دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران)

دانشکده مهندسی کامپیوتر و فناوری اطلاعات

پایان‌نامه کارشناسی ارشد

رشته مهندسی کامپیوتر – هوش مصنوعی

تشخیص بر خط دستنوشته فارسی مبتنی بر روش های یادگیری

نگارش:

محمد علی بزرگ‌زاده

استاد راهنما:

دکتر محمد رحمتی

دیماه ۱۳۸۶

تقدیم به مادر فداکار و مهربانم.

تقدیر و تشکر

پروردگار را سپاسگزارم که به من توفیق داد این کار کوچک را با موفقیت به پایان برسانم. در اینجا برخود لازم می‌دانم، از تمامی کسانی که در این مدت من را در انجام این پایان‌نامه یاری کردند، تشکر کنم. از استاد بزرگوارم، جناب آقای دکتر رحمتی، که دلسوزانه من را در انجام این پایان‌نامه راهنمایی کردند تشکر می‌کنم. از اساتید محترم، جناب آقای دکتر کبیر و جناب آقای دکتر عبادزاده که زحمت داوری این اثر را داشتند، صمیمانه سپاسگزارم. همچنین، از کلیه عزیزانی که در جمع‌آوری داده‌های مورد نیاز همکاری داشته‌اند، قدردانی می‌کنم.

این پایان‌نامه بر اساس قرارداد شماره ۵۰۰/۱۵۳۷۰/ت مورخ ۱۳۸۵/۱۲/۱۳ تحت حمایت مالی مرکز مخابرات ایران انجام شده است.

چکیده

در این پایان‌نامه، ضمن بررسی تاریخچه و تعدادی از روش‌های متداول تشخیص برخط حروف و کلمات، یک سیستم تشخیص برخط کلمات فارسی، طراحی و پیاده‌سازی شده است. ورودی این سیستم توسط قلم نوری دریافت می‌شود. اخیراً بعلت فراگیر شدن دستگاه‌های کامپیوتر جیبی و تلفن‌های همراه پیشرفته، اهمیت چنین سیستمی، بیش از پیش مورد توجه قرار گرفته است. در روش پیشنهادی ما، عمل شناسایی دستنوشته، از طریق جستجوی پرتو انجام می‌شود. این جستجو بر روی فرضیه‌هایی که با جلو رفتن گام‌های زمانی تکمیل می‌شوند، انجام می‌شود. امتیازدهی به این فرضیه‌ها از راه یافتن فاصله DTW بدنه و علائم حروف منتسب‌شده در فرضیه با الگوهای نمونه‌ای دسته مربوطه‌شان انجام می‌پذیرد.

در جستجوی پرتو، هرس کردن فرضیه‌ها، بسیار اهمیت دارد. به همین منظور، لیست نسبتاً جامعی از ۱۵ تکنیک مرتبط با هرس و کنترل رشد بی‌رویه فرضیه‌ها، ارائه شده است که می‌تواند مرجع مناسبی برای کارهای بعدی در این زمینه باشد. برخی از این تکنیک‌ها، با توجه به ویژگی‌های خاص زبان فارسی و برای اولین بار، مطرح شده‌اند. هرچند برخی از این تکنیک‌ها بطور پراکنده در مقالات مختلف مطرح شده‌اند.

در مرحله آموزش و آماده‌سازی سیستم، الگوهای نمونه‌ای برای دسته‌های مختلف بدنه و علامت حروف فارسی یافت می‌شود. الگوهای نمونه در هر دسته، می‌بایست طوری انتخاب شود که شامل انواع رسم‌الخط‌های نمونه آن دسته باشد. برای جمع‌آوری داده‌های آموزشی، از ۱۶۰ نفر، هر کدام ۳۴ کلمه و ۳۲ حرف مجزای الفبای فارسی جمع‌آوری شد. از انواع سطح تحصیلات از دانش‌آموز گرفته تا دکترا، از سنین مختلف و جنسیت مرد و زن با انواع رسم‌الخط‌های نگارشی در بین این ۱۶۰ نفر بوده‌اند. تقطیع کلیه این کلمات در مرحله آموزش برای جداکردن تکه‌های علائم و حروف هر کلمه بصورت دستی انجام پذیرفته است. سپس الگوهای نمونه‌ای در هر دسته با یک روش خوشه‌بندی پایین‌به‌بالا انتخاب شده است.

در خوشه‌بندی، برای معیار فاصله، از همان معیاری استفاده می‌شود که در مرحله شناسایی نیز برای امتیازدهی به فرضیه‌ها استفاده می‌شود یعنی فاصله DTW که تغییراتی در روش کلاسیک آن، بوجود آمده است. تعریف ارائه‌شده برای فاصله بین دو الگو در DTW کلاسیک، به تعداد تارها حساس است. چنین تعریف فاصله‌ای بطور ضمنی باعث خواهد شد که به سمت ترازبندی‌هایی با تعداد تارهای کمتر تمایل ایجاد شود. برای حل این مشکل، تغییر کوچکی در الگوریتم داده شد بطوریکه بجای مینیمم کردن مجموع فواصل نقاط متناظر، میانگین فواصل نقاط متناظر مینیمم شود.

در این پایان‌نامه، تعریف فاصله بین عناصر دو دنباله، بصورت تابعی غیرخطی از اختلاف زاویه می‌باشد. براساس خروجی تابع غیرخطی استفاده‌شده، هنگامیکه اختلاف زاویه کم باشد (خصوصاً کمتر از ۲۷ درجه)، فاصله، بسیار کم (در حدود ۰,۰۲) می‌باشد که منطقی است زیرا انتظار داریم با اختلافات اندک زاویه بین نقاط متوالی دو دستنوشته که امری طبیعی است با اغماض برخورد شود و وقتی اختلاف زاویه بین دو الگو زیاد می‌شود، رشد فاصله بین دو الگو، تشدید گردد.

