
بسم الله الرحمن الرحيم
الحمد لله رب العالمين الرحمن الرحيم مالك يوم الدين
اهدنا الصراط المستقيم صراط الذين
انعمت عليهم غير المغضوب عليهم
ولا الضالين

دانشکده برق و رباتیک

گروه الکترونیک

پایان‌نامه کارشناسی ارشد مهندسی برق-الکترونیک

تشخیص برون خطی کلمه دست نوشته فارسی با استفاده از مدل مخفی مارکوف

زهرا ایمانی

استاد راهنما :

دکتر علیرضا احمدی فرد

استاد مشاور:

دکتر حسین خسروی

پایان نامه ارشد جهت اخذ درجه کارشناسی ارشد

ماه و سال انتشار: **شهریور** 92

تقدیم به

پدر و مادر مهربانم که زندگیم را مدیون مهر و عطوفت آن‌ها می‌دانم

همسرم که نشانه لطف الهی در زندگی من است

و برادرانم که همراهان همیشگی و پشتوانه‌های زندگیم هستند.

تقدیر و تشکر

سپاس بی کران پروردگار یکتا را که هستی مان بخشید و به طریق علم و دانش رهنمونمان شد و به همنشینی رهروان علم و دانش مفتخرمان نمود و خوشه چینی از علم و معرفت را روزیمان ساخت.

از استاد با کمالات و شایسته جناب آقای دکتر علیرضا احمدی فرد که در کمال سعه صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ ننمودند و زحمت راهنمایی این رساله را بر عهده گرفتند؛ کمال تشکر و قدردانی را دارم. همچنین از رهنمون های خردمندانه جناب آقای دکتر حسین خسروی نیز متشکرم.

در انتها از همسر عزیزم جناب آقای مهندس عباس زهره وند برای همیاری در انجام این رساله سپاسگذارم.

چکیده

در این پایان نامه سیستمی برای بازشناسی برون خط کلمات دست‌نویس فارسی ارائه شده است. پیچیدگی نگارش فارسی و شکل متفاوت حروف بسته به موقعیت آن که اول کلمه، وسط کلمه، انتهای کلمه یا جدا از بخش های دیگر کلمه باشد بازشناسی کلمات در این زبان را بسیار دشوار نموده است. بطوریکه درصد بازشناسی روش های موجود کمتر از نرخ مطلوب برای تجاری شدن این سیستم‌ها می باشد. یک سیستم بازشناسی کلمه در حالت کلی شامل مراحل پیش‌پردازش، استخراج ویژگی و کلاسه‌بندی است. در این پایان نامه سه روش جهت بازشناسی پیشنهاد شده و کارایی این روش ها در مقایسه با روش های موجود ارزیابی شده است.

به دلیل نبود یک پایگاه داده مناسب و با تعداد تصویر کافی برای ارزیابی کارایی سیستم پیشنهادی، پایگاه داده فارسا، شامل 30000 تصویر از 300 کلمه متداول دست نوشته در زبان فارسی، ایجاد شد. جهت ارزیابی روش های پیشنهادی و مقایسه با روش های موجود 198 کلاس از این پایگاه داده مورد استفاده قرار گرفت.

در روش اول تصویر کلمه به پنجره‌های عمودی دارای همپوشانی تقسیم می‌شود و از هر پنجره هیستوگرام کدهای زنجیره‌ای استخراج می‌شود، هر پنجره یک بردار 20 عنصری را تولید می‌نماید. برای آموزش و ارزیابی HMM گسسته بردار ویژگی‌های استخراج شده از پنجره لغزان، با استفاده از شبکه عصبی نگاشت خودسازمانده کوانتیزه می‌شود . برای تعداد حالت‌های مخفی مدلHMM در هر کلاس کلمه ، کمترین تعداد پنجره برای تصاویر آموزش در هر کلاس بدست می‌آید. تعداد حالت‌های مخفی ضریبی از این مقدار کمینه است. با آزمایش ضرایب مختلف، سیستم در مقدار 1/8 به بهترین نرخ بازشناسی می‌رسد. برای پارامتر هموارسازی روی مقادیر مختلف آزمایش شد، سیستم در مقدار 0/001 به جواب بهتری رسید. اندازه کتاب رمز 49 تنظیم می‌شود. در نهایت روش پیشنهادی اول با تعیین ضرایب بیان شده در بالا به نرخ بازشناسی 66/57% در198 کلاس از پایگاه داده فارسا می‌رسد.

در روش پیشنهادی دوم علاوه بر هیستوگرام کدهای زنجیره‌ای، از ویژگی میانگین بلوکی برای کلاسه‌بندی استفاده می‌شود و ابعاد بردارهای ویژگی به 25 افزایش می‌یابد. نرخ بازشناسی با استفاده از همان ضرایب تنظیم شده در روش اول، 68/88% بدست می‌آید. که افزایش بیش از 2% را نسبت به روش اول نشان می‌دهد.

