وزارت علوم تحقیقات و فناوری
موسسه آموزش عالی سلمان

پایان نامه کارشناسی

پردازش تصویر

ارائه دهنده: مهدی ارباب تفتی
استاد دفاع:مهندس معارف دوست
استاد راهنما: مهندس حسن شاکری

دیماه ۱۳۸۷

چکیده

OCR یـا Optical CharacterRecognition بـه معنـای شناسایی حروف با کمک ابزار نوری است ویک شـاخه از علـم پـردازش سیگنال را تشکیل می دهد . OCR شامل تصویر برداری از روی کاغـذ واستخراج کلمات از آن است.

متن به صورت مستقیم در تصویر قابل دسترسـی نیسـت زیـرا حروف در یک متن اسکن شده ازمجموعه ای از نقاط تشکیل می شـود که در کنارهم قرار گرفتن این نقاط مطابق الگوهای معین، تصویرحروف مختلف را می سازد

فهرست

فصل ۱: مقدمه.....۸

۱-۱:آشنایی با OCR.....۹

۱-۲:پردازش تصاویر ماهواره.....۱۲

۱-۳: پردازش تصویر در صنعت و پزشکی.....۱۲

فصل ۲:مراحل OCR ۱۵

۲-۱:سیستمهای بازشناسی حروف.....۱۶

۲-۲: اسکن کردن صفحات.....۲۱

۲-۳: استخراج نواحی شامل متن ۲۳

۲-۴: شناسایی حروف.....۲۷

۲-۵: ارائه نتایج ۳۰

فصل ۳:شبکه های عصبی.....۳۱

۱-۳:معرفی شبکه های عصبی.....۳۲

۱-۱-۳: انواع شبکه های عصبی.....۳۳

۲-۳: نرون.....۳۷

۳-۳: شبکه های Multilayer Perceptron..... ۳۹

۳-۴: Overlearning.....۴۳

فصل۴: شرح یک پروژه انجام شده.....۴۸

۴-۱: شرح پروژه.....۴۹

۴-۲: تصاویر مورد استفاده.....۴۹

۴-۳: تبدیل تصاویر به بردارهای ویژگی.....۵۰

۴-۴: آموزش شبکه عصبی.....۵۴

۴-۵: طراحی ابتدایی.....۵۵

۴-۶: تغییرات و ساده سازیها.....۵۹

فصل۵: ویژگی های متون فارسی.....۶۲

۵-۱: خصوصیات متون فارسی.....۶۳

فصل ۶: برسی چند محصول OCR..... ۶۶
۶-۱: چند نمونه OCR ۶۷
۶-۲: آموزش Readiris Pro 11 ..... ۶۷
۶-۳: آموزش VajehShenas V.1..... ۷۱
۶-۴: مقایسه VajehShenas با Readiris..... ۷۵
فصل ۷: نتایج و پیشنهادات..... ۷۷
مراجع..... ۸۱

فهرست شکل ها

شکل ۱-۲ نمونه فرم اسکن شده..... ۲۵

شکل ۲-۲ یک نرون با تابع فعال ساز..... ۳۷

شکل ۳-۲. چند تاج فعال سازی رایج..... ۳۹

شکل ۴-۲ Multilayer Perceptron..... ۴۱

شکل ۵-۲ overfitting چند جمله ای..... ۴۵

شکل ۶-۲ overlearningدر شبکه عصبی..... ۴۶

شکل ۱-۳. یافتن Bounding Box ۵۱

شکل ۲-۳. روشهای تغییر اندازه..... ۵۲

شکل ۱-۵..... ۶۸

شکل ۲-۵..... ۶۹

شکل ۳-۵..... ۷۰

شکل ۴-۵..... ۷۲

شکل ۵-۵.....۷۳

شکل ۶-۵.....۷۳

شکل۷-۵.....۷۴

فصل ۱:

مقدمه

۱–۱:آشنایی با OCR

پردازش و آنالیز تصاویر می‌تواند به عنوان یک ساختار کاربردی و تکنیکی جهت تسخیرکردن، تصحیح کردن، زیاد کردن و تغییر شکل دادن تصاویری که مشاهده می‌شود تعریف کرد.

در هنگام استفاده از تکنولوژی OCR ، نرم افزار مربوطه تصویر را تجزیه وتحلیل می نماید و شکل حروف را بر اساس پراکندگی نقاط در تصویر، شناسایی می کند. متون می توانندحتی دارای عکس هم باشند که سیستم آنها را تشخیص داده، کنار می گذارد.

یک سیستم OCR به ما این امکان را می دهد که یک کتاب و یا یک مقاله را مستقیما به یک فایل الکترونیکی تبدیل نماییم و آن را با کمک یک پردازشگر تغییر دهیم. استفاده از سیستم OCR یک راه کارآمد برای تبدیل متون به فایلهای داده است که ممکن است تایپ کردن آنها زمان زیادی به طول بیانجامد.این تکنولوژی مدتهاست که به وسیله کتابخانه ها و سازمان های دولتی برای دستیابی الکترونیکی

سریع به مدارک حجیم به کار می رود و از لحاظ سرعت و هزینـه روش مناسبی است .OCR از معدود زمینه های هوش مصنوعی است که می توان در عمل به آن تکیه کرد .

از متداول ترین کاربردهای این تکنولوژی می توان جـداکردن نامه ها در مراکز پست، دسته بندی چکها
بانک، تبدیل متن به صدا برای استفاده نابینایان، امکان دسترسی سریع به بایگانیهای عظیم و تبدیل کتابهای قدیمی کتابخانه های بـه فرمتـی مناسب برای ذخیره در کامپیوتر را نام برد.

هدف از این عملیات زیاد کردن کیفیت نسبی اطلاعاتی اسـت که بعداً استخراج خواهد شد. تغییر شکل در حقیقت عملیاتی است کـه بر روی تصاویر ورودی انجام مـی‌گیـرد تـا خروجـی بـه یـک عـدد کـه اطلاعاتی از تصویر اسـت بدهـد. تکنیکهـای متفـاوت تولیـد تصـاویر در دسترس می‌باشد با وجود این تکنیکهای پردازش اطلاعات بـه عملیـات

گسترده ای اطلاق می‌شود که می‌تواند بر روی اشکال و فرمتهای مختلف تصاویر عمل کند.

ابتدا تصاویر جهت از بین بردن اعوجاجی که در مرحله جمع‌آوری تصاویر ایجاد می‌شوند تشخیص می‌گردند و تصحیح سپس پردازش تصاویر یک عملیاتی است که یک سری کاربردها را ارجحیت می‌بخشد. تکنیکهای جدید یا پروسه‌های جدید که جهت حل مسائل مشخص وجود دارند از میان کاربردهای مختلف انتخاب می‌شوند.

پردازش تصاویر در سطوح مختلف انجمن‌های مدرن انجام می‌گیرد. روشها و تکنیکها در همه این سطوح به یک شکل می‌باشد و روشهایی که در مقوله مراقبت سلامتی انجام می‌گیرند از سایر روشهای پردازش تصاویر که در علوم مختلف و صنعت انجام می‌گیرد اقتباس می‌شوند. جهت توضیح دادن عمومیت پردازش تصاویر، تعدادی از کاربرهای پزشکی و غیر پزشکی در زیر آمده شده است.

۲-۱: پردازش تصاویر ماهواره:

تصاویر زمین و سایر سطوح جهانی توسط دوربین‌هایی که برروی سفینه‌های فضایی و ماهواره‌ها نصب شده گرفته می‌شود. هنگامی که از سیارات دیگر تصویر گرفته می‌شود، هدف مطالعه سطوح آنها جهت بدست آوردن آگاهی درباره تاریخچه و مشخصات آن می‌باشد. شبیه‌سازی ماهواره‌ای زمین جهت کارهایی از قبیل کارهای فلاحتی- زراعتی، هواشناسی، محیط اطراف، پردازشهای نظامی در بین سایر موارد می باشد.

۳-۱ پردازش تصویر در صنعت و پزشکی:

تصاویر بخشهایی از سطوح زمین جهت کاهش تصویر و کوچکتر کردن، بازرسی و یا جهت نقشه‌کشی گرفته می‌شود.

- چاپ کردن و دست نوشته‌ها:

تشخیص کاراکترهای چاپ یا نوشته شده از کاربردهای جدید پردازش تصاویر است. در بین کاربردهای دیگر می‌توان به خواندن اتوماتیک کدهای پستی و کدهای بانکی اشاره کرد.

- بارکد:

مسئول وارسی در بیشتر سوپرمارکتها دارای وسایلی هستند که بارکد موجود بر روی کالاها را خوانده و تفسیر می‌کند. از استفاده‌های بارکد می‌توان به نمونه‌های مدارک، نمونه‌های بیوشیمیایی و نمونه بافت در بیمارستان اشاره کرد.

- رادیولوژی:

در مبحث پزشکی، رادیولوژی یکی از کاربردهای اصلی تکنیک پردازش تصاویر است. کاربردهای آن نه تنها در اشعه x می‌باشد بلکه کاربردهایی از قبیل مقطع‌بندی کامپیوتری PET,MRI,CT , ultra sound تمام نتایج به صورت تصاویر هستند که باید بعداً تفسیر شوند.

در میان این کاربردها، بازسازی سه‌بعدی تصاویر یکی از موضوعهای مورد علاقه است.

- مجموعه خصوصیات کروموزومی موجودات:

از آنجا که تصاویر کروموزومی یکی از ساختارهای ساده می باشد، جفت‌شدگی کروموزومی و بازسازی خودکار آنها یکی از جدیدترین کاربردهای پردازش تصویر در پزشکی است.

- سلول شناسی و بافت شناسی:

تفسیر اتوماتیک و کوانتیزهٔ تصاویر میکروسکوپیک از کاربردهای دیگر پردازش تصاویر در پزشکی است.

