علوم رایانش و فناوری اطلاعات
نشریه علمی انجمن کامپیوتر ایران
مجلد ۲۱، شماره ۱، ۱۴۰۲
صفحات ۱۹-۲۸
مقاله پژوهشی

رویکرد یادگیری انتقالی مبتنی بر فضای معنایی مشترک برای شناسایی پیوسته کاربران تلفن همراه

محدثه عالی[۱]، فاطمه سادات لسانی[۲]، فرانک فتوحی قزوینی[۳*]

*نویسنده مسئول، دریافت: ۱۴۰۲/۰۶/۰۱، بازنگری: ۱۴۰۲/۱۰/۰۱، پذیرش: ۱۴۰۲/۱۰/۰۹

[۱] کارشناسی ارشد، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران
[۲] استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی قم، قم، ایران
[۳] استادیار، گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه قم، قم، ایران

چکیده

افزایش روزافزون استفاده از تلفن‌های هوشمند و پیشرفت چشمگیر قابلیت‌های آن‌ها، این ابزار را وسیله‌ای در دسترس برای عموم مردم تبدیل کرده است. از طرفی ذخیره‌سازی اطلاعات شخصی و داده‌های محرمانه در تلفن‌های هوشمند، این وسیله را همواره در معرض تهدیدات امنیتی قرار داده است. روش‌های مرسوم احراز هویت تک‌مرحله‌ای و توانایی شناسایی پیوسته کاربران پس از تأیید هویت اولیه و بازشدن قفل تلفن همراه، امنیت کافی را ندارند. این پژوهش از روش احراز هویت زیست‌سنجی «الگوی رفتاری کاربر در زمان استفاده از برنامه‌های کاربردی[۱] تلفن همراه» برای شناسایی پیوسته کاربران تلفن‌های هوشمند استفاده می‌کند. معضل اصلی در انواع روش‌های زیست‌سنجی رفتاری، تغییر رفتار کاربر پس از گذر زمان و در پی آن نداشتن الگوی رفتاری ثابت در درازمدت می‌باشد. این امر موجب کاهش دقت سیستم احراز هویت می‌گردد و در پی آن امنیت کاربر را تحت‌تأثیر قرار می‌دهد. در این مقاله، روشی نوین برای حل مسئله شناسایی پیوسته در گوشی‌های تلفن همراه بر اساس رویکرد یادگیری انتقالی ارائه شده است. روش پیشنهادی، راهکاری برای تعریف فضای معنایی مشترک با هدف یکسان‌سازی فضای ویژگی مبدأ و مقصد و فراهم‌کردن امکان استفاده از مبحث یادگیری انتقالی را پیشنهاد می‌دهد. بر اساس آزمایش‌های انجام‌شده نشان داده‌ایم که راهکار ارائه‌شده در مقاله حاضر، هویت کاربر را به‌صورت پیوسته و ضمنی با دقت[۲]و عملکرد مناسب احراز می‌کند و نسبت به روش‌های دیگر برتری دارد.

کلمات کلیدی: شناسایی پیوسته، احراز هویت مبتنی بر زیست‌سنجی، احراز هویت در تلفن همراه، الگوهای رفتاری، استفاده از برنامه کاربردی، یادگیری انتقالی

۱- مقدمه

امروزه گوشی‌های هوشمند، جزء دائمی و جدانشدنی زندگی انسان‌ها شده‌اند؛ به‌طوری که افراد بسیاری داده‌های شخصی و محرمانه خود همچون عکس، فیلم و رمزهای بانکی خویش را در تلفن‌های همراه ذخیره می‌کنند بنابراین تلفن‌های همراه همواره مستعد تهدیدهای امنیتی و در معرض سرقت افراد سودجو هستند. به همین منظور پژوهشگران حوزه تلفن‌های هوشمند و امنیت، روش‌های مختلفی را برای حفظ امنیت و آسودگی خاطر کاربر ارائه کرده‌اند [۱, ۲]: احراز هویت مبتنی بر دانش[۳]، احراز هویت مبتنی بر توکن[۴] و احراز هویت مبتنی بر زیست‌سنجی[۵].

رایج‌ترین روشی که تاکنون مورد استفاده قرار گرفته است روش احراز هویت تک‌مرحله‌ای مبتنی بر دانش یا همان روش سنتی احراز هویت مانند رمز عبور و قفل الگو می‌باشد [۱]. در این روش پس از تأیید هویت اولیه و بازشدن قفل ابتدایی، کاربر بدون نیاز به تأیید هویت پیوسته و مجدد، اجازه دسترسی به اطلاعات گوشی هوشمند را دارا می‌گردد. این دسترسی آسان فارغ از مزایا، به علت عدم‌پیوستگی شناسایی کاربر و تأیید هویت دائمی او در طول زمان استفاده از تلفن همراه، امنیت این وسیله را همواره مورد تهدید قرار می‌دهد. از این‌رو، ارائه یک روش ایمن برای

شناسایی و احراز هویت پیوسته کاربر در گوشی‌های هوشمند، بسیار ضروری می‌نماید.

سیستم احراز هویت مبتنی بر زیست‌سنجی، یکی از ایمن‌ترین سیستم‌های حفاظت از اطلاعات شخصی شناخته‌شده است [۳]. این روش از مشخصه‌ها و صفات قابل اندازه‌گیری فیزیولوژیکی و رفتاری منحصربه‌فرد بدن انسان برای شناسایی و تأیید کاربران استفاده می‌کند [۴, ۵]. روش‌های احراز هویت مبتنی بر مشخصات فیزیولوژیکی، خصوصیات فیزیکی منحصربه‌فرد بدن انسان را که با گذر زمان، ثابت می‌مانند؛ مانند اثر انگشت، چهره، هندسه دست، عنبیه یا شبکیه اندازه‌گیری می‌کنند [۵] اما در روش‌های احراز هویت مبتنی بر خصوصیات رفتاری، یک الگوی رفتاری برای کاربر بر اساس تحت نظر گرفتن ویژگی‌های رفتاری و فعالیت‌های او در یک دوره زمانی ساخته می‌شود. روش‌های الگوی راه رفتن فرد[۶] [۶]، حرکات لمسی پویا[۷] [۷, ۸]، الگوی وب‌گردی[۸] [۹]، ضربه به صفحه کلید[۹] [۱۰] و الگوی رفتاری استفاده از برنامه‌های کاربردی[۱۰][۱۱] در دسته روش‌های احراز هویت مبتنی بر زیست‌سنجی رفتاری قرار می‌گیرند. روش‌های زیست‌سنجی رفتاری در مقایسه با روش‌های زیست‌سنجی فیزیولوژیکی دارای دقت پایین‌تر و عدم‌پایداری به علت نبود ثبات در رفتار انسان می‌باشند [۱, ۱۲]. با این وجود از دو مزیت مهم، بهره‌مند هستند: امکان احراز هویت پیوسته و ضمنی و نیازنداشتن به سخت‌افزارها و اسکنرهای مخصوص [۱]. یکی از روش‌های احراز هویت مبتنی بر زیست‌سنجی رفتاری، شناسایی کاربران بر اساس استفاده آن‌ها از برنامه‌های کاربردی و خدمات تلفن هوشمند می‌باشد. در زمان احراز هویت، رفتار و فعالیت‌های فعلی کاربر مانند ایجاد تماس یا استفاده از یک برنامه کاربردی خاص با نمایه‌ای که از تاریخچه رفتارهای کاربر ساخته شده است، مقایسه می‌شود [۱, ۱۲]. در این روش برخلاف روش‌های تشخیص صدا، تشخیص امضا یا الگوی راه رفتن، امکان تغییر رفتار تحت شرایط ناخواسته از قبیل سرماخوردگی وجود ندارد. همچنین این روش برخلاف روش‌های ضربه به صفحه کلید و صفحه لمسی پویا، ثبات دقت بیشتری را داراست [۱]. همچنین در تمامی روش‌های زیست‌سنجی رفتاری امکان تقلید رفتار و حمله وجود دارد که در روش استفاده از برنامه‌های کاربردی تلفن هوشمند احتمال این امکان بسیار پایین است [۱]. در مجموع با بررسی و مقایسه انواع روش‌های زیست‌سنجی، علی‌رغم دقت پایین‌تر روش‌های زیست‌سنجی رفتاری، به دلیل فراهم‌بودن امکان احراز هویت پیوسته بدون اختلال در عملکرد تلفن همراه، روش «شناسایی الگوی رفتاری کاربر با استفاده از برنامه‌های کاربردی تلفن هوشمند» با هدف مقاله پیش رو، یعنی حفظ پیوسته امنیت، سازگاری بیشتری را نسبت به روش‌های دیگر داراست.