روش پیشنهادی سوم در واقع بکارگیری یک سیستم دو خبره‌ای است. با استفاده از بررسی که روی نتایج ارزیابی روش دوم انجام شد یک مفهوم جدید به نام معیار اطمینان برای کلاسه‌بند HMM معرفی می‌شود. با مشاهده هیستوگرام اختلاف دو بزرگترین احتمال در خروجی HMM برای تصاویر آزمون یک مقدار آستانه به عنوان معیار اطمینان معرفی می‌شود. تصاویری که شرایط معیار اطمینان را دارند با کلاسه‌بند HMM بازشناسی می‌شوند و تصاویری که این شرایط را ندارند با استفاده از کلاسه‌بند KNN بازشناسی می‌شوند. برای کلاسه‌بند KNN از ویژگی‌های ساختاری ، تعداد مولفه‌های متصل تصویر، تعداد مولفه‌های متصل بالای خط کرسی و تعداد مولفه‌های متصل پایین خط کرسی، استفاده می‌شود. کلاسه‌بند KNN با استفاده از این ویژگی‌ها و با 11 نزدیک‌ترین همسایه و معیار فاصله بلوک شهری به نرخ بازشناسی 61/69% دست می‌یابد. در این سیستم HMM برای تصاویری که معیار اطمینان را دارند به نرخ بازشناسی 85% دست می‌یابد. در کل نرخ بازشناسی این روش برای 198 کلاس از پایگاه داده فارسا 76/49% است. که نسبت به روش اول افزایش 7% را بدنبال دارد.

کلید واژه: بازشناسی کلمه دست‌نوشته فارسی، پایگاه داده فارسا، هیستوگرام کدهای زنجیره‌ای، نگاشت خودسازمانده، مدل مخفی مارکوف، معیار اطمینان، k نزدیک ترین همسایه

لیست مقالات مستخرج از پایان‌نامه

1-"معرفی پایگاه داده فارسا: تصاویر دیجیتال از کلمات دست‌نویس فارسی"، یازدهمین کنفرانس سیستم‌های هوشمند ایران، ICIS 2013

2- "Offline handwritten Farsi cursive text recognition Using Hidden Markov Model", 8th Iranian Conference on Machine Vision and Image Processing, MVIP2013

3- "A confidence-base method for handwritten Farsi word Recognition", International Journal of Computer Applications (IJCA)

فهرست

فصل اول 1

1-1بازشناسی دستنوشته 1

2-1مشخصات دستنوشته فارسی 2

3-1روش شناسی 6

فصل دوم 10

فصل سوم 22

1-3 پیش پردازش 22

1-1-3 استخراج خط کرسی 22

2-1-3 یکسان سازی عرض حرکت 23

1-2-1-3 استفاده از همسایگی پیکسل های پیش‌زمینه[13] 23

2-2-1-3 استفاده از اسکلت‌بندی و انبساط[11] 24

2-3 استخراج ویژگی 25

1-2-3 سیستم پنجره لغزان 26

1-1-2-3کدهای زنجیره‌ای 27

2-1-2-3 میانگین بلوکی 29

3-1-2-3 ویژگی‌های انتقال[2] 29

2-2-3 ویژگی‌های ساختار- مانند 30

3-3شبکه عصبی خودسازمانده کوهنن(SOM) [19] 30

1-3-3 الگوریتم SOM : 32

4-3 کلاسه‌بند 35

1-4-3 مدل مخفی مارکوف[11] 35

1-1-4-3پارامترهای HMM 36

3-4-1-2 HMM گسسته یا پیوسته 37
3-4-1-3 سه مسئله اساسی درHMM 38
3-4-1-4 آموزشHMM برای سیستم بازشناسی کلمه دست‌نوشته 39
3-4-1-5 آزمایش سیستم HMMبرای بازشناسی کلمه دست‌نوشته 43
3-4-1-6 انواع مختلف HMM [24] 45
3-4-2 هموارسازی پارامترهای HMM با استفاده از شبکه خودسازمانده[22] 47
فصل چهارم 51
4-1 مقدمه 51
4-2 پایگاه داده فارسا[28] 52
4-2-1 جمع‌آوری و مرتب‌سازی کلمات پرکاربرد 52
4-2-2 طراحی فرم و جمع‌آوری دست‌نوشته از افراد 53
4-2-3 اسکن فرم‌ها 53
4-2-4 جداسازی کلمات از فرم‌ها و برچسب گذاری 55
4-2-5 حذف نویز و بهبود کیفیت تصویر 56
فصل پنجم 60
5-1مقدمه 60
5-1-1پیش پردازش 61
5-1-2 استخراج ویژگی 61
5-1-3 کوانتیزه سازی بردارهای ویژگی 61
5-1-4 کلاسه‌بندی 61
5-1-4-1ساختارHMM و مقداردهی اولیه پارامترها 62
5-1-4-2 هموارسازی پارامترهای HMM 64
5-2 روش‌های پیشنهادی 64