فصل ۲:

مراحلOCR

۱-۲:سیستمهای بازشناسی حروف

در این فصل، پس از ذکر تاریخچه ای از OCR ، مراحل تبدیل تصویر به متن را شرح د اده و ورودی و خروجی هر مرحله را بیان می کنیم.

- تاریخچه OCR

مبحث بازشناسی متون، همانند مبحث تشخیص صحبت از اوایل ظهور مبحث شناسایی الگو مطرح بود. در ابتدا شناسایی آنها آسان به نظر می رسید ولی در عمل بر خلاف انتظار بعد از یکسری پیشرفت کوچک، این مسأله بامشکلات بزرگی روبرو شد و توجه محققین به سمت سایر زمینه های شناخت الگو معطوف گشت. با پیدایش وسایل ارتباط جمعی الکترونیکی، نیاز شدیدی به سیستمهایی احساس شده است می شد که بتوانند با دقت و سرعت، اطلاعات نوشتاری موجود را خوانده و ذخیره نمایند.

در سال ۱۹۲۹ Taushech در آلمان و در سال ۱۹۳۳ Handle در آمریکا ابداعاتی در زمینه OCR ثبت نمودند. اینها اولین ایده های شناسایی حروف هستند. این روش، تطبیق الگو نام دارد و به این صورت کار می کند که به هر حرف، نور تابیده می شود و نور باز تابیده شده از حروف، از قالب های مکانیکی عبور داده می شود. هرگاه که نوری از قالب عبور نکرد، حرف تشخیص داده می شود.

در عمل تا زمان به عرصه آمدن کامپیوتر یعنی سال ۱۹۵۰ سیستم مطلوبی ساخته نشد. اولین کامپیوترتجاری ا UNIVAC در سال ۱۹۵۱ در آمریکا شروع به کار کرد. در این زمان بود که ایده OCR به عنوان یک پدیده قابل پیاده سازی پذیرفته شد. از این زمان به بعد تحقیقات گسترده ای بر روی این تکنولوژی انجام گرفته است و در نتیجه سیستمهای تجاری عرضه شده اند که دارای قابلیتهای خوبی هستند. سیستمهای تجاری عرضه شده را می تواند به سه نسل تقسیم بندی نمود.

سیستمهای نسل اول در اویل دهه ۱۹۶۰ بـه بـازار آمدنـد کـه نتیجـه تحقیقـات دهـه ۱۹۵۰ بودنـد. آنهـا فقـط قابلیـت تشـخیص کاراکترهای خـاص بـا انـدازه و فونـت مشخصـی را داشـتند. در آنهـا از روشهایی استفاده شده بود که نسبت به تغییر جای کاراکترها، انـدازه و دوران آنها فوق العاده حساس بودند. شاخص ترین سیسـتم ایـن نسـل، NCR 420 بود که می توانست اعداد و پـنج نمـاد دیگـر را تشـخیص دهد.

سیستمهای نسل دوم در اواسط دهه ۱۹۶۰ بـه بـازار آمدنـد و می توانستند بعضی از کاراکترهای دست نویس مانند اعداد را تشـخیص دهند .RETINA یکی از این سیستمها بود.

سیستمهای نسل سوم مربوط به اواخر دهه ۱۹۶۰ هسـتند کـه باهدف تشخیص کاراکترهایی با کیفیت چاپ پـایین و همچنـین متـون دست نویس لاتین ساخته شدند. این نسل تا بـه امـروزهم ادامـه دارد و فعالیتهـای گسـترده ای در ایـن زمینـه در حـال انجـام اسـت. امـروزه

سیستمهای OCR قادر به تشخیص دقیق کاراکترهای تایپی لاتین با انواع فونتها و در اندازه های متفاوت هستند (تا ٪۹۹) ولی هنوز در مورد تشخیص متون دست نویس لاتین و یا فونتهایی که در آنها از خطوط خمیده استفاده می شود (مثل فارسی و عربی) مشکلات زیادی وجود دارد.

روشهای بسیاری تا به امروز برای شناسایی متون فارسی و عربی به کار رفته اند. در سال ۱۳۵۹ به دنبال ابداع سیستمی برای شناسایی دست نوشته های عربی موسوم به IRAC توجهات به این زمینه نوین در قلمرو بازشناسی الگو معطوف شد. به دنبال آن سیستم دیگری برای تشخیص کلمات تایپ شده عربی پیشنهاد شد که در آن ابتدا کلمات به حروف تفکیک شده و سپس توصیف کننده های فوریه دوره خارجی حروف، مبنای شناسایی آنها قرار می گرفتند. به موازات آن دو روش ساختاری مطرح شدند:

- یکی مبتنی بر ایده دنبال کـردن کـانتور بـرای شناسـایی حـروف مجزای عربی و دیگری بر اساس استفاده از محلهـای تقـاطع و انشـعاب پاره خطها برای شکستن کلمات به زیر حرفها و دسـته بنـدی آنهـا بـه کمک ویژگیهای هندسی و توپولوژیکی.

در سال ۱۳۶۰ روشی برای شناسـایی فونتهـای بـزرگ فارسـی ابداع شد که از ویژگیهایی نظیر وجـود حفـره و دهنـه بـرای شناسـایی حروف بهره می جسـت. در سـال ۱۳۷۲ دو سیسـتم جهـت شناسـایی حروف در متن فارسی تایپی ارائه گردید که یکی با روش شکل شناسی و دیگری بر اساس توصیف کننده های فوریه عمل می نمود. امـروزه بـا کمک شبکه های عصبی که قادرنـد ارتبـاط پیچیـده ای بـین ورودی و خروجی برقرار کنند، می توان حروف فارسی را شناسایی نمود. مرجع [۷] شامل پیوندهایی به چند پایان نامـه انجـام شـده در زمینـه OCR است. مرجع [۸] نیز برخـی پایـان نامـه هـا را بررسـی و نتـایج آنهـا را مقایسه نموده است.

پس از ذکر این تاریخچه، به شرح مراحل مختلف OCR می پردازیم.

۲-۲. اسکن کردن صفحات

اسکن کردن، اولین مرحله کار یک سیستم OCR است. یک پرتو نور به صفحه مورد نظر تابیده می شود. با حرکت این پرتو روی صفحه و با تغییر مقدار سیاهی (و یا رنگ در اسکنرهای رنگی) شدت نور بازتابیده شده تغییر می کند. نور بازتابیده شده به آرایه ای از سلول های فتوالکتردیک برخورد کرده و به این ترتیب، یک فایل تصویری از صفحه اسکن شده بدست می آید. از آنجایی که با دیدن این تصاویر می توانیم محتویات صفحه را دریابیم، ممکن است این سوال پیش آید که چرا باید زمان وهزینه تبدیل تصویربه متن را تقبل کنیم؟ در پاسخ باید گفت:

- حجم فایلهای تصویری بسیار بیشتر از فایلهای متن است، حتی اگر از روشهای فشرده سا زی استفاده کنیم.

- تغییر در نحوه نمایش فایلهای تصویری مشکل و یا عملا" غیر ممکن است، مثلا" تغییر فونت، در صورت امکان بسیار مشکل خواهد بود و یا ریز و درشت کردن تصویر می تواند به ناخوانا شدن حروف بیانجامد.
- تبدیل تصویر به اشکال دیگر ارائه، مثلا" خواند متن برای نابینایان، بسیار مشکل است.
- امکان جستجوی عبارتی خاص و یا مقایسه بخشهایی از متن تصویری با متون دیگر عملا" وجود ندارد.

مجموعه علتهای بالا باعث شده تا تلاشهای بسیاری برای تبدیل تصاویر به متن صورت گیرد. روشهای دیگری، از جمله استفاده از دوربین دیجیتال، نیز برای ایجاد فایلهای تصویری وجود دارد. این روشها معمولا" در زمینه های دیگر تشخیص الگو (مثل تشخیص چهره) بکار می روند.

در مورد تبدیل متون کتابها و سایر نوشته ها، اسکن کردن مناسب ترین روش است. البته در بعضی کاربردها (مثل LPR) استفاده از این دوربینها ضروری است.

۳-۲. استخراج نواحی شامل متن و تفکیک متن به حروف

یک فایل تصویری اسکن شده، فقط شامل متن نیست بلکه ممکن است شامل تصویر اشخاص و مناظر، نمودار و سایر تصاویر غیر متنی نیز باشد. سیستم OCR باید بتواند نواحی شامل متن را شناسایی و استخراج کند.

الف) اگر چیدمان صفحه اسکن شده مشخص باشد (مثل چک بانکی، پلاک خودرو، فرم ثبت نام داوطلبین آزمونها) این کار با روشی ساده و با قابلیت اطمینان بالا امکان پذیر است. مثلا" در شکل (۱-۲) با استفاده از موقعیت مکانی مستطیل های سیاه کوچک حروف مورد نظر به سادگی استخراج می شوند.

ب) اما اگر چیدمان صفحه مشخص نباشد (مثل صفحات کتاب) باید از روشهای پیچیده تری بهره گرفت که بتوانند بر اساس تفاوتهای تصاویر متنی با تصاویر اشخاص و مناظر، متن را استخراج کنند. بدیهی است که در این کار امکان خطا وجود دارد.

یکی از روشهای مورد استفاده در این حالت، استفاده از هیستوگرام (histogram) است. هیستوگرام تصاویر متنی معمولا" حول دو رنگ (نوشته و زمینه) مقدار بالا دارند ولی تصاویر غیر متنی در بسیاری از رنگها، مولفه دارند. برای کسب اطلاعات بیشتر در مورد روشهای استخراج نواحی متنی می توانید به [۱] مراجعه کنید.

