴) :نمودار عملکرد شبکه عصبی‌مصنوعی پرسپترون چند لایه

(Figure-4): The diagram of Multilayer Perceptron Artificial Neural Network performance



(شکل -5): مقدار خطا در ۵۰ تکرار متوالی

(Figure-5): The Error value in 50 consecutive iterations

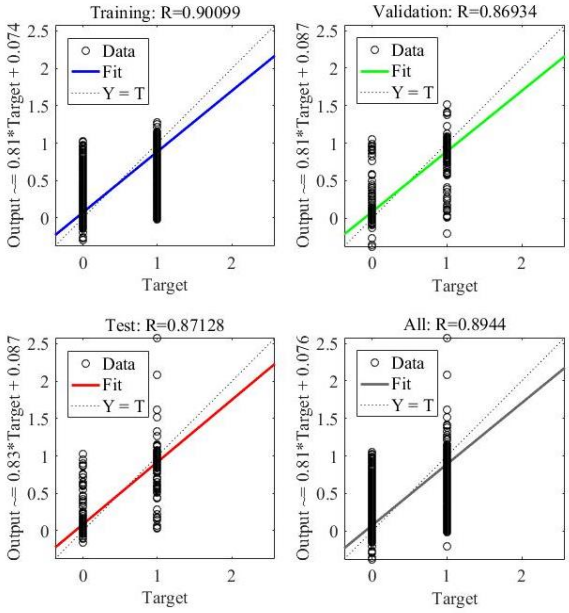
نمودار عملکرد شبکه در شکل (۴) شامل سه نمودار می‌باشد، زیرا کل مجموعه‌داده به‌صورت تصادفی به سه مجموعه آموزش، آزمایش و وارسی تقسیم شده‌است. مجموعه ارزیابی در راستای حفظ عمومیت شبکه مورداستفاده قرار می‌گیرد. روال آموزش تا زمانی ادامه می‌یابد که خطای داده‌های وارسی کاهش داشته باشد. به این ترتیب از بیش‌برازش شبکه برروی مجموعه آموزش جلوگیری می‌شود. بهترین عملکرد اعتبارسنجی برابر ۰/۰۶۰۳ در سیکل ۴۴ام می‌باشد. پس در کل می‌توان بدین نتایج دست یافت که:

- مقدار خطای میانگین مربعات نهایی کوچک است.

- خطای مجموعه آزمایشی با خطای مجموعه وارسی دارای رفتار و خصوصیات مشابهی است.

تا تکرار۴۴ام (که بهترین کارآیی در مورد مجموعه وارسی به‌وقوع می‌پیوندد) هیچ بیش‌برازشی رخ نداده است.

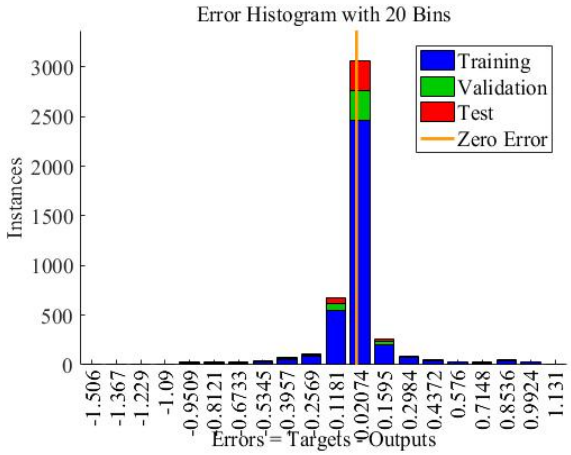
شکل (۶) ضریب همبستگی (R) را برای داده‌های آموزش، آزمایش و وارسی و کل داده‌ها را نشان می‌دهد. زمانی‌که ضریب همبستگی به یک نزدیک شود (در امتداد ۴۵ درجه) نشانگر آن است که عملکرد شبکه خوب و قابل‌اعتماد است. مقدار ضریب همبستگی برای داده‌های آموزش، آزمایش و وارسی به ترتیب برابر ۰/۹۰۹، ۰/۸۷۱ و ۰/۸۶۹ و برای کل داده‌ها برابر ۰/۸۹۴ است. پس باتوجه به نتایج، ضریب همبستگی متمایل به یک است، پس عملکرد شبکه خوب و قابل‌اعتماد است.