3-5 تعاریف اولیه: 64
4-5 روش پیشنهادی اول: استفاده از هیستوگرام کدهای زنجیره‌ای 66
1-4-5 نتایج ارزیابی عملکرد روش اول با استفاده از پایگاه داده فارسا 67
1-1-4-5 HMM ها با تعداد حالت یکسان 68
2-1-4-5 HMM ها با تعداد حالت‌های مخفی متفاوت 73
5- 5 روش پیشنهادی دوم: استفاده از بردار ویژگی 25 تایی 75
1-5-5نتایج ارزیابی عملکرد روش دوم با استفاده از پایگاه داده فارسا 75
6-5 روش پیشنهادی سوم: بازشناسی به کمک دو خبره 76
1-6-5 ارزیابی روش بازشناسی دوخبره‌ای با استفاده از پایگاه داده فارسا 81
7-5 ارزیابی روش‌های پیشنهادی با استفاده از پایگاه داده ایرانشهر 84
8-5 تحلیل نتایج و مقایسه با روش‌های موجود 86
فصل ششم 89
1-6 نتیجه گیری 89
2-6 پیشنهادات 91
مراجع 92

فهرست شکل‌ها

شکل1-1: برخی ویژگی های نگارش فارسی[1] 4
شکل1-2: نمونه هایی از ادغام حرف مجاور در متون دست نویس[1] 6
شکل1-3: فضای بین کلمه ای[3] 7
شکل 2-1: خلاصه الگوریتم استخراج نقطه پیشنهاد شده در[2] 15
شکل2-2: نحوه استخراج ویژگی انتقال[2] 16
شکل2-3: ناحیه های مورد استفاده برای استخراج ویژگی و نوار متحرک[9] 17
شکل2-4: فریم بندی تصویر برای استخراج ویژگی [11] 18
شکل2-5: چهارنوع پیکربندی تقعر برای پیکسل پس‌زمینه P [12] 19
شکل 3-1: نحوه تخمین مکان خط کرسی، (الف) تصویر کلمه، (ب) هیستوگرام افقی کلمه[12] 23
شکل3-2: (الف) شکل اصلی، (ب) شکل اصلی بعد از یک مرحله تصحیح عرض حرکت [13] 24
شکل3-3: (الف) تصویر اصلی، (ب) اسکلت تصویر، (ج) اسکلت منبسط شده 25
شکل3-4: حرکت نوار لغزان روی تصویر کلمه [11]، 27
شکل3-5 : جهت های کد زنجیره ای پایه، (الف) برای چهار جهت، (ب) برای 8 جهت [14] 28
شکل3-6 : مرز با کد زنجیره‌ای 8 تایی[17] 28
شکل3-7: 4 جهت بدست آمده از 8 جهت [17] 29
شکل3-8: ساختارشبکهSOM 31
شکل3-9: نرون‌های لایه خروجی یک بعدی 32
شکل3-10: تغییر تدریجی شعاع همسایگی[19] 32
شکل3-11: یک مدل HMM با سه حالت و چهار بردار مشاهده[22] 37
شکل 3-12 : ساختارهای متفاوت شبکه ی انتقالی HMM [24] 45
شکل3-13:(الف) بردار احتمال انتشار، (ب)نمایش بردار احتمال انتشار به صورت ماتریسی 48
شکل3-14: همسایگی‌ها برای یک هدف 48
شکل4-1: یک نمونه پر شده از فرم شماره1 54
شکل4-2: تصویر سازی افقی و عمودی برای آشکارسازی مکان سطرها و ستون‌ها در فرم 55
شکل4-3: تصویر استخراج شده از فرم شماره 1 56

شکل4-4: مراحل دو سطحی سازی تصویر، 57
شکل4-5: شش نمونه از دو کلمه در پایگاه داده فارسا 58
شکل5-2: یک مدل مخفی مارکوف با چهار حالت[11] 62
شکل5-3: پنجره بندی تصویر قبل از استخراج ویژگی 67
شکل 5-4: بردار ویژگی کد زنجیره‌ای برای یک پنجره 67
شکل 5-5: هیستوگرام چینش بردارهای ورودی شبکه SOM 49تایی 69
شکل5-6: بردارهای وزن در شبکهSOM برای ورودی ها 69
شکل5-7: هیستوگرام اختلاف احتمال دو تا بهترین مدل 79
شکل 5-7: بلوک دیاگرام مرحله کلاسه بندی در روش پیشنهاد شده دو خبره‌ای 80
شکل5-8: استخراج ویژگی ساختار-مانند 81
شکل5-9: تقسیم بندی نمونه‌هایی از تصاویر با معیار اطمینان 81
شکل 5-10: نمونه‌های از تصاویر موجود در پایگاه داده ایرانشهر 84
شکل 5-11: مقایسه نتایج بدست آمده از روش‌های پیشنهادی در پایگاه داده فارسا و ایرانشهر 86
شکل5-12: مقایسه نتایج نرخ بازشناسی روش‌های موجود و روش های پیشنهادی 87