شکل ۱-۲ نمونه فرم اسکن شده

تصاویر متن پس از استخراج، بر اساس یک آستانه که توسط روشهای آماری بدست می آید ([۱],[۹]) به تصاویر سیاه و سفید (باینری) تبدیل شده و توسط فیلترهای مثل LP و Median نویزگیری می شوند [۶] .

اگر سیستم OCR بخواهد خروجی مشابه با ورودی بدهد (مثلا" در تبدیل کتابها که خروجی از لحاظ بصری شبیه صفحه اسکن شده است) باید چیدمان صفحه اسکن شده را ذخیره کند. در

سیستمهای دیگر(مثل سیستمهایی که خروجی آنها صوتی است یا سیستمهای LPR) این کار لزومی ندارد.

پس از انجام این مرحله، متون استخراج شده باید به تعداد محدودی واحد "قابل تشخیص اتوماتیک " تفکیک شوند. معمولاً این واحدها حروف تشکیل دهنده متن هستند، هرچند در برخی الگوریتمها از زیرواژه ها (مثل "با" و "لا" در کلمه "بالا") به عنوان واحد تشخیص استفاده می شود.

در برخی از کاربردها مثل فرم آزمونها (شکل ۲- ۱) دو مرحله استخراج نواحی و تفکیک به حروف، مستقیما" انجام می شود ولی در سایر کاربردها باید از الگوریتمهای ویژه ای استفاده کرد. این کار درمورد حروف چاپی لاتین نسبتا" ساده است زیرا حروف به صورت مجزا در کنارهم قرار می گیرند و بین آنها فاصله خالی وجود دارد. در مورد متون فارسی و سایر زبانهایی که به صورت "سرهم " نوشته می

شوند باید کارهای پیچیده تری صورت گیـرد. (بـه عنـوان یـک مرجـع خوب در این مورد به [۱] مراجعه کنید).

در صـورت انجـام موفـق ایـن مرحلـه، مجموعـه ای از تصـاویر بدست می آید که هر تصویر شامل یک حرف است. این تصاویر بعنـوان ورودی به نرم افزار شناسایی کننده حروف داده می شـوند. بسـیاری از الگوریتمها تصویر رنگی یا خاکستری را نیز بـرای اسـتفاده درپردازشـها ذخیره می کنند.

۴-۲. شناسایی حروف

هدف از این مرحله عبارت است از تبدیل تصاویر "تک حرف " مرحله قبلی به کدهایی که نشانگرحروف مـورد نظـر اسـت (مثـل کـد اسکی حرف متناظر).

به طور کلی روشهای شناسایی حروف بر دو نوع هستند:

- روشهای مبتنی بر تئوری تصمیم گیری

- روشهای ساختاری

در روشهای نوع یک، ویژگیهـای خاصـی از تصـاویر ورودی بـه صورت کمی اندازه گیری و استخراج شده و به صورت درایه هـای یـک بـردار نـرم افـزار شناسـایی کننـده را تشـکیل مـی دهنـد. بـه ایـن کارFeature Extractior می گویند. این ویژگیها می توانند شدت نقاط تصاویر خاکستری (یا مقدار نقاط تصویر باینری) ، projection افقی یا عمودی تصویر، میانگین، واریانس و یا نتایج اجرای یک فراینـد پیچیده روی تصویر باشند. نرم افزار با توجه به بردار ویژگی، به هر یک از کلاسهای تعیـین شـد، امتیـازی نسـبت مـی دهـد. کلاسـهایی کـه بیشترین امتیـاز را کسـب کنـد، خروجـی نـرم افـزار اسـت. مـثلا" در تشخیص حروف لاتین، هر کدام از حروف A تا z وa تا z یک کـلاس هستند و نرم افزارهربردار ورودی را به یکی از این کلاسـها نسـبت مـی دهد. یکی از مهمترین روشهای این نوع، شبکه های عصبی است که در فصل بعد، عملکرد آن را بیان خواهیم کرد.

در روشهای نوع دو، حروف با کمک ویژگیهای ساختاری (نقطه، سرکش، دسته، حلقه،...) و نحوه قرار گیری این ویژگیها در کنارهم شناسایی می شوند.

برای آشنایی بیشتر با انوع روشهای نوع یک و دو می توانید به [۱] مراجعه نمایید. همچنین [۸] شامل بخشی در مورد بکارگیری بعضی از این روشها در کاربردهای عملی و مقایسه نتایج آنهاست.

آنچه در این پایان نامه تاکید میشود روش شبکه های عصبی بدون استفاده از الگوریتمهای پیچیده استخراج ویژگی است.

نشان خواهیم داد که شبکه های عصبی که حروف جایی را بدون بهره گیری از ویژگیهای پیچیده، تا حد قابل قبولی شناسایی می کنند، نمی توانند این کار را در مورد حروف دست نویس انجام دهند و باید ویژگیهای پیچیده ای استخراج شوند.

۵-۲:ارائه نتایج

هدف از این مرحله که مرحله پایانی کار سیستم OCR است، ارائه کدهای تولید شده در مرحله قبل به فرمت مناسب است. مثلا" اگرهدف از OCR ، تبدیل کتاب و مجله باشد، سیستم باید با استفاده از کدهای اسکی بدست آمده و اطلاعات چیدمانی که در مراحل قبل ذخیره شده، صفحه اسکن شده را شبیه سازی کند طوری که از نظر بیننده، تفاوتی بین صفحه کاغذ و فایل تصویر وجود نداشته باشد ولی از نظر رایانه، به جای تصویر حروف، کد آنها جایگزین شده باشد. به همین ترتیب اگرهدف، تبدیل تصویر به صوت است باید کدهای بدست آمده توسط الگوریتمهایی مثل Text to Speech API به اصوات شنیداری تبدیل شوند.

فصل ۳:

شبکه های عصبی

۶-۲:شبکه های عصبی

در صفحات پیشین شبکه عصبی را به عنوان یک روش متداول در شناسایی حروف معرفی کردیم. در این فصل ساختار و عملکرد شبکه های عصبی را بررسی کرده و در مورد ساختار و الگوریتم یادگیری یک نوع خاص شبکه عصبی (شبکه Multilayer Perceptron) که در این پایان نامه به کار رفته است، توضیح خواهیم داد.

۳- ۱. آشنایی با شبکه عصبی

شبکه عصبی سیستمی است متشکل از تعدادی واحد ساده به نام نرون که با اتصالات وزن دار به هم متصل شده اند. با تغییر دادن وزن این اتصالات در فرایندی به نام یادگیری ، شبکه می تواند توابع پیچیده ای را پیاده سازی کند. به طور کلی برای استفاده از شبکه عصبی، ابتدا ساختار مناسب را انتخاب می کنند (بخش ۳-۲ را ببینید)

سپس شبکه را آموزش می دهند تـا وزن اتصـالات (کـه در ابتـدا دارای مقادیر پیش گزیده و یا تصادفی هستند) طوری تنظیم شوند کـه بـرای تعداد محدودی ورودی نتیجه مطلوب حاصل شود. سپس بقیه ورودیهـا را به شبکه آموزش دیده می دهند و نتایج حاصل را دریافت می کنند.

از ویژگیهای شبکه های عصبی می توان بـه مـوارد زیـر اشـاره کرد [۷],[۱۱]:

- می توانند توابع پیچیده و غیر خطی را مدل سازی کنند.
- می توانند بر اساس نمونه ها یاد بگیرند و احتیاجی به دانستن تابع درونی شبکه بصورت تحلیلی نیست.
- به دلیل فعالیت موازی نرونها، شبکه های عصبی پس از یـادگیری، می توانند با سرعت بالایی ورودیها را پـردازش کـرده و نتـایج را بیـرون دهند. تحمل پذیری بـالایی نسـبت بـه نـویز ورودی دارنـد. در صـورت خرابی یک یا چند نرون بازهم تا حد خوبی می توانند به کار خود ادامـه دهند.

امروزه شبکه هـای عصـبی در زمینـه هـای مختلـف پزشـکی، دفاعی، صنعتی و... کاربرد دارند [۹]. در زمینه های گوناگون تشـخیص الگو (مثل تشخیص صحبت، اثر انگشـت، چهـره و OCR) نیـز شـبکه های عصبی، بسیار به کار می روند.

ایده اولیه این شبکه ها، از شبکه های عصـبی طبیعـی گرفتـه شده است. در شبکه های طبیعی، همه فعالیتهای ذهنی و حرکتـی، بـا اتصال سلولهای عصبی یا نرونها انجام می پـذیرد. سـرعت انتقـال پیـام الکتریکی در طول یک سلول و بین سلولهای عصبی، نسبت بـه سـرعت انتقال در سیستمهای الکترونیک بسیار کمتر است (حداکثر تا چند متر بر ثانیه) و قابلیت پردازش آنها با کامپیوترها قابل مقایسـه نیسـت ولـی موجودات زنده فعالیتهای بسیار پیچیده ای انجام می دهند که بسـیاری از آنها قابل پیاده سازی با فن آوری فعلی نیست. این قابلیت را به تعداد زیاد این سلولها و فعالیت موازی آنها نسبت می دهند.

شبکه های عصبی مصنوعی به منظور استفاده از این مزیتها ایجاد شده اند. هرچند به دلایل گوناگون، شبکه های عصبی مصنوعی از نظر ساختار و عملکرد، از شبکه های طبیعی فاصله گرفته اند [۷] .

۱-۶-۲ انواع شبکه های عصبی

شبکه های عصبی را از چند نظر می توان تقسیم بندی کرد. از نظر ساختمان ، شبکه ها به دو نوع تقسیم می شوند.

۱) Feedforward

۲) Recurrent یا Feed Back

در نوع اول، مسیر انتقال اطلاعات همواره رو به جلو است ولی در نوع دوم، حلقه وجود دارد. شبکه های نوع دوم ممکن است ناپایدار باشند و یادگیری در این شبکه ها، بسیار کندتر است. این شبکه ها در تحقیقات، توجه زیادی به خود جلب کرده اند.ولی در کاربردهای عملی، شبکه های Feedforward سودمندتر بوده اند[۱۱].