همچنین در این کار، بمنظور کاهش تعداد دسته‌ها، برای حروفی که بدنه مشترک دارند دسته‌های بدنه، با هم ادغام شده است. تا جایی که ما اطلاع داریم، ایده نگاه جداگانه به بدنه حروف در ٤ حالت اول، وسط، آخر و مجزا و ادغام دسته‌های بدنه مشترک، برای کاهش تعداد دسته‌ها در کار دیگری، مستندسازی نشده است.

با تنظیم بهینه پارامترها، به نرخ تشخیص صحیح ٨٤,٣٨ ٪ و میانگین زمان اجرای ۷/۵ ثانیه برای تشخیص یک کلمه می‌رسیم. بعنوان داده‌های آزمایشی، از داده‌هایی که از ۸ نفر جمع‌آوری شده بود استفاده کرده‌ایم. هر فرد، ۲۰ کلمه را می‌نویسد که بصورت تصادفی از یک فرهنگ لغت شامل ۱۱۹۰۰ لغت انتخاب می‌شود. چنانچه میانگین زمان اجرا از اهمیت بالایی برخوردار باشد، می‌توان بازاء کاهش اندک در نرخ تشخیص، میانگین زمان اجرا را تا حد زیادی کاهش داد. با تغییر اندک در مقادیر پارامترها، میانگین زمان اجرا ۵ ثانیه کاهش می‌یابد و به ۲/۵ ثانیه تقلیل پیدا می‌کند درحالیکه از نرخ تشخیص صحیح، کمتر از ۳٪ کاسته می‌شود و نرخ تشخیص صحیح به ۸۱,۸۸٪ می‌رسد. درصورتیکه، میانگین زمان اجرا، ۱ ثانیه دیگر کمتر شود و به ۱/۵ ثانیه برسد، نرخ تشخیص صحیح برابر با ۷۸,۷۵٪ خواهد بود. تنظیمات اخیر، احتمالاً برای یک سیستم تجاری با محدودیت زمان اجرا مناسب‌تر است.

از مزایای روش ارائه‌شده، می‌توان به سادگی پیاده‌سازی، امکان تنظیم دقت در مقابل سرعت شناسایی، امکان تطبیق با نویسنده، عدم وجود شرایط محدودکننده در نگارش کلمات، استفاده از تکنیک‌های فراوان برای کنترل رشد فرضیه‌ها و عدم وابستگی حداکثر مدت زمان لازم برای شناسایی کلمه به حجم فرهنگ لغات اشاره نمود.

کلمات کلیدی: تشخیص برخط دستنوشته، فارسی، کلمه، جستجوی پرتو، تارکشی زمانی پویا.

فهرست مطالب

مقدمه ۱

۱-۱. آشنایی با موضوع و معرفی اصطلاحات ۲

۱-۲. انگیزه و کاربردها ۵

۱-۳. هدف از انجام پایان‌نامه ٦

۱-٤. مقایسه و ارتباط با سیستم‌های برون‌خط ۷

۱-۵. دسته‌بندی روش‌های شناسایی برخط ۸

۱-٦. تاریخچه ۹

۱-٦-۱. سیستم‌های تجاری و تحقیقاتی ۱۱

۱-٦-۲. شناسایی برخط زبان‌های فارسی و عربی ۱٦

۱-۷. ساختار پایان‌نامه ۱۷

۱-۸. خلاصه مطالب ۱۸

مراحل متداول شناسایی دستنوشته ۲۰

۲-۱. مراحل پیش‌پردازش و پس‌پردازش ۲۱

۲-۲. تقطیع ۲۳

۲-۳. استخراج ویژگی‌ها ۲٤

۲-٤. تهیه مدل‌های آماری زبان ۲٤

۲-٤-۱. مجموعه نوشتارها ۲۵

۲-٤-۲. مدل ساده جمله ۲۵

۲-٤-۳. مدل جمله تکی ۲٦

۲-٤-٤. مدل‌های جمله دوتایی و چندتایی ۲٦

۲-٤-۵. مدل جمله نرمال‌شده ۲٦

۲-٤-٦. تهیه مدل زبان مستقل از فرهنگ لغات ۲۷
۲-۵. تطبیق با نویسنده ۲۸
۲-٦. خلاصه مطالب ۲۹
آشنایی با چند تکنیک در زمینه شناسایی برخط ۳۱
۳-۱. تطبیق با کلیشه کشسان ۳۲
۳-۲. دگردیسی شکل ۳٤
۳-۳. تکنیک‌های فازی ۳٦
۳-٤. مدل‌های قطعه‌ای ۳۹
۳-٤-۱. یک چارچوب کلی برای مدل کردن دنباله‌ها ٤۰
۳-٤-۲. حالت خاص، HMM کلاسیک ٤۱
۳-٤-۳. حالت خاص، مدل‌های پیشگوی خودبرگشتی ٤۱
۳-٤-٤. حالت خاص، مدل‌های قطعه‌ای ٤۱
۳-٤-۵. یادگیری SM ٤۲
۳-٤-٦. مدل SM چندتایی ٤۳
۳-٤-۷. کدگشایی در SM ٤٤
۳-۵. خلاصه مطالب ٤٤
مراحل آماده‌سازی و آموزش سیستم پیشنهادی ٤٦
٤-۱. دسته‌بندی بدنه حروف فارسی ٤۷
٤-۲. دسته‌بندی علائم حروف فارسی ٤۹
٤-۳. الگوهای نمونه ٤۹
٤-٤. تقطیع کلمات ۵۲
٤-۵. پیش‌پردازش ۵۳