معضل اصلی مطرح در حوزه احراز هویت مبتنی بر زیست‌سنجی رفتاری، ازجمله روش الگوی رفتاری استفاده از برنامه‌های کاربردی، تغییر رفتار کاربر با گذر زمان و کاهش میزان تطابق الگوی رفتاری جدید با الگوی رفتاری قبلی کاربر و کاهش دقت، امنیت و قابلیت استفاده در پی آن می‌باشد. در اینجا منظور از تغییر رفتار کاربر، مجموعه برنامه‌های کاربردی است که کاربر طی یک دوره از آن‌ها استفاده می‌کرده و ممکن است با گذر زمان و تحت شرایط محیطی متغیر، نوع، مدت‌زمان استفاده یا ترتیب استفاده از آن‌ها تغییر کند. یک راه‌حل این است که مدل را از ابتدا و با استفاده از داده‌های آموزشی جمع‌آوری‌شده جدید بازسازی کرد. از سویی دیگر جمع‌آوری دوباره داده‌ها برای آموزش و ساخت دوباره مدل، گاهی هزینه‌بر یا غیرممکن می‌باشد؛ از این‌رو باید به دنبال راهی برای کاهش هزینه و بدون نیاز به تلاش برای جمع‌آوری و برچسب‌گذاری مجدد داده‌ها بود. در مقاله حاضر با انتخاب روش «شناسایی الگوی رفتاری کاربر با استفاده از برنامه‌های کاربردی»، راهکار فضای معنایی مشترک به‌منظور یکسان‌کردن فضای ویژگی مبدأ و مقصد و فراهم‌شدن امکان انتقال دانش ارائه می‌گردد. بنابراین با گذر زمان و تغییر برنامه‌های کاربردی مورداستفاده کاربر، مدل احراز هویت آموزش داده‌شده، می‌تواند به‌خوبی مالک اصلی را شناسایی کند.

ساختار ادامه مقاله بدین‌گونه است: در بخش دوم مروری بر پژوهش‌های مرتبط به حوزه شناسایی پیوسته بر اساس الگوی رفتاری استفاده از برنامه‌های کاربردی تلفن هوشمند خواهد شد. در بخش سوم به بررسی دقیق شیوه عملکرد شناسایی پیوسته و راهکار ارائه‌شده به‌منظور حل معضل این پژوهش، پیاده‌سازی آن و ساخت مدل احراز هویت می‌پردازیم. نتایج حاصل از ساخت مدل و تحلیل آن‌ها در بخش چهارم بررسی خواهد شد. بخش پنجم نیز به نتیجه‌گیری و ارائه پیشنهاد برای پژوهش‌های آینده اختصاص یافته است.

۲- کارهای مرتبط

تحقیقات صورت‌گرفته در زمینه احراز هویت بر اساس الگوی رفتاری استفاده از برنامه‌های کاربردی تلفن هوشمند به دو دسته مبتنی بر شبکه و مبتنی بر میزبان تقسیم شده‌اند. آنچه درصدد بازخوانی آن هستیم مقالاتی است که مبتنی بر میزبان هستند زیرا تمرکز اصلی آن بر روی برنامه‌های کاربردی مورداستفاده کاربر می‌باشد که با هدف مقاله حاضر، همخوانی دارد.

یک رویکرد، مبتنی بر روش امتیازدهی بر اساس فعالیت‌های اخیرکاربر در [۱۳] ارائه گردیده است. با جمع‌آوری اطلاعات رفتاری مربوط به تماس‌ها، پیام‌های متنی، استفاده از مرورگر و مکان کاربر، نمایه وی ایجاد می‌گردد. در این روش، امتیاز بر اساس شناسایی رویدادهای منطبق بر عادات همیشگی افزایش و بر اساس شناسایی فعالیت‌های جدید کاهش می‌یابد. همچنین گذر زمان منجر به کاهش امتیاز می‌شود. در نهایت اگر امتیاز احراز هویت از حد آستانه کمتر شود، کاربر باید به‌طور صریح (مانند واردکردن رمزعبور) احراز هویت گردد.

لی و همکاران [۱۴] یک روش احراز هویت بر اساس مشخصات رفتاری با تمرکز بر روش مبتنی بر میزبان ارائه کرده‌اند. آن‌ها با استفاده از پیام‌های متنی و رفتارهای مربوط به تماس‌ها، کاربران را شناسایی می‌کنند. همچنین می‌توان رفتارهای غیرطبیعی را بر اساس تعامل آن‌ها با برنامه‌های تلفن همراه شناسایی کرد.

در ارتباط با روش‌های احرازهویت داده‌محور، کایاجیک و همکاران [۱۵] یک روش داده‌محور احراز هویت ضمنی مبتنی بر حسگرها را ارائه کرده‌اند. در این مقاله، صرفاً از داده‌های برنامه‌های کاربردی مورداستفاده کاربر مانند اطلاعات مکانی، زمانی یا حسگرها برای ساخت نمایه کاربر استفاده می‌گردد و توجهی به رفتار کاربر در ترتیب یا مدت‌زمان استفاده از برنامه‌های وی نخواهد شد. همچنین یک حد آستانه به‌صورت خودکار برای سیستم درنظر گرفته شده است. در صورت تغییر رفتار کاربر با گذشت زمان، مدل رفتاری از حالت پایدار خارج می‌شود و نیاز به بازآموزی پیدا می‌کند.

فریدمن و همکاران [۹] یک روش احراز هویت پیوسته مبتنی بر ترکیب روش‌های زیست‌سنجی رفتاری ارائه کرده‌اند. در این مقاله چهار روش تحلیل متن، الگوی استفاده از برنامه کاربردی، رفتار مرورگر وب و مکان فیزیکی دستگاه مطالعه شده و همچنین به توازن بین مدت‌زمان تشخیص سودجو و خطا در تشخیص مالک اصلی توجه گردیده است. نتایج ارزیابی‌ها نشان داد طبقه‌بندی‌کننده مبتنی بر مکان فیزیکی دستگاه و پس از آن مرورگر وب، بیشترین کمک را به عملکرد سیستم ترکیبی کرده‌اند.

در مطالعه دیگری [۱۶]، یک مدل احراز هویت پیوسته مبتنی بر نمایه رفتاری کاربر ارائه گردیده و بر اساس الگوریتم‌های یادگیری ماشین با نظارت، ارزیابی شده است. ویژگی‌های به‌کارگرفته‌شده در این مطالعه شامل نام برنامه کاربردی، نوع فعالیت و اطلاعات اضافی مانند طول پیامک یا ایمیل، مدت‌زمان تماس و تاریخ می‌باشد. در این مطالعه، سه طبقه‌بندی[۱۱]کننده ماشین بردار پشتیبان[۱۲]، جنگل تصادفی[۱۳] و گرادیان تقویتی[۱۴] برای ارزیابی کارآمدی آن‌ها بررسی شده است. در این پژوهش، عملیات احراز هویت به‌ازای هر فعالیت در تلفن همراه صورت می‌گیرد. نتایج

این مطالعه نشان داد روش گرادیان تقویتی با میانگین نرخ خطای برابر ۲۷ درصد، عملکرد بهتری نسبت به سایر روش‌ها داشته است.

محبوب و همکاران [۱۱] روشی تجربی برای احراز هویت پیوسته کاربران با استفاده از مدل مخفی مارکوف ارائه کرده‌اند. آن‌ها تأثیر برنامه‌های کاربردی ناشناس را بر مدل شناسایی و احراز هویت بررسی کرده‌اند. برنامه‌های کاربردی ناشناس تنها در مجموعه برنامه‌های کاربردی در زمان ارزیابی قرار گرفته‌اند. آن‌ها دریافتند که برنامه‌های کاربردی ناشناس منجر به افزایش تعداد منفی‌های کاذب می‌گردد.

آشیبانی و همکاران [۱۷] یک مدل احراز هویت پیوسته کاربر در شبکه‌های خانه هوشمند با استفاده از رویدادهای دسترسی کاربر به برنامه‌های کاربردی رایج مربوط به خانه هوشمند در طول روزهای مختلف هفته ارائه می‌کنند. شناسه کاربر، نام برنامه کاربردی، زمان و مدت دسترسی و روز هفته ازجمله ویژگی‌های استخراج‌شده هستند. نتایج ارزیابی انواع روش‌های طبقه‌بندی نشان داد الگوریتم جنگل تصادفی به معیار F1[15] بهتری نسبت به سایر روش‌ها دست یافت. معیار F1 ترکیبی از معیارهای صحت[16] و بازیابی[17] است. همچنین در مقاله دیگری از این نویسنده [۱۸]، یک روش احراز هویت پیوسته کاربر بر اساس الگوی دسترسی کاربر به برنامه‌های کاربردی تلفن همراه به‌منظور حفاظت از وسایل خانه هوشمند از دسترسی افراد غیرمجاز ارائه گردیده است.

در سال‌های اخیر ـ ۲۰۲۱ تا ۲۰۲۳ ـ پژوهش‌هایی در حوزه احراز هویت پیوسته با استفاده از روش‌های زیست‌سنجی رفتاری انجام گرفته است که در ادامه به برخی از آن‌ها اشاره می‌گردد. شایان ذکر است در مرور مقالات این سال‌ها پژوهش مرتبط به حوزه مقاله حاضر یعنی احراز هویت پیوسته بر اساس الگوی رفتاری استفاده از برنامه‌های کاربردی یافت نشد. از این‌رو به‌صورت مختصر به مطالعاتی که در حوزه دیگر روش‌های زیست‌سنجی رفتاری انجام گردیده است می‌پردازیم.