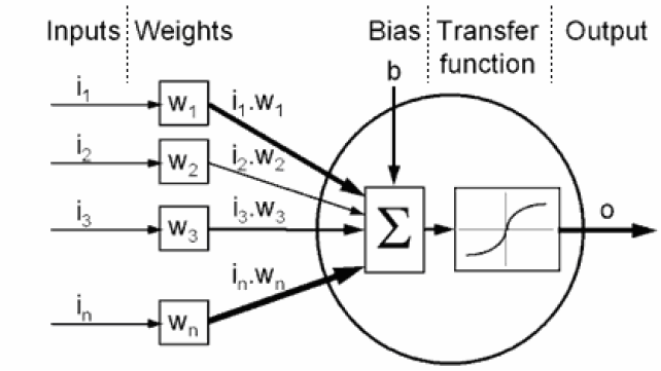
از شبکه هـای نـوع اول مـی تـوان بـه Multilayer Perceptron و Hebbian و از شـبکه هـای نـوع دوم بـه Boltzman و Hopfield اشاره کرد. از نظر نحوه یادگیری شبکه ها بـه دو نـوع Supervised و Unsupervised تقسیم می شوند.

در شبکه های نوع اول، مجموعـه ای از ورودیهـا و خروجیهـای متناظر آنها (خروجیهایی که انتظار داریم شبکه بـه ازا ورودیهـا ی داده شده، آن خروجی را تولید کند) به شبکه داده مـی شـود. آنگـاه توسـط الگوریتمهایی مثل گرادیان کاهشی ، وزن اتصالات طـوری تنظـیم مـی شود که تابع هزینه (معمولا" مجموع مجذورات تفـاوت خروجـی تولیـد شده توسط شبکه و خروجی مورد انتظار یا SSE) مینیموم شود.

در روشهای نوع دوم، اطلاعـاتی از خـارج بـه شـبکه داده نمـی شود و شبکه با استفاده از ویژگیهای مختلف ورودیهـا (مثـل ویژگیهـای آماری) خروجی مـی دهـد. ایـن روشـها بـر خـلاف روشـهای نـوع اول معمولا" در کاربردهای تشخیص و طبقه بندی به کار نمـی رونـد بلکـه

بیشتر برای فشرده سازی و حذف redundancy و یا یـافتن الگوهـای خاص بین داده های ورودی استفاده می شـوند . بحـث در مـورد انـواع شبکه های عصبی بسیار گسترده است و ما به همین مقـدار اکتفـا مـی کنیم. برای کسب اطلاعات بیشـتر در ایـن مـورد مـی توانیـد بـه [۱۰] مراجعه کنید.

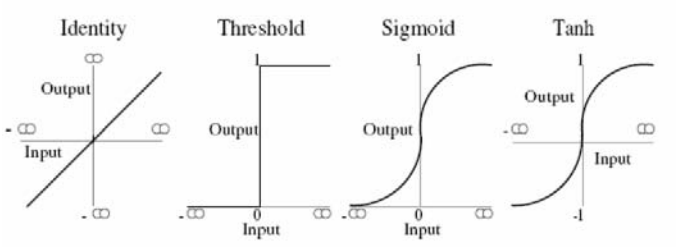
۷-۲: نرون



شکل ۲-۲ یک نرون با تابع فعال ساز

نرون واحد تشکیل دهنده شبکه عصبی است. قسمتهای تشکیل دهنده یک نرون عبارتند از:

- مجموعه ای از ورودیها (i) که می توانند از خارج از شبکه تأمین شوند و یا خروجی سایر نرونها باشند. هر ورودی وزن متناظر داردکه قابل تغییر است (W)
- یک مقدار بایاس یا آستانه که همواره مقدار ثابتی است (معمولا ۱+ وگاهی ۱-) ولی وزن آن قابل تغییر است. بایاس را می توان یک ورودی همواره ثابت فرض کرد.
- تابع ترکیب کننده که مجموع وزن دار ورودیها و بایاس $(\Sigma i_i w_i + b)$ را محاسبه می کند.
- تابع فعا لسازی یا انتقال که بر اساس مقدار خروجی تابع ترکیب، خروجی می دهد. این خروجی همان خروجی نرون است. در شکل ۳-۲ چند نمونه از توابع فعال سازی را مشاهده می کنید. برای مشاهده چند تابع دیگر می توانید به [۹] مراجعه نمایید.



شکل ۳-۲. چند تاج فعال سازی رایج

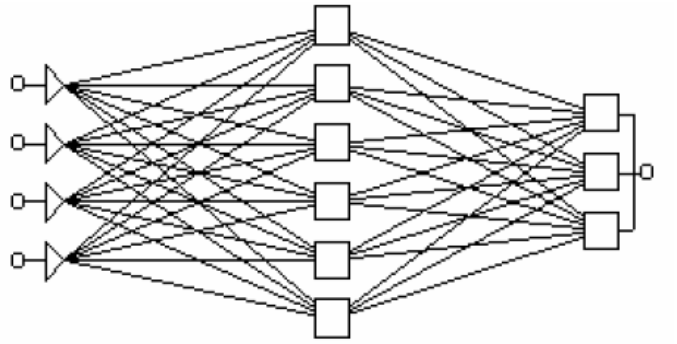
۸-۲. شبکه های Multilayer Perceptron و الگوریتم Backpropagation

Multilayer Perceptron (که از این به بعد به اختصارآن را MLP می نامیم) پرکاربردترین نوع شبکه عصبی است [۱۱]. در این نوع شبکه، نرونها به صورت لایه های متوالی در یک شبکه Feedforward قرار دارند بطوریکه هر نرونی به همه نرونهای لایه بعد متصل است ولی بین نرونهای یک لایه، ارتباطی وجود ندارد.

(شکل ۴-۲). لایه اول، لایه ورودی است. تعداد نرونهای این لایه، برابر با تعداد ورودیها (به عبارتی دیگر، تعداد درایه های بردار ورودی) است.

لایه آخر لایه خروجی نام دارد و تعداد نرونهای آن با توجه به خروجی مورد انتظار تعیین می شود. مثلا" در پیاده سازی توابع، یک نرون کافی است ولی در مسائل طبقه بندی، تعداد آنها برابر با تعداد کلاسهای خروجی است. لایه های بین این دو لایه، لایه های مخفی نام دارند.

می توان ثابت کرد که تمام مسائل قابل حل با شبکه های MLP ، بایک لایه مخفی قابل پیاده سازی هستند هرچند وجود دو لایه مخفی می تواند از جهاتی سودمند باشد [۷].تعداد نرونها در لایه های مخفی برخلاف دو لایه دیگر به سادگی قابل تعیین نیست بلکه باید با روش سعی و خطا و یا استفاده از تجربیات دیگران تعیین گردد [۴].



شکل ۴-۲ Multilayer Perceptron

فرض کنید شبکه دارای n اتصال وزندار باشد (شامل بایاسها). تابع خطای SSE را می توان به صورت سطحی در فضای n +۱ بعدی در نظر گرفت. در هر لحظه با توجه به مقادیر وزنها، شبکه را می توان با نقطه ای در این فضا نشان داد. هدف از آموزش شبکه، هدایت این نقطه به سمت نقطه مینیموم سطح است. مشخص است که اگر در جهت عکس بردار گرادیان حرکت کنیم به سمت مینیموم خواهیم رفت (روش گرادیان کاهشی). البته ممکن است این نقطه مینیموم محلی و نه مینیموم کلی باشد. این مشکل، مشکل اصلی روش گرادیان کاهشی

است که ما در ادامه راه هایی را برای کم کردن احتمال بروز این مشکل ارائه خواهیم داد. اصولا" یافتن ساختار بهینه و پارامترهای یادگیری شبکه با این مشکل مواجه هستند.

شبکه های MLP را غالبا" با الگوریتم Backpropagation که مبتنی بر روش گرادیان کاهشی است آموزش می دهند. در این روش به طورiterative با اعمال ورودیها به شبکه و بدست آوردن تابع خطا، وزنها را اصلاح کرده و مجددا" ورودیها را اعمال می کنیم و این کار را تا محقق شدن یکی از شرایط توقف ادامه می دهیم. مراحل این الگوریتم عبارتند از [۴]:

۱) وزنها را برابر با مقادیر تصادفی قرار می دهیم. این کار علاوه بر کمک به همگرایی، باعث می شود تا الگو ریتم درهر بار اجرا شدن، از وضعیت جدیدی شروع شود و لذا در صورت ناموفق بودن اجرای آن (بخصوص در صورت افتادن در مینیموم محلی) امکان موفقیت در اجرای بعدی وجود داشته باشد.

۲) اولین بردار ورودی را به شبکه اعمال می کنیم (آغاز یک گذرالگوریتم).

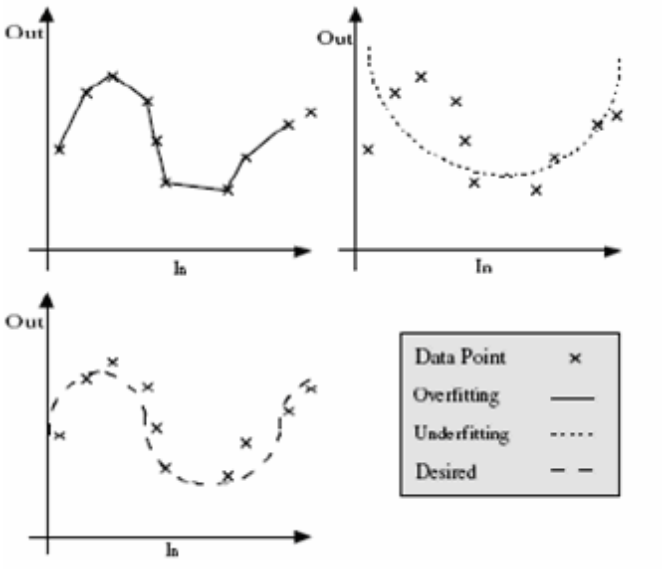
۳) بر اساس ورودیها، خروجیهای اولین لایه مخفی را محاسبه می کنیم. خروجیهای لایه دوم را بر اساس خروجیهای لایه اول محاسبه کرده و این کار را ادامه می دهیم تا خروجیهای شبکه (O ها) بدست آیند.