فهرست جداول

جدول1-1: الفبای فارسی[2] 5
جدول2-1: تنوع نقاط در نوشتار فارسی[2] 15
جدول5-1: درصد نرخ بازشناسی روش اول برای 10 کلاس از پایگاه داده فارسا برای عرض پنجره لغزان متفاوت 68
جدول5-2: نتایج آزمایش روش اول بر روی 10 کلاس اول از پایگاه داده فارسا 70
به ازای تعداد حالت‌های مخفی متفاوت HMM 70
جدول5-3: نرخ بازشناسی با استفاده از هموارسازی ماتریس احتمال مشاهدات 71
جدول5-4: ماتریس سردرگمی سیستم برای 10 کلاس اول از پایگاه داده فارسا 72
جدول5-5: نرخ بازشناسی با استفاده از ویژگی هیستوگرام کدهای زنجیره‌ای برای 198 کلاس 73
جدول 5-6: نتایج بازشناسی برای HMM ها با تعداد حالت های مخفی متفاوت 74
جدول 5-7: ارزیابی سیستم برای HMM ها با تعداد حالت‌های مختلف بعد از هموار سازی 75
جدول 5-8: نرخ بازشناسی روش پیشنهادی دوم قبل از هموار سازی پارامترهای HMM به ازای ضرایب مختلف کنترل کننده تعداد حالت‌های مخفی 76
جدول5-9: نرخ بازشناسی روش پیشنهادی دوم بعد از هموار سازی پارامترهای HMM به ازای مقادیر مختلف پارامتر SF 76
جدول5-11: نرخ بازشناسی HMM برای تصویرهایی که شرایط معیار اطمینان را دارند 83
جدول 5-12: ارزیابی نرخ بازشناسی روش های پیشنهادی با استفاده از پایگاه داده ایرانشهر 85
جدول5-13 نرخ بازشناسی روش دوخبره‌ای با استفاده از پایگاه داده ایرانشهر 86

فصل اول:

مقدمه

فصل اول

1-1بازشناسی دست‌نوشته[1]

در حدود بیش از 350 میلیون نفر در جهان به زبان‌هایی صحبت می‌کنند که برای نوشتن آن‌ها از دست‌خط عربی یا شبیه به آن استفاده می‌شود. از زبان‌های غیر عربی می‌توان به فارسی، دری، پشتو، اردو وکردی اشاره کرد.

در نیمه اول دهه 1980 میلادی کارهای انگشت شماری درباره بازشناسی کلمات دست‌نویس، حروف و ارقام دست‌نویس انجام شد. در نیمه دوم این دهه پژوهش‌هایOCR [2]رونق بیشتری گرفت. مهمترین زمینه‌ها در این دوره عبارتند از: بازشناسی ارقام، حروف و مجموعه‌های بسیار محدود کلمات دست‌نویس، ارقام و حروف مجزای چاپی و کلمات چاپی با قلم‌های محدود. از سال 2000 میلادی تاکنون، پژوهش‌هایی در موارد زیر گزارش شده‌است: بازشناسی کلمات چاپی بدون جداسازی آن‌ها به حروف یا زیر حروف، تعیین نوار زمینه یا خط کرسی[3] در متون چاپی ودست‌نویس، بازشناسی اسامی شهرها و مبالغ مندرج در چک‌های بانکی به حروف، ترکیب طبقه بندها در بازشناسی کلمات. البته کارهای مربوط به بازشناسی نویسه‌ها و کلمات دست‌نویس و شکستن کلمات چاپی و دست‌نویس به حروف و زیر حروف و بازشناسی آن‌ها همچنان ادامه دارد.

در سال 2005 اولین مسابقه OCR عربی درباره بازشناسی اسامی دست‌نویس شهرها برگزار شد.

---

[1]Handwritten Recognition

[2]Optical Character Recognition

3Baseline

مسئله بازشناسی کلمه دست‌نویس می‌تواند به دو گروه اصلی به نام‌های تشخیص برخط[1] و برون‌خط[2] مطابق با فرمت کلمه دست‌نویس تقسیم شود. ورودی سیستم بازشناسی برخط به صورت پیوسته هم‌زمان با نوشتن نویسنده بر روی یک صفحه رقمی کننده[3] به سیستم وارد می‌شود. برای بازشناسی دست‌نوشته برخط از ویژگی‌هایی مانند مختصات مکانی حرکت قلم روی صفحه، سرعت نوشتن و میزان فشار قلم بر صفحه استفاده می‌شود. ورودی سیستم بازشناسی برون‌خط تصویر اسکن شده از کلمه دست‌نویس است و از آنجایی که فقط بر مبنای تصویر است، نسبت به سیستم‌های برخط مشکل‌تر است[1].

این پایان‌نامه برروی سیستم‌های بازشناسی کلمات دست‌نوشته برون‌خط متمرکز شده است.

1-2مشخصات دست‌نوشته فارسی

متن فارسی به طور ذاتی در دست‌نوشته و فرم‌های چاپ شده پیوسته است و به صورت افقی از راست به چپ نوشته می‌شود. نوشتن فارسی از لحاظ ساختاری خیلی شبیه عربی است. بنابراین یک شناساگرکلمه فارسی, برای بازشناسی کلمات عربی هم می تواند استفاده شود.