٤-٥-١. نمونه‌برداری مجدد ٥٣
٤-٥-٢. اصلاح کجی ٥٣
٤-٥-٣. هموارسازی ٥٥
٤-٥-٤. اصلاح اندازه ٥٧
٤-٦. خوشه‌بندی برای یافتن الگوهای نمونه ٥٧
٤-٧. محاسبه فاصله بین دو الگو ٥٨
٤-٨. تارکشی زمانی پویا ٥٩
٤-٩. تارکشی زمانی پویا، نسخه تغییریافته ٦١
٤-١٠. الگوهای نمونه ٦٢
٤-١١. خلاصه مطالب ٦٣
مراحل شناسایی کلمات در سیستم پیشنهادی ٦٥
٥-١. سیستم شناسایی دستنوشته، دید کلی ٦٦
٥-٢. امتیازدهی به علائم ٦٧
٥-٣. فرهنگ لغات ٧٠
٥-٤. جستجوی پرتو ٧١
٥-٥. مفهوم و اجزاء فرضیه ٧٢
٥-٦. بسط و گسترش فرضیه‌ها ٧٩
٥-٧. تکنیک‌های کنترل رشد بی‌رویه فرضیه‌ها ٨٢
٥-٨. عملیات پایانی در آخرین گام زمانی ٨٤
٥-٩. خلاصه مطالب ٨٥
آزمایشات تجربی ٨٧
٦-١. پارامترهای پیش‌پردازش ٨٨
٦-٢. پارامترهای کنترل رشد فرضیه‌ها ٨٩

٦-٣. نمونه‌های تشخیص صحیح و عدم تشخیص کلمات ٩٤
٦-٤. خلاصه مطالب ٩٦
نتیجه‌گیری ٩٨
٧-١. نتیجه‌گیری ٩٩
٧-٢. پیشنهاد برای کارهای آینده ١٠٢
فهرست منابع: ١٠٤
ضمیمه الف ١٠٧
آ-١. آشنایی با HMM ١٠٨
آ-٢. شبکه‌های بیزین ١٠٩
آ-٣. مسائل سه‌گانه HMM ١١١
آ-٣-١. مسئله ارزیابی ١١١
آ-٣-٢. مسأله کدگشایی ١١٣
آ-٣-٣. مسأله یادگیری ١١٤
آ-٤. بکارگیری HMM در تشخیص برخط دستنوشته ١١٥
آ-٥. کاهش پیچیدگی محاسباتی با ساختار سلسله مراتبی ١١٧
آ-٦. خوشه‌بندی مبتنی بر HMM حروف ١١٧
آ-٧. متغیرهای تصادفی جهت‌دار ١١٨
آ-٨. خلاصه مطالب ١٢١
ضمیمه ب ١٢٤

فصل اول

مقدمه

در فصل مقدمه، در یک دید کلی به معرفی سیستم‌های شناسایی برخط دستنوشته می‌پردازیم. در ادامه، هدف از اجرای این پایان‌نامه نیز بیان می‌شود. تا کنون سیستم‌های تشخیص برخط، برای زبان‌های مختلف اعم از لاتین، فارسی، عربی یا چینی و ژاپنی، پیاده‌سازی گشته‌اند. در این میان برخی از آنها جنبه تجاری پیدا کرده‌اند و تعدادی تنها دارای جنبه تحقیقاتی هستند. برخی از این سیستم‌ها، به تشخیص حروف مجزا می‌پردازند و عملکرد برخی دیگر در سطح کلمات است و از روش‌های متعددی و متفاوتی در آنها استفاده شده است که تاریخچه مختصری از این سیستم‌ها نیز در انتهای این فصل، بیان می‌شود.

۱-۱. آشنایی با موضوع و معرفی اصطلاحات

سیستم‌های تشخیص حروف نوری یا به اختصار، OCR[1] قدمتی طولانی دارند و طراحی اولین سیستم‌ها در این زمینه به دهه ۶۰ برمی‌گردد. تشخیص حروف، به معنای تبدیل یک نمایش نگارش‌یافته زبان بصورت علائم گرافیکی به نمایش سمبولیک آن است. شناسایی دستنوشته[2] حالت خاصی از OCR است، هنگامیکه ورودی نوشتار انسان و نه حروف تایپی، مدنظر باشد. نمایش نوشتاری اولیه می‌تواند، با استفاده از دستگاه ورودی اسکنر و یا فرمت تصویری مهیا باشد که در اینصورت، سیستم برون‌خط[3] خوانده می‌شود و یا ممکن است با استفاده از یک قلم دیجیتال یا دستگاه دیگری که اطلاعات زمانی نوشتار را همزمان با عمل نوشتن، فراهم می‌آورد باشد که در اینصورت سیستم، برخط[4] گفته می‌شود. تفاوت ورودی‌ها در سیستم‌های برخط و برون‌خط را می‌توان

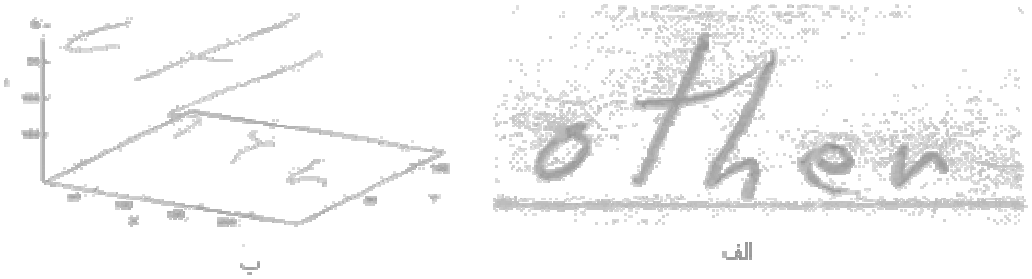
[1] Optical Character Recognition

[2] handwriting recognition

[3] offline

[4] online

می‌توان در شکل ۱-۱ مشاهده نمود. ورودی یک سیستم برخط یک سیگنال دوبعدی گسسته شده در حوزه زمان است. در حالیکه در حالت برون‌خط، با یک تصویر دیجیتال عادی روبرو هستیم. در یک نمونه برخط، معمولاً در حدود ۱۰۰ مختصات مکانی در ثانیه نمونه‌برداری می‌شود و در یک نمونه برون‌خط، تصویر، رزولوشنی در حدود ۳۰۰ نقطه در اینچ دارد.