۹-۲. Overlearning

در بخش قبلی با الگوریتمی برای کاهش خطای شبکه آشنا شدیم. نکته قابل توجه این است که این الگوریتم خطا را روی مجموعه داده های یادگیری مینیموم می کند نه بر روی داده های آزمایش که مجموعه ای متفاوت هستند (معمولا" تعدادشان ناشناخته است و ما دقیقا" نمی دانیم چه ورودیهایی به شبکه اعمال خواهند شد). در نتیجه ممکن است شبکه به جای اینکه با استفاده از داده های یادگیری، تابع کلی را که ورودی و خروجی مورد انتظار را به هم مربوط

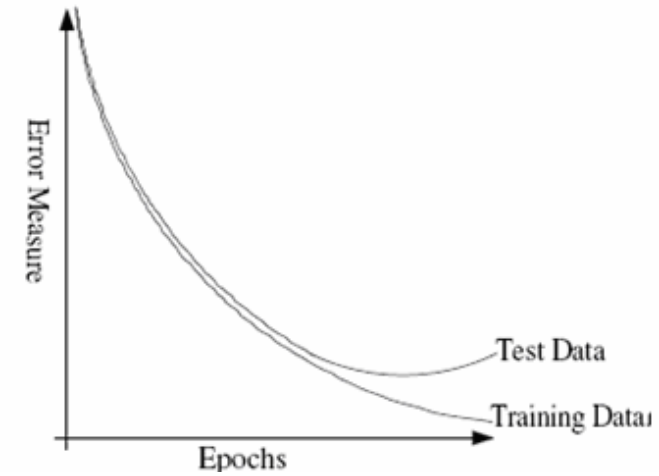
می کند تخمین بزند، تنها تابع مرحله یادگیری را پیاده سازی کند. هر چند این کار سبب کاهش خطای یادگیری می شود اما خطای آزمایش را که برای ما مهم تر است بیشتر می کند. به این پدیده Overlearning می گویند.

برای درک بهتر این پدیده می توان پدیده مشابهی (Overfitting) را در مورد تخمین چندجمله ای بر اساس چند نقطه داده شده بررسی کرد (شکل۵-۲). در صورت overfit شدن، هرچند چندجمله ای پیچیده تر، با خطای کمتری نقاط را در بر گرفته ولی این چندجمله ای با چندجمله ای مطلوب، متفاوت است



شکل ۵-۲ overfitting چند جمله ای

هرچه تعداد نرونهای شبکه و تعـداد گـذرهای یـادگیری بیشـتر شـود، خطای آموزش کمتر می شود ولی امکان overlearning و همچنـین زمان یادگیری افزایش می یابد (شکل ۶-۲).



شکل ۶-۲ overlearningدر شبکه عصبی

برای حل این مشکل، مجموعه ای از داده ها به نام داده های اعتباررا در نظر می گیریم.

ایـن داده هامتفـاوت از داده هـای آمـوزش بـوده و در فراینـد تنظیم وزنهای شبکه مورد استفاده قرار نمی گیرند. در هر گذر الگوریتم یادگیری داده های اعتبار را به شـبکه اعمـال کـرده وخطـای اعتبـار را بدست می آوریم. اگر این خطا هم مثل خطای یـادگیری کـاهش پیـدا

کـــرده بـــود، الگـــوریتم را ادامـــه مـــی دهـــیم. در غیـــر اینصـــورت overlearning رخ داده و یادگیری را متوقف می کنـیم. ایـن حالـت، نشان دهنده قدرت بیش از حد شبکه است. پـس تعـداد نرونهـای لایـه مخفی را کاهش می دهیم و یـادگیری را از نـو شـروع مـی کنـیم. اگـر الگوریتم بـدون رخ دادنoverlearning و بـدون رسـیدن بـه خطـای یادگیری مطلوب به پایان رسید، تعداد نرونهای لایـه مخفـی را افـزایش می دهیم [۱۱].

در این فصل با شـبکه هـای عصـبی بـه عنـوان یـک راه بـرای شناسایی حـروف آشـنا شـدیم. شـبکه هـای عصـبی MLF را کـه در کاربردهای تشخیص الگو رایج هستند معرفی کـردیم و در مـورد نحـوه یادگیری آنها توضیح دادیم. با اتمام این فصل، بخـش مقـدمات نیـز بـه پایان می رسد. در بخش بعد به کارهای انجام شده در ایـن پایـان نامـه می پردازیم

فصل ۴:

شرح یک پروژه انجام شده

۴-۱:شرح پروژه

این فصل شرح مختصری به یک نمونه کار انجام شده توسط آقای امید یگانه (مرجع شماره [1]) را دارد.

عملیات انجام شده در این پروژه در دو مرحله صورت می گیرد. مرحله اول شامل پردازش تصاویر و تهیه ورودی (فایل ویژگی) برای شبکه عصبی و مرحله دوم شامل ساختن و آموزش شبکه است.

در این فصل مرحله اول یعنی تهیه ورودی شبکه عصبی را بررسی می کنیم. در ابتدا در مورد مشخصات فایلهای تصویری مورد استفاده و سپس در مورد نحوه تبدیل تصاویر به بردار ویژگی و قراردادن بردارها در فایل ویژگی (به عنوان ورودی مرحله دوم) توضیح می دهیم.

۴-۲. تصاویر مورد استفاده

فایلهای تصویر استفاده شده در این پروژه ، برگرفته از بانک تصاویر حدف ۸۴ هستند. این بانک تصاویر، توسط شرکت اندیشه نرم افزار پایا و به سفارش شورای اطلاع رسانی کشور تهیه شده است. این

حروف دست نویس فارسی بانک شامل دویست هـزار تصـویر از حـروف مجزای دست نویس فارسی است که از فرم ثبت نام داوطلبـان ورود بـه مراکز استعدادهای درخشان (سمپاد) در مقاطع راهنمـایی و دبیرسـتان استخراج شده اند.

مشخصات فایلهای تصویری استفاده شده به شرح زیراست :

- Grayscale bitmap (bmp) .
- ۳۰۰ dpi .
- 77×95 pixel

۳-۴. تبدیل تصاویر به بردارهای ویژگی

ابتـدا تصـاویر خاکسـتری، سـیاه و سـفید مـی شـود. سـپس Bounding Box حروف موجود در تصویر (کوچکترین مسـتطیل در بردارنده حرف) پیدا شده و تصویر بریده می شود (شکل ۱-۳).

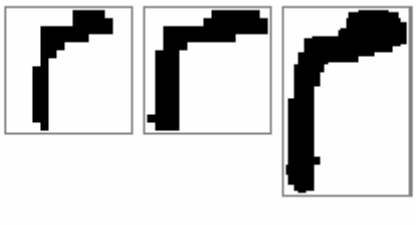
شکل ۱-۳. یافتن Bounding Box و بریدن تصویر

الف: تصویر اولیه ب: تصویر سیاه و سفید پس از بریده شدن

ابعاد تصویر بریده شده را با دو روش به اندازه ۱۵*۱۵ تبدیل می کنیم:

- روش اول، تغییر ابعاد بدون حفظ نسبت آنها است. این روش، همان روش استاندارد تغییر ابعاد است. با این روش حرف موجود در تصویر بریده شده، تمامی فضای تصویر جدید را پر می کند ولی چون ابعاد تصویر بریده شده یکسان نیستند، یک بعد بیش از بعد دیگر تغییر می یابد و این سبب تغییر شکل حرف می شود (شکل ۲-۳).
- روش دوم، تغییر ابعاد با حفظ نسبت آنها است. با توجه به اینکه بعضی از حروف (مثل " ق " و " ت ") در اثر تغییر ابعاد به روش اول ممکن است شبیه به هم شوند، یک بار نیز با حفظ نسبت ابعاد تصویر را

تغییر می دهیم [۳] . در این روش هر دو بعد تصویر را به یک نسبت تغییر می دهیم بطوریکه بعد بزرگتر برابر با ۱۵ و بعد دیگر کمتر از ۱۵ شود (مثلا" اگر ابعاد تصویر بریده شده،h و w باشند که h< w ، تصویر جدید دارای ابعاد ۱۵ ۱۵ * w/h است پس با اضافه کردن متقارن سطرها یا ستونهای سفید به بعد کوچکتر، آن را به ۱۵ نقطه می رسانیم بطوریکه حرف در وسط تصویر باشد (شکل ۲-۳).



شکل ۲-۳. روشهای تغییر اندازه

الف: تصویراولیه ب: بدون حفظ نسبت ج: با حفظ نسبت

در انتها تصویر تولید شده در مرحله اول را روی تصویر مرحله دوم می گذاریم تا تصویری به ابعاد ۳۰ * ۱۵ تشکیل شود و این تصودر را به

برداری با ۴۵۰ عضو که مقدار هر عضو صفر یـا یـک اسـت تبـدیل مـی کنیم. همانطور که قبلا" هم اشاره کردیم، مـا از روشـهای پیچیـده تـر استخراج ویژگی استفاده نمی کنیم و به تغییر اندازه اکتفا می نماییم.