(شکل -۶): نمودار همبستگی داده‌های آموزش، آزمایش و وارسی

(Figure-6): Correlation diagram of training, testing and verification data

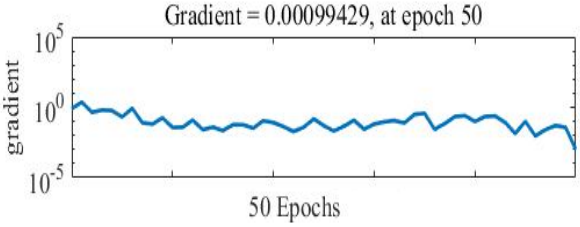
شکل (۷) نمودار هیستوگرام خطای داده‌های آموزش، ارزیابی و وارسی را نشان می‌دهد. میله آبی مربوط به داده‌های آموزش، میله قرمز مربوط به داده‌های آزمایش و میله سبز مربوط به داده‌های وارسی می‌باشد. در واقع نمودار هیستوگرام خطای داده‌های آموزش، آزمایش و وارسی هر دسته از داده‌های به‌ازای خطاهای مختلف مورد بررسی قرار می‌گیرد. نمودار هیستوگرام می‌تواند نقاط پرت را نشان دهد. در این نمودار، بهترین حالت موجود تنها یک نمودار میله‌ای در خطای صفر است که در این شبیه‌سازی بهترین حالت مربوط به خطای ۰/۰۲۰۷ می‌باشد.



(شکل -۷) : نمودار هیستوگرام خطای داده‌های آموزش، آزمایش و وارسی

(Figure-7): Error histogram diagram of training, testing and verification data

شکل (۸) نمودار گرادیان در ۵۰ سیکل متوالی را نشان می‌دهد. همان‌طور که قابل مشاهده است، مقدار گرادیان در سیکل ۵۰ام برابر۰/۰۰۰۹۹ می‌باشد.



(شکل -۸): نمودار گرادیان در ۵۰ سیکل متوالی

(Figure-8): Gradient diagram in 50 consecutive cycles

در ادامه از شبکه عصبی مصنوعی تابع پایه شعاعی نیز به‌عنوان هدف استفاده شده‌است تا نتایج آن با شبکه عصبی مصنوعی پس‌انتشار بررسی شود. این شبکه شامل دو لایه گره (به‌جز لایه ورودی که داده‌های ورودی به آن اعمال می‌شود) می‌باشد. خروجی گره‌های لایه نهان در وزن‌های مناسب ضرب شده و خروجی شبکه عصبی مصنوعی تابع پایه شعاعی را تولید می‌کند[22].

به‌طور کلی با آموزش شبکه‌های عصبی‌مصنوعی پرسپترون چند لایه و شبکه های عصبی مصنوعی تابع پایه شعاعی توانسته‌ایم خطا را کاهش داده و به تشخیص ایمیل‌های اسپم بپردازیم. اما با مقایسه عملکرد شبکه‌های عصبی‌مصنوعی پرسپترون چند لایه و تابع پایه شعاعی می‌توان نتیجه گرفت که عملکرد شبکه عصبی‌مصنوعی پرسپترون چند لایه بهتر از تابع پایه شعاعی بوده است.

برای ارزیابی عملکرد شبکه عصبی‌مصنوعی پرسپترون چندلایه و تابع پایه شعاعی در تشخیص ایمیل‌های اسپم از معیارهای میانگین خطای مطلق، انحراف استاندارد خطای مطلق، میانگین مربع‌خطا، جذر میانگین مربع‌خطا، دقت، صحت و کارآیی استفاده شده‌است، که نتایج آن‌ها به‌طور خلاصه در جدول (۲) قابل مشاهده است.

