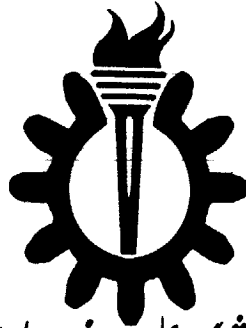


۱۳۸۰ / ۱۲ / ۲۸



دانشگاه علم و صنعت ایران

دانشکده مهندسی کامپیوتر

**تشخیص صحت امضاهای فارسی
به روش شبکه‌های عصبی فازی**

016465

علیرضا نیکیان

پایان نامه جهت اخذ درجه کارشناسی ارشد

در رشته مهندسی کامپیوتر

گرایش هوش مصنوعی و ریباتیک

استاد راهنما: دکتر محمود فتحی

تیر ۱۳۷۸

۲۹۱۰۱

تقدیم به پدر و مادر عزیز و بزرگوارم

چکیده

در این پایان نامه، با استفاده از شبکه‌های عصبی فازی، سیستم تشخیص صحت برون‌خط امضاهاى فارسى پياده‌سازى شده است. اين سيستم شامل چهار بخش است: اکتساب داده، پیش پردازش، استخراج ویژگی و فرایند مقایسه. در اکتساب داده، از یک بانک داده امضای مناسب که در یک پریود زمانی طولانی جمع‌آوری شده، استفاده می‌شود و امضاهاى اصلى و امضاهاى جعلى که به صورت ماهرانه جعل شده‌اند، توسط اسکنر به صورت تصاویر ایستا در کامپیوتر ذخیره می‌شوند. بعد از آن، چندین پیش پردازش مانند فیلتر کردن، آستانه‌گیری و نازک‌سازی تصویر به کار گرفته می‌شود. سپس، چند نوع ویژگی مانند خط مبنای سراسری، ویژگیهای شیب، ویژگیهای بر اساس الگوی مرجع و غیره از روی تصویر امضا استخراج می‌شوند.

بر خلاف اکثر سیستمهای تشخیص صحت اتوماتیک امضای موجود که تکنیکهای سنتی (مانند تکنیکهای خوشه‌بندی) را جهت انجام فرایند مقایسه به کار می‌برند، در سیستم پياده‌سازى شده، از یک شبکه عصبی فازی جدید به نام شبکه عصبی فازی بر اساس Pseudo-Outer Product (POPFNN) استفاده می‌شود. با توجه به ترکیب سیستمهای فازی و شبکه‌های عصبی، POPFNN، مزایای هر دو تکنیک را دارا می‌باشد. ویژگیهای POPFNN مانند قابلیت یادگیری، قابلیت تعمیم‌دهی و قابلیت محاسباتی زیاد، سیستم تشخیص صحت امضا را موقعی که جهت تشخیص جعلهای ماهرانه به کار گرفته می‌شود، بسیار قوی، قابل اعتماد و دقیق می‌نماید. نتایج نشان می‌دهد که سیستم تشخیص صحت امضای پياده‌سازى شده می‌تواند اکثر جعلهای ماهرانه را حتی وقتی که شبکه فقط با نمونه امضاهاى اصلى آموزش داده می‌شود، به طور صحیح تشخیص دهد.

کلمات کلیدی: امضا؛ تشخیص صحت؛ شبکه‌های عصبی فازی؛ شبکه عصبی فازی بر اساس Pseudo-Outer Product (POPFNN)؛ الگوریتم یادگیری Pseudo-Outer Product (POP)

قدردانی و تشکر

بر خود لازم می‌دانم از جناب آقای دکتر محمود فتحی که استاد راهنمای پایان نامه کارشناسی ارشد اینجانب بوده‌اند، به خاطر راهنمایی‌های ایشان در انتخاب و انجام پروژه کارشناسی ارشد و تهیه پایان نامه کارشناسی ارشد، صمیمانه تشکر نمایم. همچنین از رئیس دانشکده مهندسی کامپیوتر جناب آقای دکتر محمدرضا جاهد مطلق و استاد مدعو جناب آقای دکتر احسان الله کبیر عضو هیأت علمی دانشگاه تربیت مدرس که اعضای هیأت داورى جلسه دفاعیه اینجانب بوده‌اند، بسیار متشکرم. از مهندس محمد حسین تقی‌پور و کلیه دوستان و آشنایانی که در ایجاد بانک داده امضا به اینجانب کمک کرده‌اند، بسیار سپاسگزارم.