تنها تفاوت بین دست‌خط فارسی و عربی در مجموعه حروف است. مجموعه حروف فارسی در جدول 1 نشان داده شده است. همه 28 حرف عربی را شامل می شود (بعلاوه 4 حرف اضافه که در جدول 1-1 با*علامت گذاری شده) یک حرف فارسی به عنوان یک حرکت اصلی (واحد) نوشته می شود و در بیشتر موارد با حرکت‌های تکمیلی دیگر مانند نقاط و خطوط زیکزاک کامل می‌شود. بیشتر از نصف حروف فارسی (18 تا از32) نقطه دار هستند. 10 حرف 1 نقطه دارند. 3 حرف دو نقطه دارند و 5 حرف سه نقطه دارند این حروف یک قالب کلی واحد دارند که با حضور، نبود ویا تعداد نقاط از هم متمایز شده‌اند این مجموعه{ن،ث،ت،پ،ب}، {خ،ح،چ،ج}، {ذ،د}، {ژ،ز،ر}، {ش،س}، {ض،ص}،

[1] On line

[2]Off line

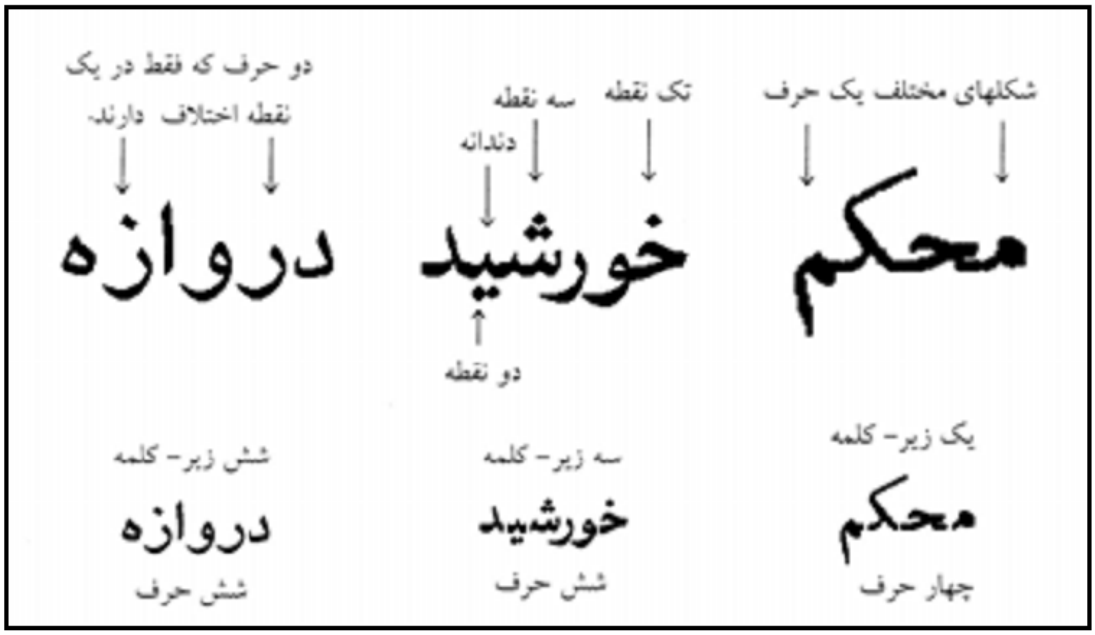
[3]Tablet Digitizer

{ظ،ط}، {غ،ع}، {ق،ف} است. بنابراین نوشتن مبهم نقاط بعضی مواقع باعث می‌شود تصویر یک کلمه در شکل‌های گوناگون با معانی کاملا متفاوت خوانده شود. برخلاف زبان انگلیسی ، حروف فارسی به دسته‌های حروف کوچک و بزرگ تقسیم نمی‌شوند در عوض حروف فارسی چندین شکل ، بسته به موقعیت قرارگیری آن‌ها در کلمه، دارند. برای مثال شکل حرفی مانند «ه» اگر در ابتدای کلمه«ھ » در وسط کلمه «ھ » یا در آخر کلمه قرار بگیرد «ه » و یا به صورت مجزا «ه» تغییرمی کند. بعضی از حروف فارسی مانند آ، د، ذ، ر، ز، و، چهار شکل مختلف ندارند، در نتیجه آن‌ها معمولا کلمه را به دو قسمت تقسیم می‌کنند[2].

بنابراین هر کلمه می‌تواند شامل یک یا چند بخش باشد که به آن‌ها زیر-کلمه[1] می‌گویند. یک زیر-کلمه از چند حرف چسبیده به هم تشکیل می‌شود. مطابق قواعد نگارش زبان فارسی باید بین کلمات فاصله وجود داشته باشد و بین زیر-کلمات فاصله اضافه وجود نداشته باشد (شکل1-1).

در برخی از شیوه‌های نوشتاری زبان فارسی، دو یا چند حرف کنار هم می‌توانند به گونه‌ای با هم ترکیب شوند که شکل حاصل شباهتی به حروف تشکیل دهنده آن‌ها ندارد. چنین مواردی نه تنها در نوشتار دست‌نویس بلکه در متون تایپی نیز وجود دارد. متداول‌ترین ترکیب در متون تایپی، ادغام دو حرف (ل)و (ا) به صورت (لا) است. در نوشته‌های دست‌نویس فارسی نیز بخاطر زیبایی بصری نوشتار وهمچنین سلیقه نویسنده، شکل برخی از حروف کنار هم، به کلی تغییر می‌کند.(شکل1-2)[1].