(جدول -۲): نتایج ارزیابی خطای شبکه‌های عصبی‌مصنوعی پرسپترون چند لایه و تابع پایه شعاعی

(Table-2): Error evaluation results of multilayer perceptron artificial neural networks and radial basis function

	پرسپترون چندلایه	تابع پایه شعاعی
میانگین خطای مطلق	۰/۰۲۸۰	۰/۰۳۶۷
انحراف استاندارد خطای مطلق	۰/۲۵۲۴	۰/۳۲۹۸
میانگین مربع خطا	۰/۰۶۴	۰/۱۰۶۷
جذر میانگین مربع خطا	۰/۲۵۳۸	۰/۳۲۶۷
دقت	۹۲/۷۸۱۶	۹۱/۰۴۲۴
صحت	۹۴/۳۷۸۲	۹۲/۲۲۷۹
کارآیی	۹۵/۱۹۶۹	۹۴/۹۲۲۳

همان‌طور که قابل‌مشاهده است، خطای شبکه عصبی‌مصنوعی پرسپترون کمتر از تابع پایه شعاعی است. پس عملکرد شبکه عصبی‌مصنوعی پرسپترون چندلایه در تشخیص ایمیل‌های اسپم بهتر از تابع پایه شعاعی می‌باشد.

همچنین دقت، صحت و کارآیی شبکه عصبی‌مصنوعی پرسپترون چند لایه بیشتر از تابع پایه شعاعی است. پس به‌طور کلی می‌توان نتیجه گرفت تشخیص ایمیل‌های اسپم براساس این ۵۳ ویژگی مشخص‌شده و شبکه عصبی‌مصنوعی پرسپترون چند لایه بهتر از تابع پایه شعاعی است.

برای بررسی بیشتر عملکرد روش پیشنهادی به مقایسه نتایج آن با نتایج Renuka و همکاران در [17] و Rusland و همکاران در [18]، Lee، Lu و Huang در[21] و Wang و همکاران در [۱۲] می‌پردازیم.

نتایج کارآیی Renuka و همکاران و Rusland و همکاران و Lee، Lu و Huang در جدول (۳) نشان داده شده است. باتوجه به جدول (۳)، کارآیی شبکه‌های عصبی‌مصنوعی پرسپترون چند لایه (۹۵/۱۹) و تابع پایه شعاعی (۹۴/۹۲) با ۵۳ ویژگی استخراج‌شده بهتر از کارآیی روش‌های بیزین ساده (۸۰/۷۷) و ارتقای گرادیان (۹۴/۱) توسط Renuka و همکاران و روش‌های بیزین ساده توسط Rusland و همکاران و درخت تصمیم‌گیری (۹۲/۶)، ماشین بردار پشتیبان (۹۱/۵)، شبکه عصبی‌مصنوعی پس‌انتشار (۹۰/۴)، الگوریتم Kنزدیکترین همسایه (۹۰/۸) و درخت تصمیم‌گیری مبتنی‌بر زنبورعسل مصنوعی(۹۳/۷) ارائه‌شده توسط Lee، Lu و Huang است.

(جدول -۳): مقایسه نتایج کارآیی روش پیشنهادی با نتایج محققان

(Table-3): Comparison of the efficiency results of the proposed method with the results of researchers

روش	کارآیی
روش بیزین ساده [۱۷]	۸۰/۷۷
روش ارتقای گرادیان [۱۷]	۹۴/۱
روش بیزین ساده و مجموعه‌داده اسپم [۱۸]	۹۱/۱۳
روش بیزین ساده و مجموعه‌داده SPAMBASE [۱۸]	۸۲/۵۴
درخت تصمیم‌گیری [۲۱]	۹۲/۶
ماشین بردار پشتیبان [۲۱]	۹۱/۵
شبکه عصبی‌مصنوعی پس‌انتشار [۲۱]	۹۰/۴
الگوریتم K نزدیکترین همسایه [۲۱]	۹۰/۸
درخت تصمیم‌گیری مبتنی‌بر زنبورعسل مصنوعی [۲۱]	۹۳/۷
پرسپترون چند لایه	۹۵/۱۹
تابع پایه شعاعی	۹۴/۹۲