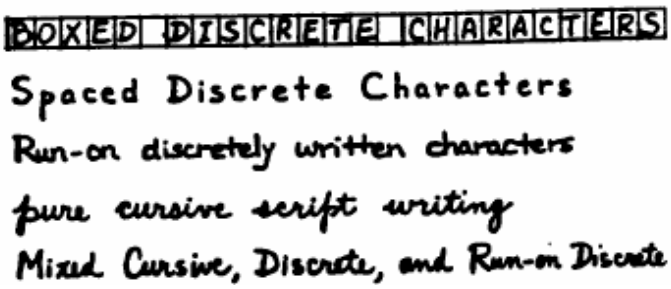


شکل۱-۱. (الف) ورودی یک سیستم برون‌خط: یک تصویر دیجیتال عادی (ب) حالت برخط: یک سیگنال دوبعدی گسسته شده در حوزه زمان، محور عمودی بیانگر زمان است.

هدف در سیستم‌های تشخیص دستنوشته، غلبه بر تنوع نگارشی است. تنوع در دستنوشته‌ها می‌تواند، منشأهای گوناگونی داشته باشد، از آن جمله، تنوع هندسی است یعنی تنوع در اندازه و موقعیت و کجی نوشتارها. و دیگر تنوع رسم‌الخط[1]ها که مربوط به حروفی می‌شود که برای آنها بیش از یک مدل رسم‌الخط وجود دارد. تنوع دیگر نوشتارها مربوط به خصوصیات بیومکانیکی است که بر روی کیفیت نوشتار افراد مختلف، تأثیر دارد و دیگر، تفاوت دستخط‌ها و یا تنوع در ترتیب نگارش اجزا و تنوع در سرعت نوشتن که در سیستم‌های برخط مطرح است و دست آخر، انواع مختلف نویز نیز بر تنوع نوشتارها، تأثیر می‌گذارند [Plamandon 2000].

دستنوشته‌ها در زبان‌های مختلف از لحاظ نحوه نگارش حروف در کنار یکدیگر و ترکیب یافتن آنها، در قالب‌های گوناگونی ارائه می‌شوند. به عنوان مثال برای زبان انگلیسی ۵ نوع قالب نگارش را در شکل ۱-۲ مشاهده می‌کنید که از بالا به پایین عبارتند از: حالتی که هر حرف در مربع مشخصی نوشته می‌شود، حالتی که حروف مجزا هستند، حالتی که حروف مجزا و چسبیده به هم هستند (Run-on)، حالت پیوسته خالص و حالت ترکیبی یا بدون قید. به تناسب برای زبان‌های دیگر نیز می‌توان فرمت‌های مختلفی تشخیص داد.

[1] allograph



شکل۱-۲. قالب‌های نگارشی برای زبان انگلیسی [Tappert 1988]

تفاوت در قالب‌ها بیش از هر چیز، در فاز تقطیع سیستم، تاثیرگذار است. در یک سیستم تشخیص دستنوشته، فازهای ***تقطیع***[1] گوناگونی می‌تواند وجود داشته باشد. تقطیع خطوط نوشتار از تصویر متن اولیه (در سیستم‌های برون‌خط)، تقطیع لغات از خطوط نوشته، تقطیع حروف از لغات و تقطیع ***تکانه‌ها***[2] یا اجزای حروف از حروف، از آن جمله‌اند. البته یک سیستم می‌تواند به تناسب یک یا چند قسمت از این فازهای تقطیع را پیاده‌سازی نکند و یا فاز تقطیع، بصورت تلویحی در الگوریتم شناسایی، در نظر گرفته شود. در اینجا، لازم به ذکر است که واژه تکانه در مقالات به معانی مختلفی از جمله به معنای یک حرف یا به معنای تکه‌ای از یک حرف بکارگرفته شده است. در سیستم‌های برخط، این واژه، به معنای ***تکه‌ای از نوشته از لحظه گذاشتن قلم بر روی صفحه تا لحظه اولین جدا شدن آن***، نیز استعمال شده است. هر جا ما نیز در این نوشتار از این واژه استفاده می‌نماییم، همین معنای اخیر مدنظرمان است.

در یک تقسیم‌بندی، می‌توان تکانه‌ها را به ***تکانه‌های عادی***، ***بست‌ها***[3] و ***تکانه‌های تأخیردار***[4] تقسیم نمود. تکانه‌های عادی، آن‌دسته از تکانه‌هایی هستند که در ترتیب عادی و بعنوان قسمتی از یک حرف نوشته می‌شوند. بست‌ها در واقع جزئی از یک حرف الفبا نیستند و فقط برای اتصال دو حرف در قالب نوشتاری پیوسته در زبان‌هایی مانند انگلیسی ظاهر می‌شوند و تکانه‌های تأخیردار، علائمی مانند نقطه برای حروف نقطه‌دار یا خط افقی برای حرف t هستند که معمولاً بعد از نگارش حروف بعدی‌شان در کلمه و با تأخیر نگاشته می‌شوند. در زبان فارسی هم علائمی چون نقطه، دسته، سرکش، مد و همزه نیز می‌توانند بصورت تکانه‌های تأخیردار ظاهر شوند. بست‌ها و خصوصاً تکانه‌های تأخیردار، هر دو پاشنه آشیل تکنیک‌های شناسایی برخط محسوب می‌شوند طوریکه در برخی روش‌ها، تکانه‌های تأخیردار شناسایی‌شده و کلاً از جریان شناسایی حذف

[1] segmentation

[2] stroke

[3] ligature

[4] delayed stroke

می‌شوند! در برخی روش‌های دیگر، تکانه‌های تأخیردار، تنها در فاز پس‌پردازش بکار گرفته می‌شوند و تنها در بعضی از تکنیک‌ها، با این تکانه‌ها، مستقیماً در هسته الگوریتم شناسایی برخورد می‌شود.