علیرضا نیکیان

تیر ۱۳۷۸

فهرست مطالب

صفحه	عنوان
۱	فصل اول- مقدمه.....
۱	۱-۱- دست نوشته و طبقه بندی آن.....
۷	۲-۱- امضا و انواع آن.....
۱۰	فصل دوم- تشخیص صحت امضا.....
۱۰	۱-۲- برخط و برون خط.....
۱۳	۲-۲- جعل و انواع آن.....
۱۵	۳-۲- طراحی سیستم تشخیص صحت امضا.....
۱۷	۴-۲- مسائل مربوط به تشخیص صحت امضا.....
۱۸	۱-۴-۲- تغییرپذیری امضا.....
۲۰	۲-۴-۲- ارزیابی کارایی.....
۲۳	۳-۴-۲- مسائل تحقیقاتی.....
۲۴	۵-۲- تشخیص صحت برون خط امضا.....
۲۴	۱-۵-۲- پیش پردازش.....
۲۷	۲-۵-۲- انتخاب ویژگی.....
۲۹	۳-۵-۲- فرایندهای مقایسه و ارزیابی کارایی.....
۳۲	۶-۲- مرور کارهای انجام شده در زمینه تشخیص صحت امضا.....
۳۶	۷-۲- تشخیص صحت ایستای امضا با استفاده از شبکه‌های عصبی.....
۴۰	فصل سوم- ویژگیهای امضا.....
۴۰	۱-۳- مقدمه.....
۴۲	۲-۳- ویژگیهای استخراج شده امضا.....
۵۱	۳-۳- ویژگیهای امضا در مراجع فارسی.....

۵۳.....	فصل چهارم- شبکه‌های عصبی فازی.....
۵۳.....	۴-۱- مقدمه.....
۵۴.....	۴-۲- شبکه عصبی فازی بر اساس Pseudo Outer - Product (POPFNN).....
۵۹.....	۴-۳- یادگیری در شبکه عصبی فازی POPFNN.....
۶۰.....	۴-۴- الگوریتم یادگیری POP.....
۶۶.....	فصل پنجم- پیاده‌سازی سیستم تشخیص صحت برون‌خط امضاهاى فارسی.....
۶۷.....	۵-۱- اکتساب داده.....
۶۷.....	۵-۲- پیش پردازش و استخراج ویژگی.....
۶۹.....	۵-۳- فرایند مقایسه.....
۷۷.....	۵-۴- نتیجه‌گیری و پیشنهادات برای آینده.....
۸۰.....	واژه نامه انگلیسی به فارسی.....
۹۰.....	مراجع.....

فهرست اشکال

صفحه	عنوان
۲.....	شکل ۱-۱- چهار نوع مسأله طبقه بندی.....
۴.....	شکل ۲-۱- بررسیهای اخیر طبقه بندی دست نوشته.....
۱۶.....	شکل ۱-۲- نمودار جریان داده یک سیستم تشخیص صحت امضا.....
	شکل ۲-۲- افراز کردن کلاس و ارزیابی نرخ خطا (از روی میزان شباهت):
۲۲.....	(الف) حالت ایده آل. (ب) حالت عملی.....
۴۰.....	شکل ۱-۳- نمایش یک امضا.....
۴۴.....	شکل ۲-۳- نمایش طرح افقی، طرح عمودی و خط مبنای سراسری یک امضا.....
۴۵.....	شکل ۳-۳- نمایش طرح افقی و الگوی مرجع یک امضا.....
۵۵.....	شکل ۱-۴- ساختار شبکه عصبی فازی POPFNN.....
۵۶.....	شکل ۲-۴- تابع عضویت Bell-Shaped.....
۵۷.....	شکل ۳-۴- ند قانون k ام در لایه پایگاه قانون.....
۵۸.....	شکل ۴-۴- ند برچسب خروجی (l, m) ام در لایه نتیجه.....
۶۰.....	شکل ۵-۴- خوشه‌ها در فضای ورودی خروجی به عنوان قوانین.....
۶۱.....	شکل ۶-۴- خوشه‌ها در فضای ورودی خروجی و تصاویر آنها.....
	شکل ۷-۴- یک POPFNN نمونه با دو متغیر زبانی ورودی و یک متغیر زبانی
۶۳.....	خروجی.....
۶۵.....	شکل ۸-۴- بکارگیری الگوریتم یادگیری POP جهت تعیین کردن قوانین فازی.....
۶۶.....	شکل ۱-۵- ساختار سیستم تشخیص صحت امضای اتوماتیک.....