Wang و همکاران در [۱۲] برای تشخیص ایمیل‌های اسپم از مدل ترکیبی ماشین‌بردار پشتیبان و الگوریتم ژنتیک براساس ۵۲ و ۵۷ ویژگی استفاده کرده‌اند. برای انجام مقایسه شبکه‌های عصبی‌مصنوعی پیشنهادی در این مقاله با نتایج Wang و همکاران از استخراج ۵۲ ویژگی استفاده شده‌است. هر چند متأسفانه آن‌ها اندیس این ۵۲ ویژگی را ارائه نکرده‌اند. بنابراین از استخراج ۵۲ ویژگی مقاله استفاده شد. نتایج میانگین مربع خطای داده‌های آزمایش Wang و همکاران و شبکه‌های عصبی‌مصنوعی پرسپترون چند لایه و تابع پایه شعاعی در جدول (۴) نشان داده شده‌است. باتوجه به نتایج، درصد میانگین مربع خطای شبکه‌های عصبی‌مصنوعی پیشنهادی کمتر از نتایج Wang و همکاران است.

روش	میانگین مربع خطا ۵۲ ویژگی	۵۷ ویژگی
Wang و همکاران [۱۲]	۸۷/۸۹	۸۷/۷۳
پرسپترون چند لایه	۹۴/۳۴	۹۳/۲۱
تابع پایه شعاعی	۹۲/۵۱	۹۲/۱۷

۵-نتیجه‌گیری وپژوهش های آتی

یکی از مشکلات شایع در دنیای فناوری اطلاعات و اینترنت وجود ایمیل‌های اسپم برای کاربران استفاده‌کننده از این فناوری می‌باشد. برای رفع این مشکل و شناسایی ایمیل‌های اسپم در این مقاله، از شبکه‌های عصبی مصنوعی پرسپترون چندلایه با الگوریتم جهش قورباغه استفاده‌شد. برای این‌منظور از مجموعه‌داده ایمیل‌های اسپم سایت UCI استفاده گردید. ابتدا به کمک الگوریتم جهش قورباغه اندیس بهترین ویژگی‌های مجموعه‌داده به‌دست آمد که شامل ۵۳ ویژگی از ۵۷ ویژگی بود. این اندیس‌ها برابر ۱ الی ۱۵، ۱۶ الی ۱۸، ۲۰ الی ۲۵، ۲۷ الی ۲۸، ۴۹ الی ۵۴ و ۵۵ الی ۵۷ بود. سپس با کمک این بهترین ویژگی‌ها به تشخیص ایمیل‌های اسپم با کمک شبکه‌های عصبی‌مصنوعی پرسپترون چند لایه و تابع پایه شعاعی پرداخته شد. میانگین مربع خطا، دقت، صحت و کارآیی شبکه عصبی‌مصنوعی پرسپترون چند لایه به‌ترتیب برابر ۰/۰۶۴، ۹۲/۷۸، ۹۴/۳۷، ۹۵/۱۹ و برای شبکه عصبی‌مصنوعی تابع پایه شعاعی به‌ترتیب برابر ۰/۱۰۶۷، ۹۱/۰۴، ۹۲/۲۲ و ۹۴/۹۲ به‌دست آمده است. که این نتایج نشان‌دهنده عملکرد بهتر شبکه عصبی‌مصنوعی پرسپترون چند لایه نسبت به تابع پایه شعاعی است.