برای آموزش شبکه عصبی، باید به طریقـی بـه شـبکه عصـبی اطلاع دهیم که هر بردار ورودی نشان دهنده چه حرفی اسـت. هرچنـد در مرحله آزمایش در یـک سیسـتم واقعـی، از قبـل نمـی دانـیم بـردار ورودی نشان دهنده چه حرفی است (اگـر مـی دانسـتیم احتیـاجی بـه OCR نداشتیم) ولی در این پروژه باید این اطلاعات را داشته باشـیم تا بتوانیم درصد تشـخیص صـحیح شـبکه را در صـورت همگـرا شـدن بدست آوریم. درهر صورت، باید جواب صـحیح مـورد انتظـار را هـم بـه شبکه بد هیم. پس عضو ۴۵۱- ام را به بردار ورودی اضافه کنیم. مقـدار این عضو، برابر کـد متنـاظر بـا حـرف مـورد انتظـار اسـت. حـال همـه بردارهای تشکیل شده را (هر بردار برای یک تصویر) در یک فایل متنی

(txt .) می نویسیم. به این ترتیب فایل ویژگی که ورودی مرحله بعد است، ایجاد می شود.

در این فصل مرحله اول عملیات پروژه را توضیح دادیم و نحوه ایجاد فایل ویژگی را بررسی کردیم. در وحله بعد مرحله دوم را توضیح داده و شبکه ای عصبی برای تشخیص حروف ساخته و سعی می کنیم آن را آموزش دهیم.

۴-۴:آموزش شبکه عصبی

در این فصل درباره مرحله دوم تشخیص حروف، یعنی آموزش شبکه عصبی با ورودیهای تولید شده درمرحله اول، توضیح می دهیم. در بخش نخست، طراحی ابتدایی شبکه را بررسی کرده و خواهیم دید که این شبکه همگرا نخواهد شد. در بخش دوم، تغییرات و ساده سازهایی را که به منظور رفع مشکل عدم همگرایی شبکه انجام دادیم، بیان خواهیم کرد (هرچند که این کارها نیز نتوانستند مشکل را برطرف

سازند). با توجه به عـدم همگرایـی شـبکه در مرحلـه آمـوزش، مرحلـه آزمایش انجام نشده است.

۵-۴. طراحی ابتدایی

شبکه به صورت MPL سه لایه (ورودی، خروجی و یـک لایـه مخفی) پیاده سازی شده است. تعداد نرونهـا در لایـه ورودی ۴۵۰ (بـه تعداد عناصر بردار ویژگی)، در لایه مخفـی ۵۰ و در لایـه خروجـی ۳۲ عدد (به تعداد حروف فارسی) است.

تابع فعال سازی همه نرونهای لایه مخفـی و لایـه خروجـی بـه این صـورت اسـت. (ایـن تـابع در نـرم افـزار MATLAB بـا پـارامتر logsig معرفی می شود). تابع فعال سازی نرونهای ورودی، تابع خطـی همانی است زیرا می خواهیم خروجی آنها، حاصل جمح وزندار ورودیهـا باشد.در حالت ایده آل، می خواهیم با اعمـال هـر حـرف ورودی، نـرون خروجی متناظر با آن، مقدار یک و بقیه نرونها مقدار صفر داشته باشند.

مثلا" اگر حرف ورودی "ج " بود، خروجی نـرون ششـم لایـه خروجـی برابر با یک و بقیه خروجیها صفر باشند.

با اعمال این ورودیها به شبکه، خروجیهای لایه ورودی به لایـه مخفی و خروجیهای این لایه به لایه خروجـی مـی رود. خروجـی هـای شبکه مقداری حقیقی بین صفر و یک دارند (مقدار تـابع logsig بـین صفر و یک است). با توجه به اینکه کـدام خروجـی بیشـترین مقـدار را دارد، حروف ورودی تشخیص داده می شود (مثلا" اگـر خروجـی نـرون ششم لایه خروجی بیش از مقدار خروجی سایر نرونهای این لایـه بـود، حرف ورودی "ج " تشخیص داده می شود).

بـرای آمـوزش شـبکه از الگـوریتم Backpropagation بـا learning rate دینامیک و بهره گیری از ممان استفاده کرده ایم (این الگوریتم در نرم افزار MATLAB با پارامتر traingdx مشخص مـی شود). سایر پارامترهای الگوریتم یادگیری عبارتند از:

- ثابت ممان (α) برابر با ۹۵/.

- تعداد گذارهای الگوریتم (epochs) برابر با ۵۰۰۰
- تابع خطای SSE کمتر از ۲٪

مقادیر عددی فوق بدون بررسی گسترده و صرفآ با توجه به استفاده از این مقادیر در پروژه ای دیگر(مثلا" [۳]) انتخاب آزمایش کردیم.

باید توجه داشت که نمی توان مقادیر SSE و epochs را به میزان زیادی تغییر داد زیرا افزایش تعداد گذرها سبب طولانی تر شدن زمان یادگیری می شود و افزایش مقدار خطای مجاز، هرچند ممکن است به همگرایی شبکه کمک کند ولی سبب افزایش درصد تشخیص اشتباه خواهد گشت.

برای آموزش شبکه دو نکته را در نظر گرفته ایم:

- آموزش، تدریجی انجام می شود. یعنی ابتدا الگوریتم را با ۲۰ بردار ورودی اجرا می کنیم. پس از اتمام یادگیری، ۲۰ بردار دیگر را به ۲۰

بردار قبلی اضافه می کنیم و مجددا" الگوریتم را با این ۴۰ ورودی اجرا می نماییم وهر بار چند بردار به مجموعه قبلی می افزاییم تا بردارهای ورودی به تعداد قبل توجهی برسند. سپس کل بردارها را اعمال می کنیم. تفاوت آموزش تدریجی با روش عادی این است که در این نوع آموزش پس از پردازش تعداد کمتری از ورودیها، وزنها تغییر می یابند (گذرها کوتاه تر هستند). این مسأله بخصوص در مراحل اولیه الگوریتم که وزنها مقدار تصادفی دارند مفید است زیرا در زمان کوتاهی، وزنها به مقدارنهایی و مطلوب خود نزدیک می شوند.

در این پروژه ، پس از اینکه تعداد ورودیها به ۱۴۰ می رسید، شکل نمی توانست همگرا شود و حتی پس از اعمال تغییرات و ساده سازیها نتوانستیم این عدد را بیش از ۲۴۰ کنیم. این مشکل نشانگر آن است که بردارهای ورودی مناسبی به شبکه اعمال نکرده ایم. یعنی بین اعضای دو کلاس یکسان (مثلا" دو حرف یکسان که توسط دو فرد نوشته شده اند) تفاوت زیادی وجود دارد ولی گاهی بین اعضای دو

کلاس متفاوت شباهت وجود دارد (مثلا" دوحـرف " ن " و " ق " کـه توسط یک نفر نوشته شده اند).

- بردارهای ورودی به ترتیب تصادفی بـه شـبکه اعمـال مـی شـوند. یعنی اینطور نیست که ابتدا "الف "ها و سـپس "ب "هـا اعمـال شـوند بلکه یک جایگشت تصادفی از حروف اعمال می شود.مـثلا" ابتـدا یـک "ج " و سپس سه " ب " و بعد یک " ل ". این کار دو علت دارد:

علت اول: این است که وقتی تعداد کمـی از ورودیهـا را اعمـال می کنیم (آموزش تدریجی) نباید همه از یک حرف باشند.

علت دیگر: این است که اگر ابتدا "الف "ها و سـپس "ب "هـا پردازش شوند، شبکه به سمت "الـف "هـا تمایـل بیشـتری دارد یعنـی خطای "الف "ها کمتر از خطای "ب "ها می شود.

۶-۴. تغییرات و ساده سازیها

برای رفع مشکل عدم همگرایی، تعداد نرونهای لایـه مخفـی را در چند مرحله افزایش دادیم تا به ۱۵۰ نرون رسید. در این حالـت نیـز

شبکه همگرا نشد. البته با توجه به تعـداد نرونهـای ورودی(۴۵۰) ،۱۵۰ تعداد کمی است ولی از آنجا که با افزایش نرونهای لایه مخفی، زمـان و حافظه لازم برای یادگیری به شدت افزایش می یابد، این تعداد را بـیش از این افزایش ندادیم .

اقدام دیگر، کم کردن تعداد کلاسهای خروجی با حذف حـروف مشابه بود. مـثلا" از بـین حـروف "ب "، "پ " ، " ت " و "ث "، تنهـا حرف "ب " به عنوان ورودی اعمـال شـد و بردارهـای وروی بقیـه ایـن حروف، حذف گردیدند. با اینکار تفاوت بین کلاسهای مختلـف افـزایش یافته و لذا امکان همگرایی شبکه بیشتر می گشت. در صـورت موفقیـت این کار، می توانستیم همانند کار انجام شده در [۷]، با حذف نقطه ها و بنابراین لازم است تعداد نرونهای ورودی(به عبارت دیگر، تعداد اعضـای بردار ویژگی) کاهش یابد. استخراج شـکل اصـلی حـروف (کـه معمـولاً بزرگترین مولفه همبندی شکل حرف است) کار تشخیص توسط شـبکه عصبی را آسان تر سازیم.

همانطور که قبلا" هم گفتیم، این تغییرات تنها سبب افـزادش تعداد بردارها در آموزش تدریجی از ۱۴۰ به ۲۴۰ شد و نتوانست شبکه را همگرا کند.

هم زمان با اعمال دو تغییـر فـوق، در مقـدار ثابـت ممـان هـم تغییراتی صورت دادیم ولی ایـن کـار تـأثیر خاصـی در عملکـرد شـبکه نگذاشت. همچنین با توجه به اینکه مقدار خطـای SSE بدسـت آمـده بسیار بیشتر از ۲٪ بود، مشخص بود که افزایش خطای مجاز نمی توانـد به همگرایی شبکه کمک کند.