در حالتی که از فرمت حروف پیوسته استفاده می‌شود یک تکه‌نوشته می‌تواند بین دو حرف یا حتی بیشتر مشترک باشد بنابراین به فاز تقطیع حروف نیاز می‌شود. چنین وضعیتی می‌تواند در حالت حروف جدای چسبیده یا حالت ترکیبی نیز روی دهد. اگر نخواهیم از تقطیع استفاده کنیم می‌بایست بجای تشخیص حرف، یک‌راست بسراغ تشخیص کلمات برویم که در این زمینه نیز کارهایی انجام شده است. این روش خصوصاً مواقعی مناسب است که تعداد لغات محدودی در کاربرد مورد نظر، وجود دارد.

سیستم‌های تشخیص حروف، بنا به نیاز می‌توانند به قابلیت‌های تشخیص علائم نوشتاری (مانند ویرگول و نقطه و ...) و قابلیت تشخیص خطوط و فرمول‌ها نیز مجهز شوند. تشخیص نوع حرف و نوع یک علامت می‌تواند یکپارچه باشد یا تحت ماژول‌های مختلف پس از تشخیص کلیت آنها (حرف یا علامت بودن) انجام شود.

۱-۲. انگیزه و کاربردها

برخی کاربردهای تشخیص برخط حروف عبارتند از: کاربردهای پرکردن فرم‌ها، شبیه‌سازی صفحه کلید خصوصاً برای زبان‌های دارای تعداد زیاد حروف [Jaegger 2003] و طراحی بوسیله کامپیوتر (CAD[1]) [Plamondon 2000]. کاربرد دیگری که مطرح شده است، تعلیم نوشتن توسط کامپیوتر است به این ترتیب که کاربر سعی می‌کند الگویی که سیستم به او نشان می‌دهد را بنویسد و از سیستم نمره قبولی یا رد می‌گیرد. تا کنون چندین سیستم تعلیم نگارش برای افراد عادی بزرگسال، افراد بزرگسال مبتلا به بیماری‌های خاص همچون پارکینسون، کودکان عادی و کودکان دارای نقص عضو پیاده‌سازی شده است. از این دسته می‌توان به سیستم‌هایی برای تعلیم نوشتن به زبان انگلیسی و با فرمت پیوسته و متد Palmer یا سیستم تعلیم حروف چینی (فرمت Kanji) اشاره نمود [Plamondon 2000] و [Tappert 1988].

کاربرد مهم دیگر که بهیچ وجه نمی‌توان از آن چشم پوشید، دستگاه‌های PDA[2] یا کامپیوترهای جیبی هستند، کوچک بودن این وسایل فراهم آوردن صفحه‌کلید سخت‌افزاری را دشوار می‌کند، ضمن آنکه اگر چنین سخت‌افزار کوچکی تهیه هم شود، کار با آن راحت نیست. یک راه حل،

[1] Computer Aided Design

[2] Personal Digital Assistant

شبیه‌سازی نرم‌افزاری صفحه‌کلید است که برای کاربران حرفه‌ای هرگز دلچسب نیست. بر این اساس قلم‌های دیجیتال جزء لاینفک کامپیوترهای جیبی هستند. در کامپیوترهای جیبی، صفحه نمایش، همان صفحه قلم نیز می‌باشد. با گذر زمان و پیشرفت تکنولوژی تلفن‌های همراه، چنین نیازی برای تلفن‌های همراه نیز، بیش از پیش احساس می‌شود. مناسب است در اینجا، به کامپیوتر قلمی[1] اشاره کنیم. مفهوم کامپیوتر قلمی، ابتدا در سال ۱۹٦۸ توسط Kay مطرح شد. در این سیستم که به Dynabook، موسوم است، صفحه نمایش به عنوان صفحه ورودی و هم خروجی سیستم محسوب می‌شود. هدف مهم این سیستم جایگزین شدن آن با قلم و کاغذ است [Plamondon 2000].

شناسایی برخط برای زبان‌هایی همانند چینی که تعداد حروف الفبای زیادی دارد، از اهمیت بیشتری برخوردار است زیرا در این زبان‌ها، ورود اطلاعات از طریق صفحه‌کلید زمانگیر است، اما برای زبانی مانند انگلیسی یک تایپیست حرفه‌ای می‌تواند حتی با سرعت بالاتری از نوشتن با قلم، تایپ کند؛ به همین علت تحقیقات بر روی زبان‌های چینی و ژاپنی گسترده‌تر است.

در مورد سیستم‌های برون‌خط البته کاربردهای زیادی مطرح است که از معروفترین آنها، می‌توان به خواندن آدرس‌های پستی بر روی پاکت نامه‌ها و کنترل چک‌های بانکی اشاره نمود. اصولاً تبدیل هر تصویر گرافیکی از یک متن موجود بر روی کاغذ یا یک فایل کامپیوتری با مزایای فراوانی روبرو است که از آن‌جمله می‌توان به کاهش حجم فضای مورد نیاز برای نگهداری، امکان دسته‌بندی و آرشیو بهتر داده‌ها، امکانات گسترده جستجو و بازیابی و انتقال تحت مکانیزم‌های مختلف جابجایی الکترونیکی داده‌ها، اشاره نمود. باید اشاره کرد که کاربردهای تشخیص برخط و تشخیص برون‌خط را نمی‌توان چندان از هم متمایز نمود و این بعلت استفاده از تکنیک‌های برون‌خط در روش‌های برخط و بالعکس است که در ادامه به آن اشاره خواهیم کرد. یک سیستم تشخیص حروف، می‌تواند گام اول از یک سیستم بزرگتر باشد و گام بعدی می‌تواند تعبیر دستنوشته[2] باشد که به درک مفهوم نوشته می‌پردازد.