۷ – مراجع

[1] *Aski, A.S & Sourati, N.K. "**Proposed efficient algorithm to filter spam using machine learning techniques**". Pacific Science Review A: Natural Science and Engineering. Vol 18. 145-149. 2016.*

[2] *Lacey, D., Salmon, P., & Glancy, P. "**Taking the Bait: A Systems Analysis of PhishingAttacks**". ProcediaManufacturing, 3 , 1109-1116. 2015.*

[3] *Rasdale, M. "**Legislating for robots & zombies"**. Computer Law & Security Review, 22(3), 222-227. 2006.*

[4] *Ryan, M. D. "**Cloud computing security: The scientific challenge**", and a survey of solutions. Journal of Systems and Software, 86(9), 2263-2268. 2013.*

[5] *Kufandirimbwa, O. & Gotora, R. "**Spam Detection Using Artificial Neural Networks (Perceptron Learning Rule)**". Online Journal of Physical and Environmental Science Research. Vol 1. pp. 22-29. 2012.*

[6]

[7] *Abdelhamid, N., Ayesh, A., & Thabtah, F. "**Phishing detection based associative classification data mining**". Expert Systems with Applications, 41(13), 5948-5959. 2014.*

[8] *Xu, H., & Yu, B. "**Automatic thesaurus construction for spam filtering using revised back propagation neural network**". Expert Systems with Applications, 37(1), 18-23. 2010.*

[9] *Zang, H., Zhang, S., & Hapeshi, K. "**A review of nature-inspired algorithms**". Journal of Bionic Engineering, 7, S232-S237. 2010.*

[۱۰] اسدپور، وحید. ***"اصول بنیادی ومرجع کاربردی شبکه های عصبی"*** .جلد اول .چاپ دوم،تهران:انتشارات آتی نگر. ۱۳۹۳.

[۱۱] منهاج، محمدباقر. ***"هوش محاسباتی (مبانی شبکه‌های عصبی)"*** . جلد اول، چاپ دوم، تهران، مرکز نشر دانشگاه صنعتی امیرکبیر. ۱۳۸۱.

[۱۲] مهرگان، مهدوی،.حبیب، آهکی. بابک، ناصرشریف. "طراحی یک سیستم تشخیص اسکناس مبتنی بر شبکه عصبی با استفاده از مشخصه های بافت و رنگ تصویر". دوره ۷، شماره ۲. ۱۳۸۹.

[13] *H. Wang, at el. "**SVM classifier incorporating feature selection using GA for spam detection**", In International Conference*

Procedia Computer Science. Vol. 47. P. 76 – 83.

*[25] Karthika Renuka, d. & Visalakshiand SP. Rajamohana, p. "**An Ensembled Classifier for Email Spam Classification in Hadoop Environment**". Applied Mathematics & Information Sciences. No. 4. 1123-1128. 2017.*

*[26] Rusland, n.f., Wahid, n., Kasim, s., Hafit H. "**Analysis of Na¨ıve Bayes Algorithm for Email Spam Filtering across Multiple Datasets**". International Research and Innovation Summit. 2017.*

*[27] Q. Fu, B. Feng, D. Guo, Q. Li, **Combating the evolving spammers in online social networks**, Computers & Security, Vol. 72, pp. 60-73, 2018.*

*[28] F.S. Gharehchopogh, M. Vafadar, M. Motaman, **Improve of Invasive Weed Optimization algorithm with K nearest neighbor for email spam classification**, Computing Science Journal (CSJ), in press, March 2018.*

[29] *H. Faris, A.M. Al-Zoubi, A.A. Heidari, I. Aljarah, H. Fujita, **An intelligent system for spam detection and identification of the most relevant features based on evolutionary Random Weight Networks**, Information Fusion, Vol. 48, pp. 67-83, 2019.*

on Embedded and Ubiquitous Computing, Springer Berlin Heidelberg, pp. 1147-1154. 2005.