در این فصل نحوه آموزش شبکه و تغییرات اعمـال شـده بـرای حل مشکل عمدم همگرایی آن را بررسی کـردیم. بـا توجـه بـه نـاموفق بودن این کارها به این نتیجه رسیدیم که مشـــکل اصـــلی، در روش استخراج ویژگی است و باید از الگوریتم پیچیـده تـری اسـتفاده کنـیم. درفصل آخر بـه بیـان نتـایج حاصـل از انجـام ایـن پـروژه پرداختـه و پیشنهاداتی را برای رفع مشکل عدم همگرایی ارائه می نماییم.

فـصل ۵ :

ویژگی های متون فارسی

۱–۵.خصوصیات متون فارسی

متون فارسی دارای خصوصیات متفاوتی نسبت به متون لاتین هستند.

۱) جهت نوشتاری این حروف از سمت راست به چپ است، در حالی که اعداد در این زبان از سمت چپ به راست نوشته می شوند.

۲) بیشتر حروف از خطوط منحنی شکل تشکیل می شوند و یا دارای حلقه هایی هستند که اغلب در جهت عقربه های ساعت بسته می شوند.

۳) کلمات این زبان به صورت سرهم نوشته می شوند. این خصوصیت یکی از مشکل سازترین عوامل در شناسایی این حروف است، زیرا جداسازی حروف سرهم خیلی مشکل است. این همان مشکلی است که در شناخت متون دست نویس لاتین هم وجود دارد.

۴) بیشتر حروف فارسی دارای اشکال متفاوت بسته به جایگاهشان در کلمه می باشند.

۵) بسیاری از حروف فارسی شبیه به هم بوده و تفاوت آنها فقط در تعداد و مکان نقطه هایشان است (مثلا"ب، پ، ت، ث)

۶) در مورد حروف سرهم، همپوشانی عمودی بعضی از حروف فارسی نیز مسأله ای است که شناخت این حروف را با مشکل روبرو می نمایند.

این خصوصیات باعث می شود که برای بازشناسی متون فارسی از روشهایی متفاوت با روشهای متداولی که برای شناسایی متون لاتین استفاده می شوند، استفاده نماییم.

یک سیستم OCR فارسی در حالت کلی باید بتواند صفحات را اسکن کرده، تصاویر قسمتهای مربوط به متن را از تصویر صفحه اسکن شده استخراج و تصویر متن را به تصاویر حروف تفکیک نماید. آنگاه هر حرف را شناسایی (تشخیص حرف از روی تصویر آن) و سپس حروف شناسایی شده را به نحو مناسبی کنارهم چیده و خروجی مطلوب را ارائه دهد.

جدا سازی حروف پیوسته کـار بسـیار مشـکلی اسـت و گرچـه پیشرفتهایی در این زمینه انجام شده ،هنوز سیستمی که بتواند با دقـت قابل اطمینان این عملیات را انجام دهد ارائه نشده است. لذا در کارهایی کـه اسـتفاده از OCR الزامـی اسـت (مـثلا" جمـح آوری مشخصـات داوطلبین در آزمونها)، متون را به صورت دنباله ای از حروف مجزا مـی نویسیم.

ورودی این سیستم، مجموعه ای از تصاویر است که هـر کـدام یک حرف دست نویس در بردارد. این تصاویر پس از پردازش مختصـری (تبدیل به تصاویر سیاه و سـفید و تغییـر انـدازه) بـرای آمـوزش شـبکه عصبی بکار می روند. چنین پردازشی برای تشخیص حروف تایپی کافی است ولی در مورد حروف دست نویس خواهیم دید کـه شـبکه علیـرغم تلاشها و ساده سازیهای انجام شده، همگرا نمـی شـود. نتیجـه خـواهیم گرفت که در مورد حروف دست نویس، باید پردازش پیچیده تـری روی تصاویر انجام شود.

فـصل ۶ :

برسی چند محصول OCR

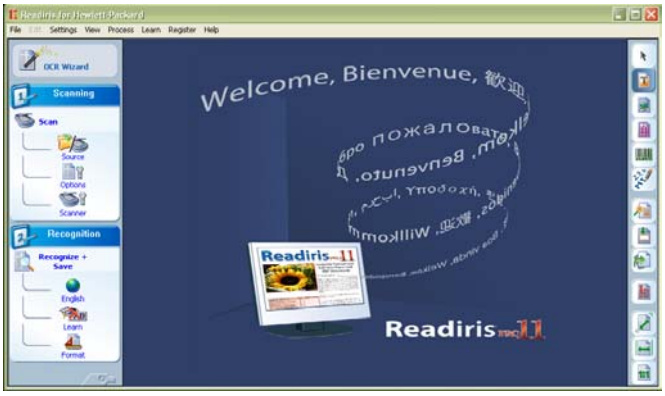
۱-۶چند نمونه OCR

از نمونه نرم افزار هایی که در این زمینه وارد بازار شده اند مـی توان به Readiris Pro 11 و VajehShenas V.1 اشاره کرد کـه VajehShenasV.1از نوع وطنی آن که به دست متخصصـین ایرانـی تولید شده است که در این فصل بر آن داریم این دو نرم افزار را آموزش دهیم

۲-۶ آموزش Readiris Pro 11

Readiris Pro 11 این نرم افزار قابلیـت شناسـایی اکثر زبان ها را دارا می باشد اما قادر به شناسایی زبـان فارسـی نیسـت. این نرم افزار قادر به شناسایی انـواع عکـس و PDF Word,Excel, Emailو .....

هنگامی که روی آیکون آن دبل کیک می کنیم شکل ۱-۵ را می بینـیم .

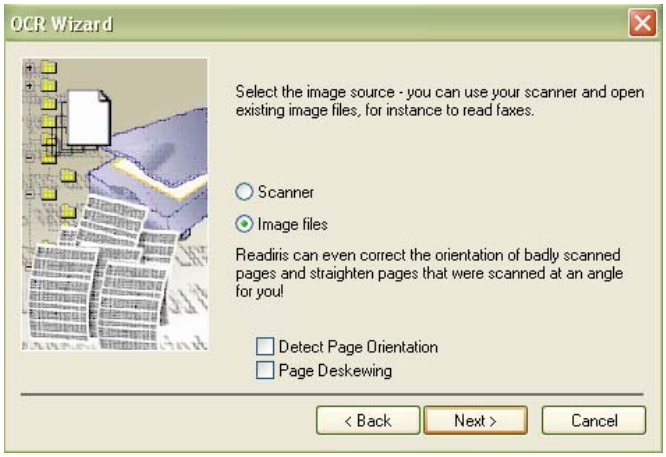


شکل ۱-۵

که به دو صورت wizard و دستی می توان با آن کار کرد:

۱) برای کار کردن بـه صـورت ویـزاردی روی مـتن ocr wizard در قسمت سمت چپ بالای کادر کلیک کرده پنجرهای باز میشود کـه بـر روی دکمه next کلیک می کنـیم .در صـفحه بعـد (شـکل ۲-۵) از مـا منبع عکس را می خواهد که اگر عکس مورد نظر ما روی هارد قرار دارد

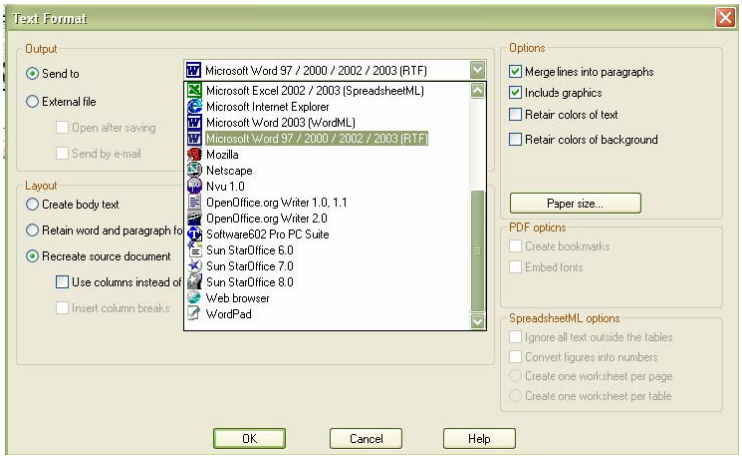
روی file image کلیـــک کـــرده در غیـــر اینصـــورت روی



شکل ۲-۵

گزینه scanner کلیک می کنیم.سپس next را کلیک می کنیم.

در صفحه بعد زبان نوشته را می پرسد.که در صورت دلخواه می توان آن را با دکمه change به راحتی عوض نمود.سپس با کلیـک بـر روی next به صفحه ای می رویم (شکل ۳-۵) که فرمت خروجـی را از ما می پرسد که به دلخواه یکی را انتخاب مینـیم ترجیحـا" بهتر است که microsoft word را انتخاب کنیم



شکل ۳-۵

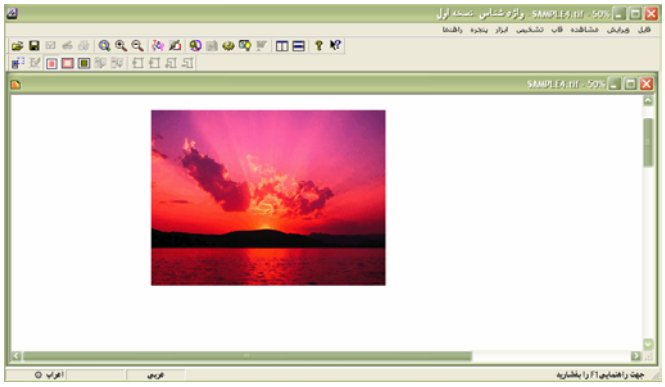
سپس با کلیک کردن بر روی next به پایان عملیات می رسیم .فقط در آخرین مرحله صفحه ای باز می شود که فایل عکس را انتخاب می کنیم که به صورت خودکار ocr فایل را مورد پردازش قرار می دهد سپس فقط نتیجه نهایی را در محل دلخواه کاربر ذخیره میکند.