۱-۳. هدف از انجام پایان‌نامه

هرچند در مورد زبان انگلیسی و ژاپنی کارهای متعددی در زمینه تشخیص برخط انجام گرفته است، در زبان‌های فارسی و عربی تعدد کارها بسیار کمتر است و بیشتر آنها، در ارتباط با تشخیص حروف مجزا می‌باشد. حتی آن تعداد از کارها که در ارتباط با تشخیص کلمات نیز هستند،

[1] pen computer

[2] handwriting interpretation

دارای محدودیت‌های فراوانی در طرز نگارش کلمات می‌باشند. مثلاً در کار [رضوی ۱۳۸۵]، نویسنده می‌بایست برای تشخیص درست، قواعد مشخصی را رعایت کند از آنجمله: هر زیرکلمه باید در یک حرکت قلم نوشته شود، علائم هر زیرکلمه می‌بایست بعد از نوشتن زیرکلمه و قبل از نوشتن زیرکلمه بعدی باشد. نقطه‌های یک حرف باید بهم بچسبند، موقعیت علائم نسبت به بدنه‌شان باید همپوشانی داشته باشد، بدنه هر زیرکلمه نسبت به زیرکلمه قبلی خود باید همپوشانی افقی داشته باشد. تلاش ما در اینکار این است که سیستمی با عملکرد مناسب برای تشخیص کلمات برخط زبان فارسی ارائه‌کنیم و تا جایی که ممکن است قاعده محدود کننده‌ای برای نوشتن کلمات نداشته باشیم.

۱-٤. مقایسه و ارتباط با سیستم‌های برون‌خط

یک مزیت سیستم‌های برخط، برخورداری از اطلاعات زمانی نوشتار است که بر مبنای آن تعداد قطعه‌های تشکیل‌دهنده یک حرف، جهت هر کدام از آنها و سرعت نوشتن آنها مشخص می‌شود. در مقابل سیستم‌های برون‌خط، نیاز به عملیات پیش‌پردازشی دارند که به یافتن کانتورهای حروف و نازک‌سازی یا یافتن اسکلت آنها می‌پردازد. البته ممکن است عنوان شود، که یک حرف می‌تواند با قطعاتی که از لحاظ ترتیب و جهت یکسان نیستند، تولید شود. از این لحاظ اطلاعات زمانی نه تنها مفید نیست بلکه گیج‌کننده نیز خواهد بود اما عمدتاً می‌توان راه‌حل‌های مناسبی یافت و نتایج قوی‌تری نسبت به سیستم‌های برون‌خط کسب نمود.

گزارش‌هایی که در مورد نرخ تشخیص دستنوشته‌ها در سیستم‌های برخط وجود دارد بر نمونه‌های برون‌خط، برتری دارند. بعنوان مثال، در [Plamondon 2000] به این مطلب اشاره شده است که برای تشخیص لغت از یک فرهنگ لغت ۱۰۰، ۱۰۰۰ و ۱۰۰۰۰ کلمه‌ای به ترتیب نرخ تشخیص ۹۵٪ ، ۸۵٪ و ۷۸٪ گزارش شده است در حالیکه برای سیستم‌های برخط می‌توانیم این نرخ‌های تشخیص را در فرهنگ‌های بزرگتر داشته باشیم بعنوان مثال، برای فرهنگ لغت ۲۱۰۰۰ کلمه‌ای نرخ تشخیص ۸۰٪ ارائه شده‌است، ضمن آنکه در سال‌های اخیر آمار قابل توجهی نیز ارائه شده که البته به مجموعه تست استفاده شده وابسته است. بعلت دقت بالای سیستم‌های برخط، این ایده مطرح می‌شود که در تشخیص برون‌خط، ابتدا خط سیر[1] نوشته را تخمین بزنیم و سپس از الگوریتم‌های برخط استفاده کنیم [Guadian 2005]، اما تحقیقات نشان می‌دهد تهیه این اطلاعات زمانی کار دشواری است [Plamondon 2000].

[1] trajectory

مزیت دیگر سیستم‌های برخط قابلیت تطبیق است. البته در اینجا منظورمان قابلیت تطبیق نویسنده با سیستم شناسایی است نه برعکس آن، یعنی اینکه نویسنده می‌تواند با مشاهده نتایج شناسایی بطور همزمان، فرم نوشتن اجزایی را که درست تشخیص داده نشده‌اند اصلاح کند. اما یکی از مواردی که در سیستم‌های برخط، باید در نظر گرفته شود، سرعت آنهاست. اگر بخواهیم ، نتایج شناسایی بدون تأخیر ارائه شود، باید سیستم بتواند در یک ثانیه در حدود ۱ الی ۲,۵ کاراکتر را برای زبان انگلیسی و ۰,۲ الی ۲,۵ کاراکتر را برای زبان چینی، تشخیص دهد [Tappert 1988] .

در زمینه تشخیص برخط حروف، یکی از مشکلاتی که یک محقق با آن روبرو است، دشواری مقایسه با نتایج کارهای قبلی است. زیرا عموماً افراد مستقیماً دست به انتخاب نویسنده و تهیه نمونه‌های آزمایشی خود می‌زنند و تنوع دستخط و حیطه‌های کاربردهای مختلف این سیستم‌ها، مقایسه آنها با یکدیگر را دشوار می‌سازد.