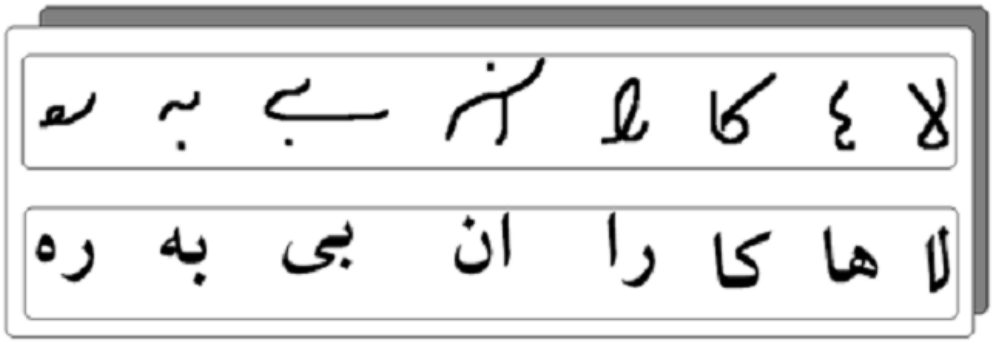
[1]Sub-word



شکل1-1: برخی ویژگی های نگارش فارسی[1]

جدول1-1: الفبای فارسی[2]

N0	Character	Isolated	First	Middle	Last
1	Alef	ا(آ)	ا(آ)	ـا	ـا
2	Be	ب	بـ	ـبـ	ـب
3	* Pe	پ	پـ	ـپـ	ـپ
4	Te	ت	تـ	ـتـ	ـت
5	Se	ث	ثـ	ـثـ	ـث
6	Jim	ج	جـ	ـجـ	ـج
7	* Che	چ	چـ	ـچـ	ـچ
8	He	ح	حـ	ـحـ	ـح
9	Khe	خ	خـ	ـخـ	ـخ
10	Dal	د	د	ـد	ـد
11	Zal	ذ	ذ	ـذ	ـذ
12	Re	ر	ر	ـر	ـر
13	Ze	ز	ز	ـز	ـز
14	* Zhe	ژ	ژ	ـژ	ـژ
15	Sin	س	سـ	ـسـ	ـس
16	Shin	ش	شـ	ـشـ	ـش
17	Sad	ص	صـ	ـصـ	ـص
18	Zad	ض	ضـ	ـضـ	ـض
19	Ta	ط	طـ	ـطـ	ـط
20	Za	ظ	ظـ	ـظـ	ـظ
21	Ayn	ع	عـ	ـعـ	ـع
22	Ghayn	غ	غـ	ـغـ	ـغ
23	Fe	ف	فـ	ـفـ	ـف
24	Ghaf	ق	قـ	ـقـ	ـق
25	Kaf	ک	کـ	ـکـ	ـک
26	* Gaf	گ	گـ	ـگـ	ـگ
27	Lam	ل	لـ	ـلـ	ـل
28	Mim	م	مـ	ـمـ	ـم
29	Noon	ن	نـ	ـنـ	ـن
30	Waw	و	و	ـو	ـو
31	He	ه	هـ	ـهـ	ـه
32	Ye	ی	یـ	ـیـ	ـی



شکل1-2: نمونه هایی از ادغام حرف مجاور در متون دست نویس[1]

1-3روش شناسی[1]

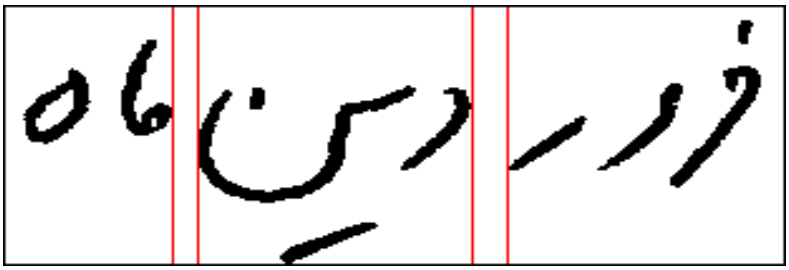
روش‌های بازشناسی متون دست‌نویس و تایپ شده به دو گروه تقسیم بندی می‌شوند :

1- روش کلی‌نگر[2] که شناسایی به صورت کلی روی همه نمایش کلمه انجام می‌شود، بنابراین نیازی به تقسیم کردن یک کلمه به کاراکترهای مجزا نیست اما لازم است که ما بتوانیم یک خط متن را به کلمات تقسیم کنیم، که این همیشه ممکن نیست زیرا فضای داخل کلمه[3] بعضی مواقع بزرگتر از فضای بین کلمه[4] است. (شکل1-3)

2- روش تحلیلی[5] که در آن کلمه به طور صریح[6] یا ضمنی[7] تقسیم بندی می‌شود. در روش تقسیم بندی صریح تلاش برای جداکردن حروف تنها ، که آن‌ها به طور مجزا شناسایی می-شوند انجام می‌شود. اما در روش ضمنی تصویر متن(خط یا کلمه) به یک دنباله از واحدهای کوچک(یک دنباله از مشاهدات) تبدیل می‌شود و شناسایی در یک سطح متوسط نسبت به کلمه یا حرف انجام می‌شود.