فهرست جداول

صفحه	عنوان
۳۱.....	جدول ۱-۲- کارایی‌های سیستمهای تشخیص صحت ایستای امضا.....
	جدول ۲-۲- سیستمهای تشخیص صحت ایستای امضا با استفاده از شبکه‌های
۳۷.....	عصبی (S = نمونه، W = نویسنده، F = جعل).....
۷۴.....	جدول ۱-۵- طبقه بندی POPFNN های طراحی شده.....
۷۵.....	جدول ۲-۵- مقایسه کارایی‌های POPFNN های نوع I و POPFNN های نوع II.....
۷۵.....	جدول ۳-۵- متوسط نرخ‌های خطا بر روی کلیه امضاهاى بانک امضا.....
	جدول ۴-۵- مقایسه خطاهای FRR و FAR در POPFNN های نوع I و
	POPFNN های نوع III از امضاهاى یک فرد با در نظر گرفتن
۷۶.....	ویژگی خط مبنای سراسری.....
	جدول ۵-۵- مقایسه خطاهای FRR و FAR در POPFNN های نوع I و
	POPFNN های نوع III از امضاهاى یک فرد با در نظر گرفتن
۷۷.....	ویژگی شیب.....

فصل اول

مقدمه

۱-۱- دست نوشته و طبقه بندی آن

طبقه بندی شیء به معنای انتساب شیء به یک کلاس می باشد. همان طور که در شکل ۱-۱ نشان داده شده است، مسائل طبقه بندی شیء به طور کلی به چهار نوع: شناخت یا یادگیری، شناسایی، تشخیص هویت و تشخیص صحت تقسیم می شوند که بستگی به دانش پیشین^۱ موجود درباره کلاسهای تحت مطالعه و نوع اطلاعات استخراج شده از اشیای مختلف دارد. هر شیء شامل بخش معنایی^۲ و بخش منفرد^۳ است [۷].

در اینجا، دست نوشته به عنوان شیء در نظر گرفته می شود. دست نوشته شامل نقاشی، رسم، متن، امضا و غیره است. در این نوشتار، منظور از دست نوشته، متن و امضا می باشد.

^۱ *a priori knowledge*

^۲ *semantic part*

^۳ *singular part*

طبقه بندی: انتساب یک شیء به یک کلاس

شیء: بخش معنایی + بخش منفرد

کلاسهای از قبل شناخته شده	کلاسهای از قبل ناشناخته	نوع دانش شیء
شناسایی	شناخت یا یادگیری	بخش معنایی
تشخیص صحت	تشخیص هویت	بخش منفرد

شکل ۱-۱- چهار نوع مسأله طبقه بندی.