ژانگ و همکاران [۱۹] یک روش احراز هویت چندوجهی بر اساس الگوی تعاملی ایستا و پویا در تلفن همراه ارائه کرده‌اند. ویژگی‌های زیست‌سنجی رفتاری HMHP، که حرکت ترکیبی دست[18] (HM) و حالت نگه‌داشتن[19] (HP) است اساساً بر روی صفحه نمایش لمسی و شتاب‌دهنده ایجاد می‌شود و مدل تغییر حرکات ریز دست و الگوهای نگه‌داشتن ایجادشده در هر دو حالت پویا و ایستا را ثبت می‌کند. با ترکیب دو ویژگی HM و HP، ویژگی ترکیبی HMHP به دقت ۹۷ درصد و نرخ خطای برابر ۳.۷۹ درصد دست می‌یابد. با گذشت زمان و تغییر رفتار کاربر، میزان تشابه به الگوی اولیه، کاهش می‌یابد و منجر به کاهش دقت می‌گردد. از این‌رو مدل ساخته‌شده، تغییرات و عادات جدید را یاد می‌گیرد و هر زمان دقت عملکرد مدل به کمتر از حد آستانه برسد، نمایه رفتاری کاربر به‌صورت خودکار به‌روزرسانی می‌گردد.

در سال ۲۰۲۲، Dybczak و همکاران [۲۰] یک روش احراز هویت پیوسته زیست‌سنجی رفتاری با توانایی شناسایی کاربران بر اساس حرکات دست آن‌ها در زمان استفاده از تلفن هوشمند و با استفاده از حسگرهای داخلی تلفن هوشمند و APIهای عمومی ارائه کردند. از آن‌جایی که این نوع احراز هویت از نظر دقت قابل مقایسه با بهترین روش‌های احراز هویت ایستا نمی‌باشد، نباید به‌عنوان یک سازوکار امنیتی جایگزین معرفی گردد بلکه این روش به‌عنوان لایه دیگری از امنیت اضافه‌شده به سیستم احراز هویت ایستا و سنتی موجود درنظر گرفته شود.

ونگ و همکاران [۲۱] به جنبه کمترپرداخته‌شده در حوزه احراز هویت پیوسته توجه کرده‌اند. آن‌ها احراز هویت پیوسته در چند دستگاه را مورد بررسی قرار داده‌اند. در این روش، کاربران پس از دسترسی اولیه به یک دستگاه، امکان انتقال و دسترسی به دستگاه دیگر را پیدا می‌کنند و در طول این فرایند به‌صورت پیوسته احراز هویت می‌گردند. در صورتی که کاربر از یک وسیله مانند تلفن هوشمند استفاده کند، دیگر نیازی به احراز هویت صریح در زمان استفاده از وسایلی مانند تبلت یا ساعت هوشمند ندارد. در این پژوهش با زیرنظرگرفتن حرکات کاربر هنگام‌راه‌رفتن، از داده‌های حسگرهای جمع‌آوری‌شده از حسگر شتاب‌سنج و ژیروسکوپ در زمان حرکت دست و وضعیت و حالت دست در تلفن هوشمند و تبلت استفاده می‌شود. این داده‌های حرکتی زیست‌سنجی رفتاری در چند دستگاه به تصاویر خاکستری تبدیل می‌شود. تصاویر دامنه دو بعدی برای مشخص‌کردن ویژگی‌های اساسی سیگنال‌های حسگر بین دستگاه‌های مختلف ساخته شده و به‌عنوان ورودی به مدل شبکه عصبی که ترکیبی از شبکه عصبی پیچشی[20] و LSTM می‌باشد به‌منظور کلاس‌بندی داده می‌شود. با استفاده از این سازوکار، به میانگین دقت ۹۹.۸ درصد برای تلفن هوشمند و ۹۹.۲ درصد برای تبلت در حدود ۲.۳ ثانیه که بیانگر دقت و سرعت احراز هویت پیوسته می‌باشد دست می‌یابند.

نویسندگان [۲۲, ۲۳] مقالات مروری مختلفی در موضوع احراز هویت پیوسته بر اساس زیست‌سنجی رفتاری در تلفن‌های هوشمند ارائه کرده‌اند. در [۲۲] به‌طور خلاصه در مورد مجموعه داده‌های عمومی در دسترس، انواع حملات و تکنیک‌های دفاعی در زیست‌سنجی رفتاری بحث شده است. همچنین در [۲۳] PK Rayani و همکاران با بررسی مجموعه داده‌های عمومی دردسترس، انواع روش‌های احراز هویت تک‌وجهی و چندوجهی، روش پیشنهادی مقالات مرتبط به این حوزه و گزارش عملکرد آن‌ها، مرور جامعی بر مقالات احراز هویت پیوسته بر اساس زیست‌سنجی رفتاری ارائه کرده‌اند. همچنین در نهایت به بررسی انواع حملات بر روی روش‌های زیست‌سنجی رفتاری پرداخته‌اند.

با بازخوانی مقالات حوزه احراز هویت پیوسته با استفاده از الگوی رفتاری استفاده از برنامه‌های کاربردی تلفن هوشمند، این نتیجه حاصل می‌گردد که در بیشتر پژوهش‌های پیشین، راهکاری برای حل معضل اساسی این روش، یعنی متغیر بودن الگوی رفتاری کاربر با گذشت زمان ارائه نگردیده است. محبوب و همکاران [۱۱] با درنظرگرفتن نماد U برای برنامه‌های دیده‌نشده در زمان ارزیابی، استفاده کاربر از برنامه‌های کاربردی جدید را مدیریت کرده‌اند. با گذشت زمان و تغییرات بیشتر در مجموعه برنامه‌های کاربردی مورداستفاده کاربر، این روش چاره‌ساز نیست و منجر به کاهش دقت سیستم احراز هویت می‌گردد. همچنین همان‌طور که پیش‌تر اشاره شد در [۱۵] ناپایداری رفتار کاربران با گذشت زمان بررسی گردیده است. کایاجیک و همکاران دریافتند بازآموزی مدل رفتاری کاربر از انجام‌نشدن آن یا آموزش یک نمایه جدید، بهتر خواهد بود. در روش ارائه‌شده در این مقاله، رفتار کاربر با برنامه‌های کاربردی مثل نوع برنامه کاربردی مورداستفاده یا مدت‌زمان استفاده از آن در زمان آموزش یا بازآموزی مدل درنظر گرفته نمی‌شود.

در روش پیشنهادی این مقاله، داده‌های مربوط به استفاده کاربر از برنامه‌های کاربردی مانند نوع برنامه مورداستفاده، زمان و ترتیب استفاده و مدت‌زمان استفاده از آن‌ها برای ساخت الگوی رفتاری کاربر مورد استفاده قرار می‌گیرد. با بهره‌گیری از دسته‌بندی‌های معنایی برنامه‌های کاربردی مورداستفاده کاربر، در صورت استفاده کاربر از برنامه‌های جدید، هر برنامه کاربردی در دسته مربوطه جای می‌گیرد و تغییری در مجموعه برنامه‌های مورداستفاده وی رخ نخواهد داد. همچنین در صورت تغییر در ترتیب، زمان استفاده یا مدت‌زمان استفاده از برنامه‌های کاربردی، با استفاده از داده‌های اخیر کاربر و بهره‌مندی از راهکار فضای معنایی مشترک، مدل رفتاری وی را بازآموزی می‌کنیم.

۳- روش پیشنهادی

در بخش گذشته، با بررسی مقالات حوزه احراز هویت پیوسته با استفاده از الگوی رفتاری استفاده از برنامه‌های کاربردی تلفن هوشمند، متوجه شدیم که متغیربودن رفتار کاربر در طول زمان که یکی از معضلات اساسی این روش می‌باشد، تاکنون بررسی نشده است.

مسئله شناسایی پیوسته کاربر گوشی هوشمند، یک مسئله طبقه‌بندی با نظارت به‌حساب می‌آید. در روش طبقه‌بندی با نظارت، فضای ویژگی، توزیع نمونه‌ها و فضای برچسب در مجموعه داده‌های آموزش و ارزیابی باید یکسان باشند [۲۴]. فضای برچسب که در این مسئله معادل مالک اصلی گوشی هوشمند است، در زمان آموزش

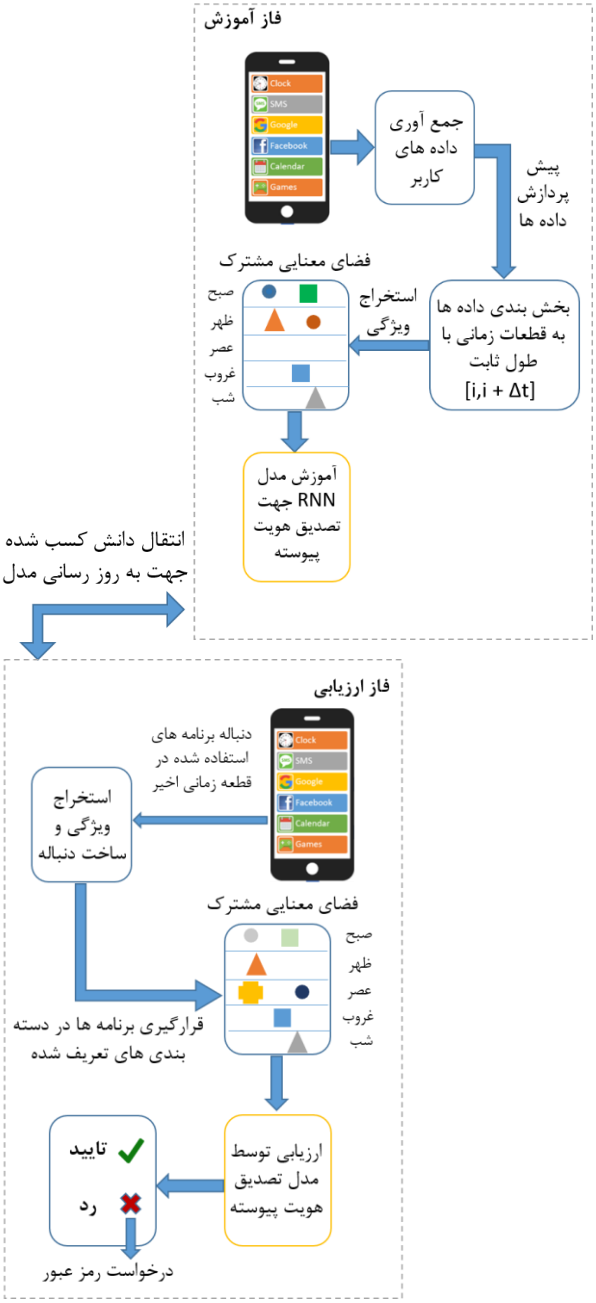
و آزمون، یکسان خواهد بود. در حالی‌که در این مسئله، به دلیل امکان جایگزین‌شدن برنامه‌های کاربردی جدید با برنامه‌های کاربردی قدیمی در مجموعه برنامه‌های کاربردی مورداستفاده کاربر با گذشت زمان یا نصب برنامه کاربردی جدید، فضای ویژگی نمونه‌های مقصد، متفاوت خواهند شد.