۱-۵. دسته‌بندی روش‌های شناسایی برخط

دو دسته مهم از روش‌های تشخیص برخط دستنوشته، روش‌های ساختاری مبتنی بر قاعده و روش‌های آماری هستند. روش‌های ساختاری و مبتنی بر قاعده، اولین بار در دهه ٦٠ مطرح شدند اما بعلت دشوار بودن یافتن قواعدی که بتواند مدل یک حرف را تعیین کند و تعداد زیاد آنها و درضمن دشواری ارائه روشی برای خودکار نمودن استخراج قواعد برای مدتی، کنار گذاشته شدند [Plamondon 2000]. از دهه ۹۰ به بعد، شاهد رجوع مجدد به این روش‌ها و عمدتاً با ساختارهای قواعد فازی و تولیدشده بصورت خودکار از حجم بالای داده‌های آموزشی هستیم مانند کارهای [Bouslama 1999] و [Malavtia 1994].

در روش‌های آماری و شبکه‌های عصبی، همانطور که می‌دانید، تعداد مشخصی ویژگی استخراج می‌شود و عملیات دسته‌بندی بر روی فضای چندبعدی ویژگی‌ها انجام می‌گیرد. از میان انواع شبکه‌های عصبی، شبکه‌های نگاشت ویژگی خودسازمانده کوهونن[1] و شبکه‌های عصبی تأخیر زمانی کانولوشنی[2] بیشتر مورد توجه قرار گرفته است. از میان تکنیک‌های آماری شناسایی الگو، شبکه‌های باور بیزین و از آن‌جمله، HMM[3] بسیار پرکاربرد است (برای آشنایی با HMM و کاربرد آن در تشخیص برخط، به ضمیمه الف مراجعه شود).

[1] self-organized feature maps

[2] convolutional time-delay

[3] Hidden Markov Model

البته تکنیک‌ها، به مطالب بالا، منحصر نمی‌شود و روش‌های زیاد دیگری مطرح است که برخی از آنها عبارتند از : برنامه‌نویسی پویا (DP[1]) [Lin 2002]، روش تطبیق با کلیشه کشسان[2] [Connel 2001]، روش‌های یادگیری فازی [Malavtia 1994] ،سیستم‌های یادگیری ژنتیک [Verma 2003]، شبکه‌های عصبی [Seen 1999] و [Lee 1989]، استفاده از شبکه‌های بیزین [Hasegava 2001]، شبکه‌های نوروفازی [Alimi 1997] و دگردیسی اشکال[3] [Pavidis 1998].

۱-٦. تاریخچه

از جنبه تاریخی، سیستم‌های OCR تا کنون سه مرحله تکاملی را پشت سر گذاشته‌اند[4] [Nazif 2001] که عبارتست از: مرحله تکوین، مرحله توسعه و مرحله بهبود.

الف) مرحله تکوین (از ۱۹۰۰ تا ۱۹۸۰). رد پای اولیه اقدامات صورت‌گرفته در زمینه بازشناسی حروف را در سال‌های اول دهه ۱۹۰۰ می‌توان یافت و آن زمانی است که تورینگ، دانشمند روسی بر آن بود که به افراد مبتلا به نارسایی‌های بینایی کمک نماید [Mantas 1986]. اولین اختراع‌های ثبت‌شده در این زمینه مربوط به سال‌های ۱۹۲۹ و ۱۹۳۳ میلادی هستند [Mori 1992]. این سیستم‌ها حروف چاپی را با روش تطابق قالبی شناسایی می‌کردند؛ به این صورت که ماسک‌های مکانیکی مختلفی از مقابل تصویر حرف عبور می‌کردند (مکانیکی) و نور از یک سو به آن تابانده می‌شد و از سوی دیگر توسط یک آشکارساز نوری دریافت می‌گردید (اپتیکی). وقتی یک انطباق کامل صورت می‌گرفت، نور به آشکارساز می‌رسید و حرف ورودی بازشناسی می‌شد. این اختراع به دلیل فناوری اپتومکانیکی مورد استفاده در آن، کاربردی نبود. تصور دسترسی به دستگاهی برای بازشناسی حروف تا دهه ۱۹٤۰ میلادی و ظهور رایانه‌های دیجیتال، به صورت یک رؤیا باقی ماند.

اقدامات اولیه در زمینه بازشناسی حروف، بر متون چاپی یا مجموعه کوچکی از حروف و نمادهای دستنوشت که براحتی قابل تشخیص بودند، متمرکز گردیده بود. سیستم‌های بازشناسی حروف چاپی که در این مقطع زمانی عرضه شدند، عمدتاً از روش تطابق قالبی استفاده می‌نمودند که در آن، تصویر ورودی با مجموعه بزرگی از تصاویر حروف، مورد مقایسه قرار می‌گرفت. در مورد متون دستنوشت نیز الگوریتم‌های پردازش تصویر که ویژگی‌های سطح پایین (ویژگی‌هایی که مستقیماً

---

[1] Dynamic Programming

[2] template matching

[3] shape metamorphosis

[4] مراحل تکامل (قسمت‌های الف، ب و ج در این بخش) عیناً از [فرامرزی ۱۳۸٤] نقل شده است.

و بدون اعمال هیچ تبدیلی، از تصاویر استخراج می‌شوند) را از تصاویر استخراج می‌کنند، در مورد تصاویر دوسطحی اعمال می‌شدند تا بردارهای ویژگی استخراج گردند. سپس این بردارهای ویژگی به طبقه‌بندی‌کننده‌های آماری سپرده می‌شدند.