---

[1]Methodology
[2]Holistic
[3]Intra-word space
[4]Inter -word space
[5]Analytical Strategy
[6]Explicit
[۷]Implicit



شکل1-3: فضای بین کلمه‌ای[3]

هر روش کلی‌نگر از یک واژه‌نامه[1]، یک لیست از تفسیرهای مجاز از تصویر کلمه ورودی، استفاده می-کند. معمولا نرخ خطا در روش کلی‌نگر با اندازه واژه نامه افزایش می‌یابد زیرا تعداد بالای کلاس‌ها احتمال اشتباه طبقه‌بندی شدن[2] را افزایش می‌دهد. با داشتن یک واژه‌نامه بردار ویژگی استخراج شده از کلمه ورودی با بردارهای ویژگی مدل کلمات در واژه‌نامه مقایسه می‌شود و ورودی که بیشترین رتبه (کمترین فاصله یا بیشترین احتمال) را دارد به عنوان کلمه تعیین کننده کلاس ورودی فرض می‌شود. روش کلی‌نگر زمانی که واژه‌نامه کوچک باشد موفق خواهد بود. زمانی که اندازه واژه‌نامه افزایش می‌یابد تعداد کلمات سازگار برای یک تصویر ورودی افزایش می‌یابد و طبقه‌بندی صحیح مشکل می‌شود. بنابراین روش کلی‌نگر به مسائلی مانند تشخیص آدرس‌های پستی یا خواندن چک‌های بانکی که واژه‌نامه محدود شده و کوچک است اعمال می‌شود.

در سیستم‌هایی که در ارتباط با یک واژنامه بزرگ می‌باشند برای کاهش پیچیدگی محاسباتی و افزایش دقت و سرعت، تعداد کلمات مورد بررسی در فرآیند بازشناسی کلمه ورودی را کاهش می‌دهند که به آن کاهش واژه‌نامه[3] یا هرس کردن[4] گفته می‌شود. برای مثال زمانی که تصویر کلمه ورودی نقطه ندارد تمام لغات بدون نقطه از واژنامه برای بررسی استخراج می‌شود.

[1]Lexicon
[2]Misclassification
3Lexicon reduction
[4]Pruning

روش‌های تحلیلی به صورت تقسیم بندی صریح و ضمنی روی تصاویر ورودی انجام می‌شود. در روش جداسازی ضمنی کلمات به واحدهای کوچک تقسیم بندی می‌شود، سپس به یک دنباله از بردارهای مشاهدات تبدیل می‌شود هر واحد معمولا یک بخش از حرف است. در مقابل در قطعه بندی صریح کلمات به طور مستقیم به حروف جدا تبدیل می‌شود .

ازآنجایی که حروف فارسی/ عربی در یک فونت چاپی یا دست‌نوشته طول‌ها و شکل‌های متفاوتی ، بسته به موقعیت آن‌ها در کلمه، دارند به راحتی نمی‌توان کلمه را دقیقا به حروف جداسازی کرد. بنابراین قطعه بندی صریح بیشتر در معرض خطا است و تقریبا همه سیستم‌های بازشناسی دست‌نوشته از قطعه بندی ضمنی استفاده می‌کنند. [3]

اکثر روش‌های بازشناسی که برای زبان‌های لاتین و چینی دقت بالایی دارند، به دلیل تفاوت‌های زبان فارسی و عربی مناسب برای این زبان‌ها نمی‌باشند. بازشناسی متون فارسی و عربی به مراتب پیچیده‌تر از بازشناسی متون لاتین است. کارهای انجام شده در زمینه بازشناسی متون دست‌نوشته عربی و فارسی به دلیل این پیچیدگی‌ها هنوز به درصد قابل قبولی برای تجاری شدن نرسیده است.

این پایان‌نامه برروی روش‌های کلی‌نگر بازشناسی کلمات دست‌نوشته برون‌خط متمرکز شده است.

فصل دوم:

مروری بر فعالیت‌های انجام شده

فصل دوم

در این فصل تحقیقات انجام شده در زمینه بازشناسی برون‌خط کلمه دست‌نوشته مبتنی بر شکل کلی کلمات در زبان‌های لاتین ، عربی و فارسی بررسی می‌شود.

در حالت کلی یک سیستم بازشناسی کلمه دست‌نوشته شامل مراحل پیش پردازش ، استخراج ویژگی و طبقه بندی است.

هدف پیش‌پردازش حذف تمام عناصری است که برای فرایند بازشناسی مفید نیست. معمولا پیش پردازش شامل عملیات باینری کردن[1]، هموار سازی[2]، تخمین خط کرسی[3] و نازک‌سازی[4] است.[2]

ویژگی‌های استخراج شده می‌تواند به دو صورت 1- ویژگی‌های سراسری 2- ویژگی‌های محلی باشد.

ویژگی‌های سراسری به دو صورت ساختاری و آماری هستند. برای مثال تعداد مولفه‌های متصل، حفره‌ها، حروف بالای خط[5] و پایین خط[6] ویژگی‌های سراسری ساختاری هستند. ضرایب تبدیل فوریه و گشتاورهای ثابت مثال‌هایی از ویژگی‌های سراسری آماری هستند. ویژگی‌های ساختاری به سبک نوشتن وابسته نیستند و می‌توانند یک درجه بالایی از تغییرات را تحمل کنند، اما در برابر نویز مقاوم نیستند و استخراج آن‌ها مشکل است. ویژگی‌های آماری راحت‌تر استخراج می‌شوند وکلمه را بصورت دقیق‌تر بیان می‌کنند، اما آن‌ها مانند ویژگی‌های ساختاری نسبت به تنوع نوشتار در خط پیوسته پایدار نیستند.