در این مقاله، روشی نوین برای حل مسئله شناسایی پیوسته در تلفن همراه بر اساس رویکرد یادگیری انتقالی ارائه شده است. در روش پیشنهادی این پژوهش، ایده تعریف فضای معنایی مشترک با هدف یکسان‌سازی فضای ویژگی مبدأ و مقصد و فراهم‌شدن امکان استفاده از روش‌های طبقه‌بندی با نظارت مطرح می‌گردد. با بهره‌مندی از فضای معنایی مشترک تعریف‌شده، مجموعه برنامه‌های کاربردی مورداستفاده کاربر در زمان آموزش و ارزیابی در دسته‌بندی‌های معنایی مشترک قرار می‌گیرند و امکان استفاده از مبحث یادگیری انتقالی فراهم می‌گردد. در این روش از ویژگی معنایی برنامه‌های کاربردی برای دسته‌بندی آن‌ها در دسته‌های توصیفی مرتبط استفاده می‌شود. در بخش ۳-۱ توضیحات کاملی از شیوه ساخت فضای معنایی مشترک ارائه می‌گردد.

معماری سیستم پیشنهادی در شکل ۱ به نمایش درآمده است. به‌منظور حفظ حریم خصوصی کاربر در زمان جمع‌آوری اطلاعات، اطلاعات محدودی در زمان استفاده از برنامه‌های کاربردی همانند عنوان برنامه کاربردی و زمان ورود و خروج از آن از کاربر جمع‌آوری می‌گردد. به دلیل محدودبودن داده‌های دریافتی، نیاز به کسب اجازه دسترسی به اطلاعات استفاده کاربر از برنامه‌های کاربردی نمی‌باشد. سپس عملیات پیش‌پردازش بر روی داده‌های خام جمع‌آوری‌شده آغاز می‌گردد. ابتدا داده‌های خام با استفاده از رویکرد پنجره کشویی[۲۱] به قطعه‌های زمانی با طول ثابت بخش‌بندی می‌گردند. سپس برنامه‌های کاربردی با توجه به عناوین و دسته‌بندی‌های پیش‌فرض، در دسته‌بندی‌های معنایی که از پیش تعریف کرده‌ایم جای می‌گیرند. اطلاعات به‌دست‌آمده از مرحله قبل، وارد فاز استخراج ویژگی برای استخراج ویژگی‌های زمانی و اطلاعات مربوط به فعالیت برنامه‌های کاربردی می‌گردند. در پایان این مرحله، دنباله‌های زمانی ساخته‌شده حاوی اطلاعات رفتاری کاربر به شبکه برای ساخت نمایه رفتاری وی داده می‌شوند. پس از ساخت مدل رفتاری کاربر، شناسایی پیوسته در تلفن همراه وی آغاز می‌گردد. به همین منظور، در قطعه‌های زمانی مشخص، یک دنباله از رفتارهای کاربر با برنامه‌های کاربردی، مشابه روشی که در زمان آموزش مدل اتخاذ گردید، ایجاد می‌شود و با مدل ساخته‌شده از کاربر اصلی مقایسه می‌گردد. درصورتی‌که برچسب کاربر نامعتبر برای دنباله پیش‌بینی گردد، سیستم با تقاضای تصدیق هویت صریح، مانند درخواست رمز عبور، کاربر را رد می‌کند. اگر دنباله متعلق به کلاس کاربر اصلی پیش‌بینی گردد؛ کاربر، شناسایی و تأیید می‌گردد.

۱-۳- فضای معنایی مشترک

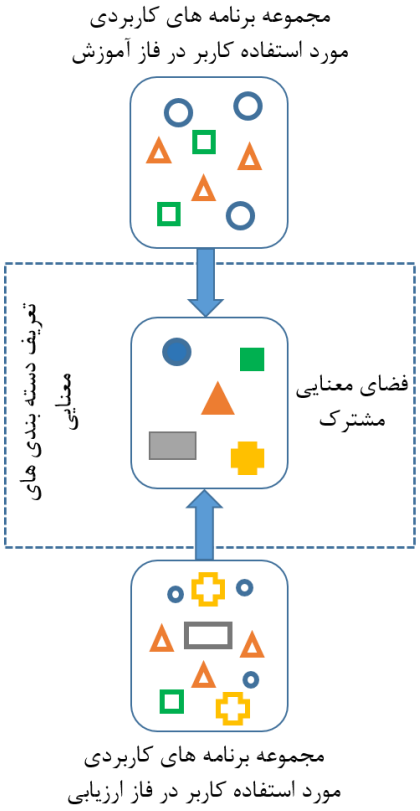
همان‌طور که در بخش پیشین بیان گردید، در مسئله شناسایی پیوسته، با گذشت زمان، به علت تغییر رفتار احتمالی کاربر و نصب یا استفاده از برنامه‌های کاربردی جدید که در زمان آموزش و ساخت مدل اولیه دیده نشده بودند، فضای ویژگی نمونه‌های استخراج‌شده ثانویه نیز متفاوت می‌گردد. در روش پیشنهادی که در این مقاله ارائه کرده‌ایم، از ویژگی معنایی و نوع کاربری برنامه‌های کاربردی برای ساخت فضای معنایی مشترک استفاده می‌شود. به عبارتی نمونه‌های استخراج‌شده از الگوی رفتاری اولیه و ثانویه کاربر در یک فضای ویژگی مشترک قرار می‌گیرند. ویژگی معنایی، به نوعی از دسته‌بندی توصیفی از برنامه‌های کاربردی اطلاق می‌گردد. برای مثال، برنامه‌های کاربردی کروم، فایرفاکس و سافاری را می‌توان در دسته مرورگرها قرار داد. با این امکان، درصورتی‌که کاربر یک یا چند برنامه کاربردی از مجموعه برنامه‌های کاربردی مورداستفاده خود را با برنامه‌های مرتبط و مشابه جایگزین کند، در عملکرد مدل احراز هویت، تغییری ایجاد نخواهد شد و مدل اولیه بدون نیاز به بازآموزی یا ساخت مجدد، توانایی شناسایی و احراز هویت کاربر را خواهد داشت. این امر موجب افزایش قابلیت استفاده و امنیت در سیستم پیشنهادی می‌گردد. همچنین در زمان ارزیابی و احراز هویت پیوسته در صورت مواجهه مدل با یک برنامه کاربردی دیده‌نشده، این برنامه کاربردی در دسته‌بندی معنایی مرتبط جای می‌گیرد و مسئله ناشناس‌بودن برنامه‌های کاربردی برای مدل آموزش داده شده حل خواهد شد.



شکل ۱. معماری سیستم پیشنهادی

شیوه ساخت فضای معنایی مشترک در شکل ۲ به تصویر درآمده است. همان‌طور که پیش از این اشاره گردید، از مزایای دیگر استفاده از راهکار فضای معنایی مشترک، فراهم‌گشتن امکان انتقال دانش از مدل اولیه به مدل جدید به‌منظور به‌روزرسانی مدل اولیه می‌باشد. به عبارتی با یکسان‌سازی فضای ویژگی نمونه‌های استخراج‌شده از داده‌های مبدأ و مقصد، راهکار فضای معنایی مشترک مانند یک زبان مشترک بین مدل پیشین و مدل جدید عمل می‌کند. در این صورت می‌توانیم با بهره‌گیری از رویکرد یادگیری انتقالی، دانش رفتاری آموخته‌شده از الگوی رفتاری پیشین را برای بازآموزی و به‌روزرسانی نمایه رفتاری کاربر به مدل رفتاری جدید وی

انتقال دهیم. در این صورت، نیاز به جمع‌آوری و برچسب‌گذاری مجدد داده‌ها و ساخت مدل از ابتدا با داده‌های محدود مرتفع می‌گردد.