بخش معنایی دست نوشته، اشاره به محتوی پیام^۱ داده دستنویس دارد و بخش منفرد دست نوشته، تعدادی از مشخصات ویژه نویسنده را منعکس می نماید. با فرض اینکه هر دو نوع اطلاعات (بخش) بتوانند توسط کامپیوتر به طور جداگانه پردازش شوند، چهار نوع سیستم می تواند فرق گذاشته شود. سیستمهای شناخت و شناسایی به واسطه های انسان- کامپیوتر پردازش کننده اطلاعات معنایی یک پیام اشاره می کنند تا محتوی آن پیام را صرفنظر از نویسنده، بازیابند. موقعی که کلاسها از قبل شناخته شده باشند (مثلاً نشانه های حرفی عددی برای کاراکترهای مجزا) شناسایی دست نوشته، به کار برده می شود و زمانی که کلاسها شناخته شده نیستند (همان طور که در بعضی راههای مطالعه حروف الفبای شکسته است) یک مرحله شناخت قبل از رفتن به مرحله شناسایی، برای تعریف این کلاسها لازم است [۷].

به طرز مشابهی، سیستمهای تشخیص هویت و تشخیص صحت، واسطه های پردازش کننده اطلاعات منفرد دست نوشته را توصیف می نمایند تا صرفنظر از محتوی دستنویس، هویت نویسنده را ثابت نمایند. یک سیستم تشخیص هویت نویسنده باید به وسیله مقایسه کردن تعدادی از خصوصیات بخصوص دست نوشته نویسنده با آن

^۱ message content

خصوصیات از همه نویسندگان ثبت شده در یک پایگاه داده مرجع، هویت یک نویسنده را ثابت نماید. از طرف دیگر، یک سیستم تشخیص صحت، بر اساس یک فرایند مقایسه یک به یک، هویت ادعا شده یک نویسنده را مورد تصمیم‌گیری قرار می‌دهد [۷].

به طور تاریخی، اکثر پروژه‌های طبقه‌بندی دست‌نوشته به سوی تحلیل متن یا امضا، گرایش پیدا کرده‌اند. از یک طرف، سه ائیل شناسایی متن محبوبترین بوده‌اند، اگر چه چندین گروه در مسائل تشخیص هویت و تشخیص صحت، فعال بوده‌اند. موقعی که امضاها به کار برده می‌شوند، بررسی تصدیق صحت به یک عمل تشخیص صحت تبدیل می‌شود و زمانی که متن به کار برده می‌شود، معمولاً آزمایشهای تشخیص هویت انجام می‌گردد. از طرف دیگر، مسائل شناخت اکثراً در رابطه با روانشناسی تجربی بررسی شده‌اند. در حقیقت چندین مدل خیالی، فیزیکی و تجربه‌ای برای تحلیل نمودن پیچیدگی فرایندهایی که در ایجاد دست‌نوشته دخیل هستند، ایجاد شده‌اند. اکثر آنها از یک نظر قبول شده عمومی از فرایندهای بیوفیزیکی وارد شده در ایجاد دست‌نوشته سریع، ناشی می‌شوند. اینکه چه بخشهایی از فرایند ایجاد دست‌نوشته وابسته به نویسنده (اطلاعات منفرد) و چه بخشهایی مستقل از نویسنده (اطلاعات معنایی) است، یک مسأله می‌باشد [۷].

به هر حال مطالعه این مدلها خارج از بحث این نوشتار می‌باشد و خواننده می‌تواند به مراجعی که در مرجع [۷] به آنها اشاره شده است، مراجعه نماید.