*[14] Dinh, S., Azeb, T., Fortin, F., Mouheb, D., & Debbabi, M. "**Spam campaign detection, analysis, and investigation**". Digital Investigation, 12, S12-S21. 2015.*

*[15] Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. "**Detecting spammers on social networks". ***Neurocomputing, 159, 27-34. 2015.*

*[16] Alsmadi, I., & Alhami, I. "**Clustering and classification of email contents**". Journal of King Saud University-Computer and Information Sciences, 27(1), 46-57. 2015.*

*[17] Ghosh, S., Biswas, S., Sarkar, D., & Sarkar, P. P. "**A novel Neuro-fuzzy classification technique for data mining**". Egyptian Informatics Journal,15 (3), 129-147. 014.*

*[18] Karthika Renuka, d. & Visalakshiand SP. Rajamohana, p. "**An Ensembled Classifier for Email Spam Classification in Hadoop Environment**". Applied Mathematics & Information Sciences. No. 4. 1123-1128. 2017.*

*[19] Rusland, n.f., Wahid, n., Kasim, s., Hafit H. "**Analysis of Na¨ıve Bayes Algorithm for Email Spam Filtering across Multiple Datasets**". International Research and Innovation Summit. 2017.*

*[20] Gupta, D.K. & Goyal, S. "**Email Classification into Relevant Category Using Neural Networks**". arXiv preprint arXiv:1802.03971. 2017.*

*[21] EUSUFF, M.; LANSEY, K.; PASHA, F. "**Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization**". Engineering Optimization. Vol. 38, No. 2. 2006.*

*[22] Lee, J.Z., Lu, T.H., Huang, H. "**A novel algorithm applied to filter spam e-mails for iPhone**". Vietnam J Comput Sci. Vol.2, pp. 143–148. 2015.*

*[۲۳] قانعی یخدان، حسین. **"روشی جدید برای اختفای خطا در فریم‌های ویدئو با استفاده از شبکه عصبی RBF"**. پردازش علائم و داده‌ها. ۱۰ (۱) :۱۲-۳. ۱۳۹۲.*

*[24] T. Santhanam a, M.S Padmavathi b. (2015). "**Application of K-Means and Genetic Algorithms for Dimension Reduction by Integrating SVM for Diabetes Diagnosis**".*

روش ارجاع: ا.حیدریان، ف.سلیمانیان قره‌چپق، ارائه راهکاری نوین جهت بهبود شبکه های عصبی مصنوعی پرسپترون چند لایه مبتنی بر ترکیب با الگوریتم جهش قورباغه برای تشخیص ایمیل‌های اسپم، فصلنامه محاسبات و سامانه های توزیع شده ،سال پنجم، شماره۲، شماره پیاپی۱۰، صفحه ۸۸ تا ۱۰۱، سال ۱۴۰۱

How to cite: A.Heydariyan ,F.Soleymanian QareChopoq, Presenting a novel method to improve multi-layered perceptron artificial neural networks based on combination with frog leaping algorithm to detect spam emails , Journal of Distributed Computing and Systems(JDCS), Vol 5, Issue 2, Page 88-101, 2023.

احمد حیدریان مدرک کارشناسی خود را در رشته مهندسی کامپیوتر–نرم افزار از دانشگاه ارومیه درسال ۱۳۹۴وکارشناسی ارشد خود را در همان رشته از دانشگاه آزاد اسلامی واحد مراغه در سال ۱۳۹۶اخذ نموده است درحال حاضردانشجوی دکتری مهندسی کامپیوتر گرایش نرم افزارمیباشند.زمینه های پژوهشی مورد علاقه ایشان داده کاوی والگوریتمهای یادگیری ماشین ،پردازش زبان طبیعی وشبکه های اجتماعی است.

نشانی رایانامه ایشان عبارت است از :

heydariyan1661@gmail.com

فرهاد سلیمانیان قره چپق مدرک کارشناسی ارشد ودکترای تخصصی خود را در رشته مهندسی کامپیوتر به ترتیب از دانشگاه چو کوروا ودانشگاه حاجت تپه کشور ترکیه اخذ نموده است.وتخصص وی پردازش زبان طبیعی،داده کاوی والگوریتمهای یاد گیری ماشین است .ایشان عضوهیئت علمی دانشکده مهندسی کامپیوتر دانشگاه آزاد اسلامی واحد ارومیه بوده وبه تدریس دروس مختلف در حوزه کاری خویش مشغول است.

نشانی رایانامه ایشان عبارت است از :

farhad@iaurmia.ac.ir