شکل ۲. فضای معنایی مشترک

۲-۳- استخراج ویژگی

در بحث شناسایی الگوی رفتاری کاربر، عامل زمان، اهمیت ویژه‌ای در ساخت الگوی رفتاری منحصربه‌فرد کاربر و احراز هویت وی را دارد. زمان استفاده کاربر از برنامه‌های کاربردی در روز و ترتیب استفاده آن‌ها در ساخت الگوی رفتاری منحصربه‌فرد کاربر مؤثر خواهد بود. در همین راستا از استراتژی پنجره کشویی برای تقسیم روز به قطعه‌های زمانی با طول ثابت و بدون هم‌پوشانی استفاده می‌گردد. استفاده از این استراتژی یک طراحی کلیدی در این پژوهش به‌حساب می‌آید زیرا که یک رفتار غیرطبیعی با تلفن هوشمند در یک لحظه رخ نمی‌دهد بلکه در یک دوره زمانی هرچند کوتاه اتفاق می‌افتد [۲۵]. سپس نمونه‌ها از داده‌های خام به‌دست‌آمده از این قطعه‌های زمانی استخراج می‌گردد. هر قطعه زمانی حاوی: ۱ـ داده‌های زمانی استفاده از برنامه‌های کاربردی و ۲ـ داده‌های مربوط به فعالیت در برنامه‌های کاربردی می‌باشد و ویژگی‌ها بر اساس این دو دسته اطلاعات استخراج می‌گردند.

۱-۲-۳- اطلاعات زمانی استفاده از برنامه‌های کاربردی

کاربران گوشی‌های هوشمند به‌طور معمول یا طبق سبک زندگی خود، در هر بخش از روز برنامه‌های کاربردی متفاوتی استفاده می‌کنند. برای مثال از برنامه Clock معمولاً در شب استفاده می‌گردد. شناسایی این الگوی زمانی به شناخت بهتر الگوی رفتاری کاربر کمک می‌کند. از همین رو، زمان در یک شبانه‌روز به ۵ بخش تقسیم خواهد گشت: صبح، ظهر، عصر، غروب و شب. این نوع تقسیم‌بندی زمان در روز، الگوی رفتاری کاربر را انعطاف‌پذیرتر می‌کند و اجازه ایجاد تغییرات جزئی در الگوی رفتاری را به کاربر می‌دهد. برای نمونه اگر کاربر از یک برنامه کاربردی در ساعت ۹:۳۰ به‌جای ساعت ۸:۳۰ برخلاف عادت همیشه استفاده کند، این تغییر رفتاری در الگوی رفتاری او تغییری ایجاد نمی‌کند.

۲-۲-۳- اطلاعات مربوط به فعالیت در برنامه‌های کاربردی

همان‌طور که در بخش قبل ذکر گردید، با گذشت زمان، به علت تغییر رفتار احتمالی کاربر و استفاده از برنامه‌های کاربردی جدید که در زمان آموزش و ساخت مدل اولیه دیده نشده بودند، فضای ویژگی نمونه‌های استخراج‌شده ثانویه نیز متفاوت می‌گردد. در روش پیشنهادی ارائه‌شده، از ویژگی معنایی برنامه‌های کاربردی برای ساخت فضای معنایی مشترک استفاده می‌شود. به همین منظور دسته‌بندی‌های توصیفی از برنامه‌های کاربردی مورداستفاده کاربر ایجاد می‌گردد. سپس هریک از برنامه‌های کاربردی مورداستفاده کاربر در دسته‌بندی‌های معنایی مرتبط جای می‌گیرند.

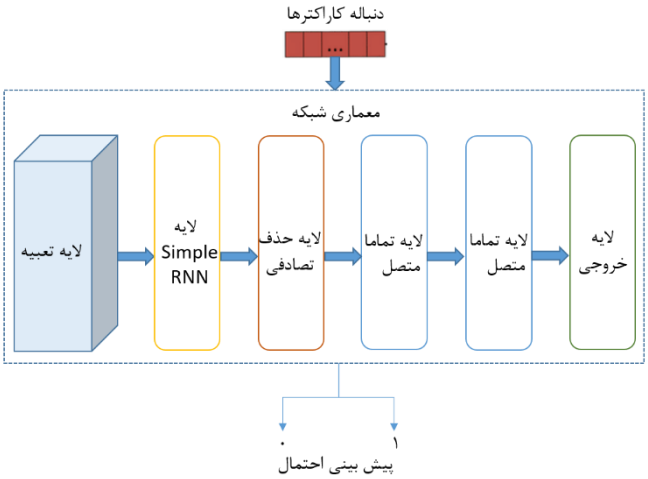
پس از استخراج اطلاعات زمانی و اطلاعات مربوط به فعالیت در برنامه‌های کاربردی و استخراج اطلاعات کاربردی دیگر از داده‌های خام مانند مدت‌زمان استفاده از یک برنامه کاربردی و ترتیب استفاده از آن‌ها توسط کاربر در یک بازه زمانی با طول ثابت، به مجموعه داده آموزشی کامل‌تری دست خواهیم یافت. در حال حاضر به کمک اطلاعات به‌دست‌آمده و ویژگی‌های استخراج‌شده می‌توان به سؤالات زیر برای ساخت الگوی رفتاری کاربر و شناسایی او پاسخ داد:

- کاربر در طول یک بازه زمانی مشخص از کدام دسته از برنامه‌های کاربردی استفاده کرده است؟
- کاربر چه مدت از یک برنامه استفاده کرده است؟
- کاربر در چه بخشی از روز از یک برنامه استفاده کرده است؟
- ترتیب استفاده از برنامه‌های کاربردی در یک بازه زمانی مشخص به چه صورت است؟

۳-۳- شناسایی و احراز هویت پیوسته

در مسئله شناخت الگوی رفتاری کاربر در زمان استفاده از برنامه‌های کاربردی در طول روز، اطلاعات زمانی استفاده از برنامه‌های کاربردی از اهمیت بالایی برخوردار بوده و به شناخت الگوی رفتاری منحصربه‌فرد کاربر در طی روز و شناسایی رفتار غیرطبیعی کمک می‌کند. در مسائل مرتبط با سری‌های زمانی، داده‌های ترتیبی یا دنباله‌های متنی، شبکه عصبی بازگشتی[۲۲] (RNN) یکی از کاربردی‌ترین ابزار این حوزه به‌شمار می‌آید [۲۶]. بنابراین در پژوهش حاضر، با استفاده از روش طبقه‌بندی شبکه عصبی بازگشتی، دنباله‌های زمانی استخراج‌شده وارد مدل می‌گردد و نمایه رفتاری کاربر آموزش داده می‌شود.

همان‌طور که در شکل ۳ قابل‌مشاهده است، در زمان آموزش، ابتدا عملیات پیش‌پردازش روی دنباله‌های ترتیبی از کاراکترها انجام می‌گیرد. این عملیات شامل تبدیل توکن به عدد[۲۳] و لایه‌گذاری با صفر[۲۴] می‌باشد. دنباله‌های پیش‌پردازش‌شده برای آموزش وارد شبکه می‌شوند. به‌منظور پیاده‌سازی این شبکه، از لایه‌های تعریف‌شده در کتابخانه کراس استفاده کرده‌ایم. در این معماری ابتدا لایه تعبیه[۲۵]، یک بردار با مقادیر عددی به هر کاراکتر اختصاص می‌دهد. خروجی این لایه به لایه Simple RNN به‌عنوان ورودی معرفی می‌گردد. پس از این لایه، یک لایه حذف تصادفی[۲۶] در مدل قرار داده می‌شود که وظیفه آن جلوگیری از بیش‌برازش[۲۷] مدل می‌باشد. در ادامه دو لایه تماماً متصل[۲۸] قرار می‌گیرند که در لایه دوم، تابع فعال‌ساز[۲۹] Sigmoid با ایجاد خروجی بین ۰ و ۱، کلاس دنباله ورودی را پیش‌بینی می‌کند.



شکل ۳. معماری شبکه مدل پیشنهادی

۴- نتایج ارزیابی

در این بخش به ارزیابی روش پیشنهادی خواهیم پرداخت. بدین ترتیب ابتدا مجموعه داده مورداستفاده برای پیاده‌سازی و انجام آزمایش‌ها معرفی می‌گردد. سپس با پیاده‌سازی روش پیشنهادی به ارزیابی آن خواهیم پرداخت.

۴-۱- مجموعه داده

احراز هویت کاربران بر اساس نحوه استفاده آن‌ها از برنامه‌های کاربردی تلفن هوشمند نیازمند یک مجموعه داده با اطلاعات کافی و مرتبط می‌باشد. با وجود رویکردهای تحقیقاتی متنوع در حوزه داده‌های مربوط به استفاده از برنامه‌های کاربردی، مجموعه داده‌های عمومی مرتبط با این حوزه بسیار کمیاب هستند [۲۳]. همچنین مجموعه داده‌های موجود گاهی دارای تعداد محدودی برنامه کاربردی هستند یا بازنمایی درستی از رفتار واقعی کاربران با برنامه‌های کاربردی را ندارند و صرفاً داده‌ها توسط یک ناظر یا طبق یک دستورالعمل مشخص تولید شده‌اند [۱۱].

در این ارزیابی، از مجموعه داده بسیار بزرگ shared_records [۲۷] که شامل ۶ میلیون داده بایگانی[۳۰] شده است استفاده گردید و اطلاعات دسترسی کاربران به مجموعه برنامه‌های کاربردی فیس‌بوک، واتس‌اپ، اینستاگرام، یوتیوب، مرورگرها (کروم، فایرفاکس، اپرا، اپرا مینی و مرورگر پیشفرض اندروید)، waze و IBM Notes Traveler را دارا می‌باشد. اطلاعات این مجموعه داده به‌صورت خلاصه در جدول ۱ بیان گردیده است.