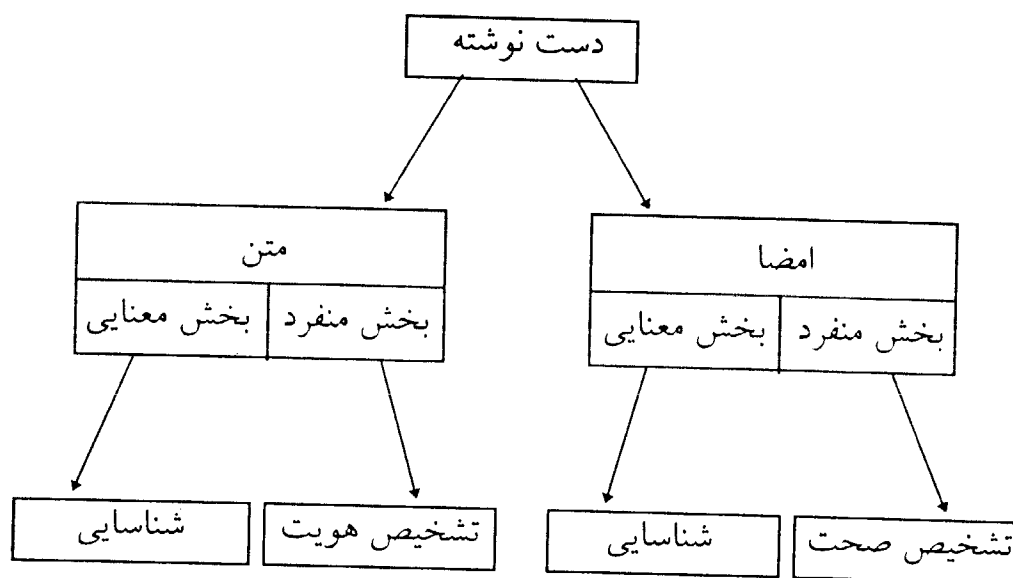
در مرجع [۷] که مقاله‌ای با استفاده از جدیدترین و پیشرفته‌ترین تکنیکها و روشها^۱ می‌باشد و مقاله‌ای با نظر اجمال و با مرور کارهای انجام شده تا قبل از سال ۱۹۸۹ است و مورد مراجعه اکثر مقالات می‌باشد و توسط پلاموندون^۲ و لورت^۳ در سال ۱۹۸۹ ارائه شده، هیچ اشاره‌ای به شناسایی امضا نشده است در حالیکه در تعدادی از

^۱ the state of the art paper

^۲ R. Plamondon

^۳ G. Lorette

مقالات، در مورد شناسایی امضا، بحث شده است. بنابراین با توجه به این مطلب و [۷]، بررسیهای اخیر طبقه بندی دست نوشته را می توان به صورت شکل ۱-۲ نشان داد. باید توجه داشت که برای شناسایی امضا، یک مرحله شناخت قبل از رفتن به مرحله شناسایی، برای تعریف کلاسهای مختلف امضا، لازم است.



شکل ۱-۲- بررسیهای اخیر طبقه بندی دست نوشته.

شناسایی دست نوشته انسان، شامل کاراکترها، اعداد و امضاها برای مدت طولانی یک موضوع تحقیقی مهم بوده است. بعد از حدود ۳۰ سال، فن آوری OCR^۱ به نقطه ای که هم رسم و هم متن در اسناد به وسیله ماشین ایجاد شده و با تنوع وسیعی از سبکهای چاپ و اندازه های صفحه، بتوانند شناسایی شوند، ارتقا یافته است، ولی شناسایی کاراکترها، اعداد و امضاها دستنویس به صورت یک مسأله مشکل، باقی مانده است. شناسایی حروف الفبای دستنویس بسیار شکسته به علت پیوستگی ذاتی بین کاراکترها و تنوع در سبک، مشکلتر می باشد. توپولوژی و سبک یک کاراکتر دستنویس از شخصی به

^۱ Optical Character Recognition technology

شخص دیگر تفاوت دارد و برای یک شخص خاص هم، در آزمایشهای مختلف، اندکی تفاوت وجود دارد [۱۵].

تشخیص صحت وابسته به معیار حیاتی^۱ یک زمینه تحقیق مهم در کاربردهای اتوماتیک تشخیص صحت هویت، به حساب می‌آید. تجربه‌های امنیتی فعلی معمولاً شامل استفاده از اعداد PIN^۲ (شماره تشخیص هویت شخصی) و کلمات عبور و کارتهای اجازه ورود هستند. این علامتها به دلیل اینکه می‌توانند فراموش شوند و یا گم شوند، خیلی قابل اعتماد نیستند و محدودیتهای بیشتری برای اینکه بتوان جلوی یک شخص غیر مجاز را برای بکارگیری آن علامتها در یک محیط تشخیص صحت اتوماتیک ماشینی، گرفت، وجود ندارد. از طرف دیگر، از اندازه‌گیریهای وابسته به معیار حیاتی به راحتی المثنی برداشته نمی‌شود و نمی‌توانند گم شوند و یا دزدیده شوند، بنابراین مطمئنتر هستند [۱۳].