۲) اما به صورت دستی این گونه عمل می کنیم که از منوی فایل گزینه open را انتخاب کرده سپس عکس مورد نظر را انتخاب می کنیم.سپس (با توجه به عکس۱-۵) بر روی گزینه save recognize

واقع در سمت چپ پایین صفحه را انتخاب می کنیم و اگر می خـواهیم فرمت خروجی یا زبان را تغییـر دهـیم در پـایین منـوی recognize save به راحتی قابل تغییر است. سپس محلی کـه مـی خـواهیم مـتن عکس ذخیره شود را انتخاب کرده و دیگر کار تمام است

۳-۶ آموزش VajehShenas V.1

VajehShenas V.1 این نرم افزار کـاملا" ایرانـی توسـط شرکت هوش مصنوعی رایورز تهیه شده است که بر خلاف Readiris Pro 11 کـه فرمـت هـای متنـوعی را بـه عنـوا ورودی مـی پـذیرد VajehShenas V.1 فقط قادر به شناسایی بعضی از فرم های عکس مانند BMP TIF,PCX,ART را دارا می باشد.



شکل ۴-۵

۱) برای WIZARD باید مراحل زیر را انجام دهیم ابتدا به منوی ابزار رفته و گزینه ی را هبر اتو ماتیک را انتخاب می کنیم که از پنجره باز شده نوع عکس را مشخص می کنیم (شکل ۵-۵) که از نوع ورد به کامپیو تر از طریق اسکن یا از نوع فایلی که درون هارد قرار دارد.

سپس به دلخواه یکی را انتخاب و به صفحه بعدی می رویم و در این صفحه محل عکس را مشخص می کنیم.سپس گزینه دکمه بعدی را می زنیم در صفحه جدید که مهمترین صفحه می باشد از ما می پرسد که این عکس دارای چه زبانی می باشد مثلا" فارسی یا

انگلیسی و.... در صورتی که متن عکس با گزینه پـیش فـرض مغـایرت دارد حتما" باید آن را تغییر دهیم(شکل ۶-۵).

شکل ۵-۵

شکل ۶-۵

با زدن دکمه بعدی به صفحه جدیدی می رویـم کـه مربـوط بـه دوران عکس است که ما با آن کاری نداریم.سـپس دکمـه بعـدی را زده که به آخرین مرحله که محل ذخیـره سـازی سـند و فرمـت آن را مـی پرسد کار دیگر تمام است(شکل ۷-۵)

شکل۷-۵

۲) روش دستی این نرم افزار این گونه است. که از منوی فایل گزینه باز کردن را انتخاب کرده و عکس مورد نظر را انتخاب کرده و سپس از منوی قاب ، قاب بندی اتوماتیک را انتخاب کرده سپس به منوی تشخیص انتخاب موتور تشخیص رفته و زبان عکس که بسیار مهم است را با دقت انتخاب کرده و در آخر گزینه شروع از منوی تشخیص را سمهم این نرم افزار آن است که عکس با رزولو یشن ۳۰۰dpi سیاه و سفید و با فرمت TIF گرفته شود و همین طور عمق رنگ آن ۱ باشد در غیر این صورت با مشکلات بسیاری مواجه هستین

۴–۵مقایسه VajehShenas با Readiris

در مقایسه ردریز و واژ شناس ذکر موارد زیر الزامی است:

- ردریز نرم افزاری بسیار تواناست که از فرمت های ورودی بیشتری نسبت به واژه شناس بهره می برد مانندpdf و..

- تعداد حروف شناسایی نشده توسط ردریز کمتـر از واژه شـناس در زبان انگلیسی است.
- واژه شناس بعد از شناسایی حروف دارای یک نرم افـزار غلـط یـاب حرفه ای ا ست که ردریز از آن بی بهره است
- ردریز دارای ظا هری شکیل تر نسبت به واژه شناس است

فـصل ۷ :

نتایج و پیشنهادات

در فصول گذشته دیدیم که شبکه عصبی طراحی شده، همگـرا نشد. با توجه به عملگرد موفق شبکه های عصبی در این پایان نامـه هـا مشابه از جمله [۳] و [۸] و بخصوص [۷] که کاری مشابه با این پایان نامه را بررسی کرده، انتخاب روش شبکه عصبی برای تشخیص حروف و همین طـور انتخـاب نـوع شـبکه (MLP) انتخـاب مناسـبی اسـت. بـا

توضیحاتی که در فصل های قبل ارائه شد، می توان نتیجه گرفت که روش مورد استفاده برای استخراج ویژگی از تصاویر ورودی، روش مناسبی نیست. برای این منظور باید روش دیگری در نظر گرفت که دارای سه ویژگی زیر باشد:

۱) الگوریتم باید بتواند برای حروف مختلف، خروجیهایی با تفاوت قابل توجه بیرون دهد. این امر بخصوص در مورد حروف مشابه (مثل "ب " و " پ ") مهم است.

۲) الگوریتم باید بتواند برای نمونه های مختلف از یک حرف (مثلا" دست خطهای مختلف از حرف "ج ") خروجی های مشابه بدهد.

۳) تعداد عناصر خروجی الگوریتم (که اعضای بردار ویژگی هستند) باید تا حد ممکن کم باشد

زیراهرچه تعداد بیشتر باشد، تعداد نرونهای لازم در لایه مخفی نیز بیشترخواهد شد و افزایش تعداد نرونها و اتصالات، زمان و حافظه یادگیری و آزمایش نیز بیشتر خواهد شد. یک راه ساده و موثر برای

تحقق این شرط، حذف redundancy بین اعضای بردار ورودی با الگوریتمهایی مثل PCA است. یک ایده خوب برای کمک به تحقق دو شرط اول، برداشتن نقطه ها و سرکش حروف ورودی است.

ابتدا با پردازش تصاویر ورودی، بزرگترین مولفه همبندی تصویر را انتخاب کرده و با اعمال روشهای استخراج ویژگی و سپس بکارگیری شبکه عصبی، فرا- کلاس (مجموعه ای از چند حرف مشابه مثلا" فرا-کلاس "ک -گ") متناظر با حرف ورودی را شناسایی می کنیم. سپس با توجه به شکل و مکان قرارگیری نقطه ها و سرکش، حرف ورودی را تعیین می کنیم. یک ایده خوب برای این مرحله، استفاده از یک شبکه عصبی دیگر است. به این صورت که یک شبکه عصبی اصلی برای شناسایی شکل اصلی حرف قرار می دهیم. درهر فرا-کلاس خروجی نیز یک شبکه عصبی دیگر قرار می دهیم. تصویری که نقطه ها و سرکش آن حذف شده اند به شبکه عصبی اصلی وارد شده و فرا کلاس آنها شناسایی می شود.

سپس تصویر اولیه و یا خروجی الگوریتمی که اطلاعـات نقطـه ها سرکش تصویر اصلی را دربر دارد، به شبکه عصبی مخصوص این فرا- کلاس وارد شده و حرف مورد نظر شناسایی می شود.

آنچه در تبدیل کلاسها به فرا- کلاسها مشکل ساز است، امکان متصل یا جدا بودن نقطه ها و سرکش دردست خطهـای مختلـف اسـت که باید روشی برای شناسایی این نوع نقطه ها پیدا کرد.

مراجع

[۱] امید یگانه، مونا. " جداسازی حروف تایپی مسـتقل از نـوع فونـت و اندازه ". پایان نامه کارشناسی ارشد،
دانشکده مهندسی برق، دانشگاه صنعتی شریف، ۱۳۸۳.

[۲] اندیشه نرم افزار پایا، شرکت. "راهنمای بانک اطلاعاتی حـدف۸۴ "

۱۳۸۴

[۳] بهنام قادر، احسان." تشخیص کاراکترهای پلاکهای خودرو". پایـان

نامه کارشناسی، دانشکده مهندسی

برق، دانشگاه صنعتی شریف، ۱۳۸۳.

[4] Beeman, Dave. "Multi-layer Perceptrons (Feed-forward Nets), Gradient Descent, and Back Propagation". University of Colorado, 2001.
http://ece-www.colorado.edu/~ecen4831/lectures/NNet3.html

[5] Burrow, Peter. "Arabic Handwriting Recognition". Master of Science Thesis, School of Informatics, University of Edinburgh, 2004.

[6] Gonzalez, Rafael C., Woods, Richard E. "Digital Image Processing". Second Edition, New Jersey: Prentice-Hall, 2002.

[7] Johnston, Alexander. "Classifying Persian Characters with Artificial Neural Networks and Inverted Complex Zernike Moments". 2005.
http://www.doc.ic.ac.uk/teaching/projects/Distinguished05/AlexJohnston.pdf

[8] Klassen, Tim. "Towards Neural Network Recognition of Handwritten Arabic Letters". A Project Submitted to the Faculty of Computer Science in Partial Fulfillment of the Requirements for the Degree of Master of Computer Science, Dalhousie University, 2001.

[9] MathWorks Inc. "MATLAB Help". MATLAB Version 7.0 (R14) Documentation, 2004.

[10] Sarle, Warren S. "Archive of Neural Network FAQ". 2002.
ftp://ftp.sas.com/pub/neural/FAQ.html

[11] Stat Soft Inc. "Neural Networks". 2003.
http://www.statsoft.com/textbook/stneunet.html

[12] Philippides, Andy. "Multi-Layer Perceptron (MLP)". Neural Networks Lectures, 2003.
http://www.informatics.sussex.ac.uk/users/andrewop/Courses/NN/NNs5_6_MLP.ppt