جدول ۱. اطلاعات مجموعه‌داده shared_records

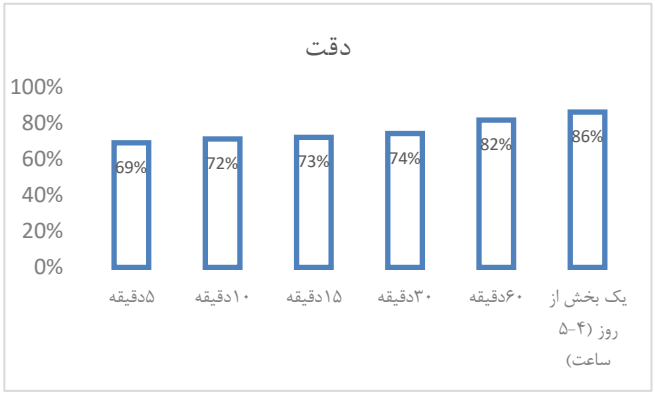
مجموعه‌داده	تعداد کاربران	تعداد لاگ‌های بایگانی شده	اطلاعات جمع‌آوری‌شده
shared_records	۵۳۴۲ نفر	۶.۸۰۵.۹۳۲	آی دی کاربر، تاریخ و ساعت شروع و پایان دسترسی به اپ، مجموع بایت‌های دانلود/آپلود شده

۴-۲- آماده‌سازی داده‌ها و استخراج ویژگی

به‌منظور آغاز فاز پیاده‌سازی، ابتدا لازم است عملیات پیش‌پردازش بر روی داده‌های خام مجموعه‌داده shared_records با هدف ارتقای کیفیت داده‌ها اعتماد انجام پذیرد. در ابتدا داده‌های خام با استفاده از استراتژی پنجره کشویی به قطعه‌های زمانی با طول ثابت بخش‌بندی می‌گردند. طول قطعه‌های زمانی (Δt) یک ابر پارامتر بوده و انتخاب عدد مناسب برای پارامتر Δt از اهمیت بالایی برخوردار است. برای انتخاب مناسب‌ترین Δt برای هر قطعه زمانی، مقادیر مختلفی برای این پارامتر درنظر گرفته شد که نتایج به‌دست‌آمده در شکل ۴ به نمایش درآمده است.

باتوجه به نتایج به‌دست‌آمده این نتیجه را دریافتیم که هرچه طول زمانی قطعه‌ها بیشتر باشد، مدل آموزش داده‌شده با دقت بالاتری کاربر را شناسایی می‌کند. از سوی دیگر تشخیص زودهنگام سوءاستفاده افراد سودجو در مسئله شناسایی پیوسته مطرح می‌باشد. از این‌رو باید بین طول زمان و دقت مدل، تعادل را برقرار کرد. ازاین‌رو مدت‌زمان ۱۰ دقیقه را می‌توان به‌عنوان مناسب‌ترین مقدار از لحاظ کوتاه‌بودن مدت‌زمان تشخیص و همچنین دقت مناسب مدل درنظر گرفت. سپس نمونه‌ها از قطعه‌های زمانی ایجادشده استخراج می‌گردند.

شایان ذکر است محدودیت اصلی این مسئله به‌حداقل‌رساندن زمان شناسایی کاربر می‌باشد. بنابراین درصورتی‌که به دنبال کاهش زمان شناسایی نباشیم، می‌توان به نتایج درخور توجهی تا حدود دقت ۸۷ درصد دست‌یافت.



شکل ۴. عملکرد مدل به‌ازای Δtهای متفاوت

در روش طبقه‌بندی شبکه عصبی بازگشتی، ورودی شبکه دنباله‌ای از کاراکترها یا کلمات دارای ارتباط زمانی هستند. ازاین‌رو در بخش پیاده‌سازی پژوهش از دنباله‌ای از کاراکترهای انگلیسی به‌منظور بازنمایی ویژگی‌ها مانند نوع برنامه کاربردی، ترتیب یا مدت‌زمان استفاده از برنامه‌های کاربردی و ... استفاده می‌گردد. همچنین با اختصاص کاراکتر به هر دسته برنامه کاربردی بر اساس زمان استفاده از آن برنامه کاربردی در روز، مشخص می‌گردد که هر قطعه زمانی $[i, i + \Delta t]$ در کدام بخش از روز رخداده است. به‌عنوان نمونه درصورتی‌که کاربر از برنامه مرورگر در بازه زمانی صبح، یعنی ۴:۰۰ تا ۱۱:۰۰ استفاده کند کاراکتر d و در صورتی که در بازه زمانی غروب، یعنی ۱۸:۰۰ تا ۲۲:۰۰ استفاده کند، کاراکتر s به آن تعلق می‌گیرد. در بخش اطلاعات مربوط به فعالیت در برنامه‌های کاربردی، از ویژگی معنایی برنامه‌های کاربردی برای ساخت فضای معنایی مشترک استفاده می‌گردد. از این‌رو با تعریف پنج دسته‌بندی توصیفی از برنامه‌های کاربردی، ۱۱ برنامه کاربردی موجود در مجموعه‌داده در دسته‌بندی‌های زیر قرار می‌گیرند: شبکه اجتماعی، سرگرمی، پیام‌رسان، مرورگر و مسیریاب. شیوه تخصیص کاراکتر به برنامه‌های کاربردی در جدول ۲ قابل‌مشاهده می‌باشد. همچنین به‌منظور نمایش مدت‌زمان استفاده کاربر از برنامه کاربردی به‌ازای هر یک دقیقه استفاده، یک کاراکتر مربوطه به دنباله اضافه می‌شود. در زمان درج کاراکترها، ترتیب استفاده از برنامه‌های کاربردی نیز موردتوجه قرار خواهد گرفت.

جدول ۲. شیوه اختصاص‌دهی کاراکترها به برنامه‌های کاربردی

دسته برنامه کاربردی / بخش روز	صبح	ظهر	عصر	غروب	شب
شبکه اجتماعی	a	f	k	p	u
سرگرمی	b	g	l	q	v
پیام‌رسان	c	h	m	r	w
مرورگر	d	I	n	s	x
مسیریاب	e	J	o	t	y

در این پژوهش مجموعاً چهار ویژگی از مجموعه‌داده shared_records استخراج گردیده است: دسته برنامه کاربردی، زمان استفاده از برنامه کاربردی در روز، مدت‌زمان استفاده از برنامه کاربردی و ترتیب استفاده از برنامه‌های کاربردی.
الگوریتم ۱ شبه کد فرایند ساخت مدل رفتاری کاربر بر اساس استفاده از برنامه‌های کاربردی را نمایش می‌دهد.

Algorithm 1: App-Based User Verification Model Building

Input: Dataset shared-records

Output: user's classification model

procedure()

input ← **D**=[1, 2, . . . , n] Read data from dataset - n is the number of samples

divide the time in one day into 5 periods: morning, noon, afternoon, evening and night

define semantic descriptions for applications in Dataset D to Categorize them

Label all sequences using one-vs-all classification approach

Extract Features - Δt = 10 minutes

Assign character to each app in a sequence based on category and time of day (According to Table 2) - Append a Character per minute in a sequence

Tokenize and Pad sequences

shuffle data and split data into train 60% , validation 20% , test 20%

build the RNN model using train data and choose optimized hyperparameters using validation data

Predict classes of test data

Invalid User ← Predict class: lable 0 / valid User ← Predict class: lable 1

end procedure

۳-۴- تنظیمات آزمایش[۳۱]

به‌منظور انتخاب روش طبقه‌بندی از بین روش دو کلاسه یا چند کلاسه، آزمایش‌هایی انجام گردید و روش دو کلاسه دقت بالاتری دریافت کرد. درصورتی‌که نیاز به احراز هویت چند کاربر باشد، برای مثال تلفن همراه به‌صورت اشتراکی بین والد و فرزند مورد استفاده قرار گیرد، می‌توان از داده‌های رفتاری کاربر دوم برای شناسایی و ساخت الگوی رفتاری ایشان استفاده کرد. در نهایت با استفاده از طبقه‌بندی دو کلاسه و تکنیک یکی در مقابل همه، کاربر یا کاربران اصلی به‌عنوان کلاس ۱ و سایر کاربران کلاس ۰ درنظر گرفته می‌شوند. شایان ذکر است پژوهش حاضر در صدد شناسایی الگوی رفتاری یک کاربر به‌عنوان کاربر اصلی می‌باشد. به‌منظور آموزش مدل و ارزیابی آن، از روش hold-out استفاده می‌گردد. باتوجه به اینکه همه داده‌های این مجموعه‌داده به‌ترتیب کلاس‌ها مرتب شده‌اند، ابتدا داده‌ها را بُر زده[۳۲] و سپس به سه بخش آموزش، اعتبارسنجی و ارزیابی با نسبت ۶۰، ۲۰، ۲۰ تقسیم می‌گردند.

تنظیم ابرپارامترها نقش مهمی در عملکرد مدل‌های یادگیری عمیق دارد. پارامترها و ابرپارامترهای مورداستفاده در هر لایه در جدول ۳ آورده شده‌اند.