دو نوع معیار حیاتی وجود دارد: معیار حیات فیزیولوژیکی مانند الگوی عنبیه چشم و اثر انگشت و معیار حیات رفتاری مانند صحبت و دست نوشته [۱۳].
یک نمونه امضا (امضای دستنویس) یک الگوی امضا می‌باشد که یک شخص در آرامش کامل و موقعیت جسمی و روحی پایدار ایجاد می‌نماید و آن شخص می‌تواند در همان شرایط، به دفعات قطع نظر از اثرات جانبی قلم و کاغذ، مشابه‌های آن را ایجاد نماید ولی البته به طور طبیعی در طی دوره‌های زمانی طولانی، در امضای او، تغییرات جزئی رخ خواهد داد. این مشخصات، امضای دستنویس را در بسیاری از سیستمهای مالی و امنیتی یک نشانه اعتبار قرار می‌دهد [۱۸].

اگر چه پایداری در امضاها مشاهده می‌شود، ولی امضا کردن یک فرایندی که برای آزمایشهای متوالی، مشخصات کاملاً دقیق، قابل تکرار و یکسان بدهد، نیست. تنها اطمینانی که در این زمینه وجود دارد این است که وقتی دو امضا یکسان هستند، یکی از آنها امضای جعلی است؛ یعنی، احتمالاً یک کپی است [۷].

^۱ biometric verification

^۲ Personal Identification Number

گمان می‌شود که متن دستنویس نسبت به یک امضا (امضای دستنویس) پایدارتر است چرا که طولانی‌تر است و شامل اطلاعات بیشتری می‌باشد اگر چه یک فرایند کاملاً پایدار و ثابت نیست [۷].

غالباً از افراد خواسته می‌شود تا هویت خود را موقع نوشتن اسناد مالی ثابت نمایند. این کار برای بازداشتن آنها از بکارگیری اثر شخص دیگر و همچنین بازداشتن آنها از رد کردن سندی که آنها قبلاً قبول کرده‌اند، انجام می‌پذیرد. به دلیل اینکه متن‌ها غالباً تایپ می‌شوند، امکانپذیر نیست تا صحت و سقم این سندها را از روی متن دست نوشته، معلوم ساخت. به هر حال، مرسوم است که یک علامت برای ثابت نمودن نویسنده سند، الحاق می‌کنند تا نشان دهد که او با متن سند موافق است. امروزه این علامت معمولاً یک امضای دستنویس است [۲۳]. مورد دیگری که وجود دارد این است که افراد در نوشتن یک متن، حساسیت زیادی ندارند تا همانند دستخطهای گذشته خود بنویسند.

اینها همه حاکی از این است که امضا به عنوان یک دست نوشته، بیشتر از یک متن دستنویس کاربرد دارد و بهتر از آن است و می‌تواند به عنوان یک معیار حیات رفتاری مناسب در نظر گرفته شود. در مورد امضا، دو تحلیل وجود دارد که عبارتند از: تشخیص صحت امضا و شناسایی امضا.

تشخیص صحت امضا و شناسایی امضا با انواع مختلفی از مسائل سر و کار دارند. اولی تحقیق می‌کند که یک امضای داده شده به شخصی که مشخصات امضای او قبلاً تهیه شده است، تعلق دارد یا نه، در حالیکه دومی برای تشخیص هویت یک امضای داده شده، یک بانک داده امضا را جستجو می‌نماید [۱۵]، [۱۴].

امروزه نیاز به تشخیص صحت اتوماتیک امضا و شناسایی اتوماتیک امضا، بیشتر و بیشتر آشکار می‌شود. یک فرایند تشخیص صحت اتوماتیک امضا، بانکها و سایر مؤسسه‌های مالی را قادر می‌سازد تا به طور چشمگیری، جعلهای چک و جعلهای حواله را کاهش دهند. از طرف دیگر یک سیستم شناسایی اتوماتیک امضا، برای شرکتها یک