در این دوره، تحقیقات موفق اما مقید (منظور از مقید، مفروض دانستن شرایط و پیش‌فرض‌های خاص برای کاراکترهای ورودی است)، بیشتر بر روی حروف و اعداد لاتین انجام گرفت. با این حال مطالعات چندی نیز بر روی حروف ژاپنی، چینی، عبری، هندی، سیریلیکی، یونانی و عربی در هر دو زمینه حروف چاپی و دستنوشت آغاز گردید. با ظهور صفحات رقومی‌کننده در دهه ۱۹۵۰ که قادر به تشخیص مختصات حرکتی نوک یک قلم مخصوص بودند، سیستم‌های «اُسی‌آر» تجاری نیز امکان عرضه یافتند. این نوآوری سبب شد که محققان بتوانند در زمینه بازشناسی برخط حروف دستنوشت، فعالیت خود را آغاز نمایند. مرجع [Suen 1990] یک منبع مناسب درباره اقدامات صورت‌گرفته بر روی بازشناسی برخط حروف تا سال ۱۹۸۰ می‌باشد.

ب) مرحله توسعه (از ۱۹۸۰ تا ۱۹۹۰): مطالعات صورت گرفته تا قبل از سال ۱۹۸۰ بدلیل فقدان سخت‌افزارهای رایانه‌ای قدرتمند و دستگاه‌های اخذ داده‌ها با مشکل همراه بودند. در این دهه بواسطه رشد انفجارگونه فناوری اطلاعات، وضعیت بسیار مناسبی برای تحقیقات مختلف از جمله بازشناسی حروف فراهم گردید. روش‌های ساختاری به همراه روش‌های آماری در بسیاری از سیستم‌ها استفاده شدند.

تحقیقات در زمینه «اُسی‌آر» اساساً به توسعه روش‌های بازشناسی معطوف گردید، بی آنکه مسئله استفاده از اطلاعات معناشناختی به منظور افزایش دقت بازشناسی مورد توجه قرار گیرد. این امر سبب گردید که دقت بازشناسی (نرخ بازشناسی) از یک حد خاص فراتر نرود، که در بسیاری از کاربردهای «اُسی‌آر»، قابل قبول نبود. مروری بر تحقیقات و پیشرفت‌های حاصل‌شده در مورد «اُسی‌آر» در این برهه زمانی را می‌توان در [Suen 1990] و [Mori 1992] بترتیب برای بازشناسی برخط و برون‌خط ، جستجو نمود.

ج) مرحله بهبود (از ۱۹۹۰ به بعد): در این مقطع زمانی بود که با تکوین ابزارها و تکنیک‌های پردازشی جدید، پیشرفت واقعی در سیستم‌های «اُسی‌آر» محقق گردید. در اوایل دهه ۹۰، روش‌های پردازش تصویر و بازشناسی الگو با تکنیک‌های کارآمد هوش مصنوعی ادغام گشتند. محققان، الگوریتم‌های پیچیده‌ای را در بازشناسی حروف ابداع نمودند که قادر بودند داده‌های ورودی با تفکیک‌پذیری بالا را دریافت کنند و در مرحله پیاده‌سازی، محاسبات بسیار زیادی را بر روی داده‌ها انجام دهند. امروزه علاوه بر وجود رایانه‌های قدرتمندتر و تجهیزات الکترونیکی دقیق‌تر مانند اسکنرها، دوربین‌ها و صفحات رقومی‌کننده، استفاده از تکنیک‌های پردازشی مدرن و توانمند همچون

شبکه‌های عصبی، مدل‌های مارکوف پنهان، منطق‌های مجموعه فازی و مدل‌های پردازش زبان طبیعی امکان‌پذیر گشته است.

جدول ۱-۱. جایگاه کنونی پیشرفت‌های حاصل‌شده در زمینه سیستم‌های تشخیص متون لاتین

		متون چاپی			متون دستنوشته		
		یک نوع فونت	چند نوع فونت	همه نوع فونت	گسسته	پیوسته	مخلوط
برخط	مقید						
	نامقید						
برون‌خط	بدون نویز						
	نویزی						

■ نیازمند تحقیقات بیشتر ▒ نیازمند بهبود □ درحد مطلوب

سیستم‌های جدید «اُسی‌آر» برون‌خط متون چاپی و برخط متون دستنوشت با واژگان محدود و وابسته به نویسنده، در کاربردهای محدود به نحو کاملاً رضایت‌بخشی عمل می‌کنند [Nazif 2001]. اما به منظور دستیابی به هدف نهایی در شبیه‌سازی ماشینی نگارش انسانی و متون چاپی، هنوز راه درازی در پیش است. «آریکا»[Nazif 2001] برخی از تکنیک‌های عرضه‌شده در این دوره را مرور کرده است. جدول ۲-۱ جایگاه کنونی پیشرفت‌های حاصل‌شده در زمینه سیستم‌های «اُسی‌آر» برای متون لاتین را به نمایش می‌گذارد [Nazif 2001]. توجه شود که برای متون چاپی، پردازش برخط تعریف نمی‌شود.

۱-٦-۱. سیستم‌های تجاری و تحقیقاتی

اکنون، تشخیص برخط دستنوشته، تنها یک زمینه تحقیقاتی دانشگاهی نیست و خصوصاً در مورد زبان انگلیسی به یک تکنولوژی کاربردی در دسترس عموم تبدیل گشته است؛ برای مثال می‌توان به موفقیت محصولات CrossPad از شرکت‌های AT.Cross و IBM یا Calligrapher از Paragraph Int اشاره نمود. خلاصه‌ای از مشخصات سیستم‌های تجاری و تجربی اولیه را می‌توانید در جدول‌های ۱-۲ الی ۱-٤ مشاهده نمایید. جدول‌های ۱-۲ و ۱-۳ مربوط به حروف مجزا و جدول ۱-٤ مربوط به قالب پیوسته است.