جدول ۳. تنظیمات لایه‌ها در رویکرد پیشنهادی

مشخصه شبکه	مقدار پارامتر / ابرپارامتر
ابعاد لایه تعبیه	۶۴
لایه RNN	تعداد نورون‌ها: ۶۴
ابعاد لایه تماماً متصل	لایه اول: ۳۲
	لایه دوم: ۱
نرخ لایه حذف تصادفی	۰.۲
تابع فعال‌ساز	RELU[33]
	Sigmoid
تابع بهینه‌ساز	Adam[34]
تابع زیان[۳۵]	اختلاف آنتروپی دودویی[۳۶]
تعداد تکرارهای شبکه[۳۷]	۱۵
تعداد نمونه‌های هر دسته[۳۸]	۱۲۸

معیار ارزیابی منتخب به دلیل بررسی و ارزیابی نتایج در این پژوهش نظر به توزین متعادل کلاس‌ها، دقت (Accuracy) و نرخ خطای برابر[۳۹] (EER) می‌باشد.

۴-۴- ارزیابی و تحلیل نتایج

پس از آموزش مدل رفتاری کاربر و ارزیابی آن، سیستم احراز هویت پیوسته به دقت ۷۲ درصد و نرخ خطای برابر ۲۹ درصد دست یافت. همان‌طور که پیش از این ذکر گردید، در زیست‌سنجی‌های رفتاری ازجمله روش الگوی رفتاری استفاده از برنامه‌های کاربردی، مهم‌ترین معضل، تغییر در مجموعه برنامه‌های کاربردی مورداستفاده کاربر مانند اضافه‌شدن، کم‌شدن یا جایگزین‌شدن برنامه‌های کاربردی جدید با برنامه‌های کاربردی قدیمی یا تغییر الگوی رفتاری وی با گذشت زمان می‌باشد. همان‌گونه که در مرور کارهای مرتبط اشاره گردید، مطالعات انجام‌شده در این حوزه، معضل تغییر رفتار کاربر را بررسی نکرده‌اند یا راهکار کاربردی با این هدف ارائه ننموده‌اند. به عنوان نمونه در مقاله [۱۱] نماد U برای برنامه‌های دیده‌نشده در زمان ارزیابی معرفی می‌گردد. در صورت استفاده کاربر از مجموعه برنامه‌های کاربردی جدید با گذشت زمان، این راهکار چاره‌ساز نمی‌باشد و منجر به کاهش دقت سیستم شناسایی و احراز هویت می‌گردد. در ادامه با طرح سه سناریو به بررسی

نتایج عددی راهکار پیشنهادی، مقایسه آن با روش دسته‌بندی برنامه‌های کاربردی بر اساس نام و عنوان و همچنین با روش ارائه‌شده در مقاله [۱۱] می‌پردازیم.

در ابتدا یک مدل رفتاری بر اساس دسته‌بندی برنامه‌های کاربردی مبتنی بر نام و عنوان آموزش می‌دهیم. به همین منظور، با استفاده از راهکار اختصاص کاراکتر به هر برنامه بر اساس نام و عنوان آن‌ها، برای هریک از ۱۱ برنامه کاربردی موجود در این مجموعه‌ی داده، یک کاراکتر از مجموعه حروف زبان انگلیسی و کاراکترهای قابل‌چاپ اسکی[۴۰] با توجه به زمان استفاده از برنامه کاربردی در روز درنظر گرفته می‌شود. سپس مدل موردنظر، آموزش داده می‌شود. با گذشت زمان و تغییر در مجموعه برنامه‌های کاربردی مورداستفاده‌ی کاربر، عملکرد مدل احراز هویت مبتنی بر دسته‌بندی معنایی برنامه‌های کاربردی و مدل احراز هویت مبتنی بر دسته‌بندی براساس عنوان برنامه‌های کاربردی در سناریوهای جایگزینی، اضافه یا حذف یک برنامه کاربردی ارزیابی می‌گردد. نتایج به‌دست‌آمده، در جدول ۴ قابل مشاهده می‌باشد.

در اولین سناریو، در زمان ارزیابی، یک برنامه کاربردی جدید جایگزین یک برنامه کاربردی پرتکرار در مجموعه داده ارزیابی می‌گردد. برنامه جدید مورداستفاده کاربر از دسته‌بندی مرتبط و مشابه با برنامه کاربردی قبلی و بدون تغییر در الگوی استفاده انتخاب می‌گردد. همان‌گونه که انتظار می‌رفت، سیستم احراز هویت مبتنی بر روش دسته‌بندی معنایی برنامه‌های کاربردی در مواجهه با رفتار جدید کاربر مانند استفاده وی از یک برنامه کاربردی دیده‌نشده و مرتبط با دسته برنامه اول، این برنامه را در دسته‌بندی معنایی مرتبط جای داده و بدون تغییر در عملکرد شناسایی، با دقت حدود ۷۲ درصد کاربر را شناسایی می‌کند. در مدل دوم، استفاده کاربر از یک برنامه جدید و مرتبط را به‌عنوان یک رفتار جدید تلقی می‌کند و حدود ۷ درصد کاهش دقت را نسبت به مدل اول در پی دارد.

سناریو دوم و سوم به بررسی تأثیر حذف یا اضافه یک برنامه کاربردی بر عملکرد دو مدل موردمقایسه می‌پردازد. به همین منظور، ابتدا یک برنامه کاربردی کم‌تکرار به مجموعه داده ارزیابی اضافه می‌گردد. همچنین در مقایسه‌ای دیگر یک برنامه کاربردی پرتکرار از مجموعه داده ارزیابی حذف می‌گردد. نتایج هر دو مقایسه نشان می‌دهد روش پیشنهادی به‌طور قابل‌توجهی از مدل احراز هویت مبتنی بر دسته‌بندی برنامه‌ها بر اساس عنوان برنامه کاربردی بهتر عمل می‌کند.

جدول ۴. مقایسه نتایج بر اساس روش و سناریو انتخابی (Δt = ۱۰ دقیقه)

سناریو / روش	دسته‌بندی بر اساس فضای معنایی مشترک		دسته‌بندی بر اساس عنوان برنامه‌ی کاربردی		روش برگرفته از مقاله [۱۱]	
	دقت	EER	دقت	EER	دقت	EER
جایگزین‌شدن یک برنامه	۷۲٪	۲۹٪	۶۵.۷٪	۳۲٪		
اضافه‌شدن یک برنامه	۶۹٪	۳۲٪	۶۷٪	۳۴٪	۶۸.۵٪	۳۵٪
کم‌شدن یک برنامه	۶۶.۵٪	۳۱٪	۶۳.۵٪	۳۳٪		

نتایج به‌دست‌آمده با روشی که در مقاله [۱۱] ارائه گردیده است نیز مقایسه می‌شود. مشابه راهکار ارائه‌شده در [۱۱]، یک سیستم احراز هویت مبتنی بر نام و عنوان برنامه‌های کاربردی آموزش داده می‌شود. به‌منظور رسیدگی به برنامه‌های کاربردی ناشناس که ممکن است در زمان ارزیابی دیده شوند، یک برنامه کاربردی با عنوان U نیز درنظر گرفته می‌شود. در زمان آموزش، به‌منظور اختصاص شاخص[۴۱] به برنامه کاربردی U، احتمال بسیار پایین برای این برنامه کاربردی درنظر گرفته می‌شود. در زمان ارزیابی، داده‌های جدید مربوط به استفاده از یک برنامه کاربردی جدید تحت عنوان برنامه کاربردی U توسط کاربر، به مجموعه داده ارزیابی اضافه می‌گردد. نتایج این روش نشان می‌دهد سیستم احراز هویت در مواجهه با داده‌های رفتاری مرتبط با یک برنامه کاربردی جدید حدود ۳ درصد کاهش دقت را در پی دارد. با افزوده‌شدن برنامه‌های کاربردی جدید به مجموعه برنامه‌های کاربردی قبلی کاربر، شاهد کاهش بیشتری در دقت سیستم احراز هویت پیوسته خواهیم بود.

۵- نتیجه‌گیری

در این مقاله از روش شناسایی بر اساس الگوی رفتاری استفاده از برنامه‌های کاربردی تلفن همراه به‌منظور شناسایی و احراز هویت پیوسته کاربران تلفن‌های هوشمند در راستای بهبود امنیت اطلاعات ذخیره‌شده در تلفن همراه کاربران استفاده گردید. باتوجه به طبیعت روش‌های زیست‌سنجی رفتاری، ممکن است رفتار کاربران و مجموعه برنامه‌های کاربردی مورداستفاده به‌تدریج و تحت شرایط محیطی متغیر، تغییر کنند. از این‌رو به‌منظور یکسان‌سازی فضای ویژگی مبدأ و مقصد، فراهم‌شدن امکان استفاده از روش‌های طبقه‌بندی با نظارت و همچنین ایجاد شرایط استفاده از انتقال دانش برای به‌روزرسانی الگوی رفتاری کاربر، ایده ساخت فضای معنایی مشترک برای حل معضل اساسی این روش مطرح گردید.

در ادامه با طرح سه سناریو عملکرد مدل طبقه‌بندی در مواجهه با یک رفتار جدید ارزیابی گردید. نتایج نشان داد احراز هویت پیوسته بر اساس دسته‌بندی‌های معنایی برنامه‌های کاربردی، موجب جای‌گیری برنامه‌های ناشناس در دسته‌بندی‌های مرتبط گردیده و منجر به تغییرنکردن در دقت و عملکرد مدل در رویارویی با برنامه‌های ناشناس و مرتبط می‌گردد. همچنین در صورت اضافه یا کم شدن یک برنامه کاربردی، مدل پیشنهادی این پژوهش عملکرد بهتری نسبت به استفاده‌نکردن از راهکار پیشنهادی خواهد داشت. در صورت تغییر در ترتیب یا مدت‌زمان استفاده از برنامه‌ها، دقت سیستم شناسایی کاهش می‌یابد و مدل اولیه نیاز به بازآموزی پیدا می‌کند. در نهایت با مقایسه روش پیشنهادی این مقاله با روش ارائه‌شده در مقاله [۱۱] در رویارویی با یک برنامه دیده‌نشده در زمان ارزیابی، برتری عملکرد روش ارائه‌شده در این پژوهش اثبات می‌گردد.

۶- منابع

[1] W. Meng, D. S. Wong, S. Furnell, and J. Zhou, "Surveying the development of biometric user authentication on mobile phones," *IEEE Communications Surveys & Tutorials,* vol. 17, no. 3, pp. 1268-1293, 2014.

[2] P. Sari, G. Ratnasari, and A. Prasetio, "An evaluation of authentication methods for smartphone based on users' preferences," in *IOP Conference Series: Materials Science and Engineering*, 2016, vol. 128, no. 1: IOP Publishing, p. 012036.

[3] D. Bhattacharyya, R. Ranjan, F. Alisherov, and M. Choi, "Biometric authentication: A review," *International Journal of u-and e-Service, Science and Technology,* vol. 2, no. 3, pp. 13-28, 2009.

[4] A. Eng and L. A. Wahsheh, "Look into my eyes: A survey of biometric security," in *2013 10th International Conference on Information Technology: New Generations*, 2013: IEEE, pp. 422-427.

[5] I. M. Alsaadi, "Physiological biometric authentication systems, advantages, disadvantages and future development: a review," *international journal of scientific & technology research,* vol. 4, no. 12, pp. 285-289, 2015.

[6] D. Gafurov, K. Helkala, and T. Søndrol, "Biometric Gait Authentication Using Accelerometer Sensor," *JCP,* vol. 1, no. 7, pp. 51-59, 2006.
[7] T. Feng *et al.*, "Continuous mobile authentication using touchscreen gestures," in 2012 IEEE Conference on Technologies for Homeland Security (HST), 2012: IEEE, pp. 451-456.
[8] Y. Yang, B. Guo, Z. Wang, M. Li, Z. Yu, and X. Zhou, "BehaveSense: Continuous authentication for security-sensitive mobile apps using behavioral biometrics," Ad Hoc Networks, vol. 84, pp. 9-18, 2019.
[9] L. Fridman, S. Weber, R. Greenstadt, and M. Kam, "Active authentication on mobile devices via stylometry, application usage, web browsing, and GPS location," *IEEE Systems Journal,* vol. 11, no. 2, pp. 513-521, 2016.
[10] F. Bergadano, D. Gunetti, and C. Picardi, "User authentication through keystroke dynamics," *ACM Transactions on Information and System Security (TISSEC),* vol. 5, no. 4, pp. 367-397, 2002.
[11] U. Mahbub, J. Komulainen, D. Ferreira, and R. Chellappa, "Continuous authentication of smartphones based on application usage," *IEEE Transactions on Biometrics, Behavior, and Identity Science,* vol. 1, no. 3, pp. 165-180, 2019.
[12] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine,* vol. 33, no. 4, pp. 49-61, 2016.
[13] E. Shi, Y. Niu, M. Jakobsson, and R. Chow, "Implicit authentication through learning user behavior," in *International Conference on Information Security*, 2010: Springer, pp. 99-113.
[14] F. Li, N. Clarke, M. Papadaki, and P. Dowland, "Behaviour profiling for transparent authentication for mobile devices," 2011.
[15] H. G. Kayacik, M. Just, L. Baillie, D. Aspinall, and N. Micallef, "Data driven authentication: On the effectiveness of user behaviour modelling with mobile device sensors," *arXiv preprint arXiv:1410.7743,* 2014.
[16] S. Alotaibi, A. Alruban, S. Furnell, and N. Clarke, "A novel behaviour profiling approach to continuous authentication for mobile applications," in *The 5th International Conference on Information Systems Security and Privacy*, 2019, pp. 1-6.
[17] Y. Ashibani and Q. H. Mahmoud, "A machine learning-based user authentication model using mobile App data," in *International Conference on Intelligent and Fuzzy Systems*, 2019: Springer, pp. 408-415.
[18] Y. Ashibani and Q. H. Mahmoud, "A multi-feature user authentication model based on mobile app interactions," *IEEE Access,* vol. 8, pp. 96322-96339, 2020.
[19] Zhang, X.; Zhang, P.; Hu, H. Multimodal continuous user authentication on mobile devices via interaction patterns. Wirel.Commun. Mob. Comput. 2021,2021, 567797
[20] J. Dybczak and P. Nawrocki, "Continuous authentication on mobile devices using behavioral biometrics," 2022 22nd IEEE International Symposium on Cluster, Cloud and Internet Computing (CCGrid), Taormina, Italy, 2022, pp. 1028-1035, doi: 10.1109/CCGrid54584.2022.00125.
[21] Wang, Y., Zhang, X., & Hu, H. (2023). Continuous User Authentication on Multiple Smart Devices. Information, 14(5), 274. https://doi.org/10.3390/info14050274
[22] I. Stylios, S. Kokolakis, O. Thanou, and S. Chatzis, "Behavioral biometrics & continuous user authentication on mobile devices: A survey," Information Fusion, vol. 66, pp. 76-99, 2021.
[23] P. K. Rayani and S. Changder, "Continuous user authentication on smartphone via behavioral biometrics: a survey," Multimedia Tools and Applications, vol. 82, no. 2, pp. 1633-1667, 2023.
[24] W. Wang and C. Miao, "Activity recognition in new smart home environments," in *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care*, 2018, pp. 29-37.
[25] Y. Yu *et al.*, "Detecting abnormal behaviors in smart home," in *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems Workshops (MASSW)*, 2019: IEEE, pp. 37-42.
[26] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," ed: Institute for Cognitive Science, University of California, San Diego La …, 1985.
[27] F. A. Silva, A. C. Domingues, and T. R. B. Silva, "Discovering mobile application usage patterns from a large-scale dataset," *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 12, no. 5, pp. 1-36, 2018.

محدثه عالی در سال ۱۳۹۷ مدرک کارشناسی خود را در رشته‌ی مهندسی کامپیوتر با گرایش نرم‌افزار در دانشگاه قم دریافت و پس از آن مدرک کارشناسی ارشد خود را در رشته‌ی مهندسی فناوری اطلاعات با گرایش تجارت الکترونیک از همان دانشگاه اخذ نمود. موضوع پایان‌نامه کارشناسی ارشد او "تصدیق هویت پیوسته افراد در گوشی‌های هوشمند بر اساس الگوهای رفتاری" بوده است. از علائق تحقیقاتی ایشان می‌توان به زمینه‌های یادگیری عمیق و احرازهویت پیوسته اشاره کرد. نشانی رایانامه ایشان عبارت است از:

Aali.mk96@gmail.com

فاطمه سادات لسانی مدرک کارشناسی خود را در رشته مهندسی کامپیوتر از دانشگاه قم و مدرک کارشناسی ارشد و دکتری خود را در رشته مهندسی فناوری اطلاعات از دانشگاه قم دریافت کرده است. ایشان از سال ۱۴۰۲، استادیار گروه مهندسی کامپیوتر دانشگاه صنعتی قم است. علایق پژوهشی ایشان شناسایی الگو، محاسبات فراگیر، سیستم‌های زمینه‌آگاه و اینترنت اشیا است. نشانی رایانامه ایشان عبارت است از:

lesani@qut.ac.ir

فرانک فتوحی قزوینی مدرک کارشناسی خود را در رشته مهندسی برق گرایش مخابرات در سال ۱۳۷۹ از دانشگاه لندن انگلستان و مدرک کارشناسی ارشد خود را نیز در سال ۱۳۸۰ از همان دانشگاه و در همان رشته دریافت کرده است. ایشان در سال ۱۳۹۰ دکترای خود را در رشته سیستم‌های اطلاعاتی چندرسانه‌ای موبایل و فراگیر از دانشگاه برادفورد انگلستان اخذ کرد. وی هم‌اکنون استادیار گروه مهندسی کامپیوتر و فناوری اطلاعات دانشگاه قم می‌باشد. زمینه‌های پژوهشی مورد علاقه ایشان سیستم‌های سیار، سیستم‌های چندرسانه‌ای، محاسبات فراگیر و سیستم‌های پزشکی از راه دور است. نشانی رایانامه ایشان عبارت است از:

f-fotouhi@qom.ac.ir

[1] Application
[2] Accuracy
[3] Knowledge-based
[4] Token-based
[5] Biometrics
[6] Gait recognition
[7] Touch Dynamics
[8] web browsing patterns
[9] Keystroke Dynamics
[10] Application Usage
[11] Classification
[12] Support vector machine (SVM)
[13] Random forest (RF)
[14] Gradient boosting (GB)
[15] F1-Measure
[16] Precision
[17] Recall
[18] Hand motion
[19] Hold posture
[20] Convolutional Neural Network
[21] Sliding Window
[22] Recurrent Neural Network
[23] Tokenizing
[24] Zero Padding
[25] Embedding Layer
[26] Dropout Rate
[27] Overfitting
[28] Dense
[29] Activation Function
[30] Records
[31] Experimental Setup
[32] Shuffle
[33] Rectified Linear Unit
[34] Adaptive Moment Estimation (Adam)
[35] Loss Function
[36] Binary Cross Entropy
[37] Epoch
[38] Batch Size
[39] Equal Error Rate
[40] ASCII printable characters
[41] Index