---

[۱] Binarization
[۲] Smoothing
[۳] Baseline estimation
[4] Thining
[۵] Ascender
[6] Descender

برای استخراج ویژگی‌های محلی ابتدا تصویر کلمه را به بلوک‌های کوچکتر تقسیم می‌کنند و از هر یک از این بلوک‌ها، ویژگی‌های لازم استخراج می‌شود. برای مثال درصد پیکسل‌های پیش‌زمینه در پنجره، آمارهای انتقال از پیش‌زمینه به پس‌زمینه ، ناحیه‌های بالای خط و پایین خط به عنوان ویژگی‌های محلی استفاده می‌شود[3].

در مرحله طبقه بندی ویژگی های تصویر کلمه ورودی استخراج شده، با نماینده کلاس‌ها در فرهنگ لغت مقایسه می‌شود و الگویی با بیشترین شباهت (کمترین فاصله یا بیشترین احتمال) انتخاب می‌شود. عمل مقایسه توسط یک کلاسه‌بند انجام میشود. از جمله کلاسه‌بندهای پرکاربرد در این زمینه K نزدیک‌ترین همسایه(KNN)، شبکه های عصبی[1](NN) ، مدل مخفی مارکوف[2](HMM)، ماشین بردار پشتیبان[3] (SVM)،کلاسه بندی بیزین[4] و ترکیبی از روش‌ها است.

الگوریتم پیشنهاد شده در [4] یک سیستم بازشناسی کلمه دست‌نوشته روی یک پایگاه داده از 30 اسم رایج زبان فارسی است، که در آن شبکه عصبی RBF[5] به عنوان کلاسه‌بند به‌کار رفته و برای یافتن معماری بهینه در این شبکه عصبی از الگوریتم ژنتیک استفاده شده است. در این روش ویژگی‌ها از تصاویر پروفایل استخراج می‌شود. پروفایل‌های استخراج شده برای هر تصویر کلمه در 4 جهت (بالا، پایین، چپ و راست) است. دلیل انتخاب این ویژگی ساختار ویژه متن فارسی است. در نوشتن فارسی ساختار حروف پروفایل‌های متفاوتی در کلمات مختلف تولید می‌کنند. بنابراین پروفایل‌ها تقریبا حاوی تمام ویژگی‌های یک کلمه و حروف آن است. پس از استخراج پروفایل‌ها از تبدیل ویولت گسسته یک

[1] Neural Network
[2] Hidden Markov Model
[3] Support Vector Machines
[4] Bayesian Classification
[5] Radial Basis Functions

بعدی(DWT[1]) برای هموار سازی و کاهش ابعاد ویژگی‌های استخراج شده استفاده می‌شود. این ویژگی‌ها به عنوان ورودی شبکه عصبی RBF استفاده می‌شود. نرخ بازشناسی این سیستم در انتخاب اول 96% و در 2 انتخاب اول 97% گزارش شده است.

درتحقیق[5] برای بازشناسی کلمه دست‌نوشته زبان اردو، پس از انجام پیش‌پردازش‌هایی مانند حذف نویز و نرمال کردن ابعاد تصویر دو نوع ویژگی استخراج می‌شود:

1- ویژگی گرادیان: یک ویژگی جهتی است که از گرادیان تصویر سطح خاکستری بدست می‌آید. دراین تحقیق ابتدا اپراتور روبرتز به تصویر اعمال می‌شود و پس از بلوک‌بندی تصویر، در هر بلوک طول گرادیان در جهت‌های کوانتیزه شده با هم جمع می‌شود.
2- ویژگی ساختاری: ویژگی‌های ساختاری شامل پروجکشن پروفایل[2]، ویژگی‌های توپولوژیکی و پروفایل‌های بالایی و پایینی است. در این تحقیق ویژگی پروفایل بالایی که ساختار بیرونی کلمه را نشان می‌دهد استفاده شده است.

با استفاده از کلاسه بند SVM برای یک پایگاه داده با 57 کلمه از زبان اردو نرخ بازشناسی 97% گزارش شده است.

در [6] بعد از باینری کردن و نرمال‌سازی تصویر، عملگر پرویت[3] به تصویر کلمه دست‌نوشته اعمال می‌شود. برای استخراج ویژگی ابتدا تصویر را به بلوک‌هایی با اندازه 12 پیکسل و طول همپوشانی 2 پیکسل تقسیم می‌کنند. بردار ویژگی با محاسبه قدرمطلق اندازه میانگین همه ضرایب از هر بلوک در تصویر کلمه بدست می‌آید.

[1]Discrete Wavelet Transform
[2]profile projection
[3]Prewitt

برای طبقه بندی در این تحقیق از روش KNN[1] استفاده شده است، KNN یک الگوریتم یادگیری ماشین سریع است که برای طبقه بندی مجموعه تست از یک مجموعه آموزش برچسب زده شده استفاده می‌شود. برای این منظور تصویر یک کلمه از مجموعه تست با ویژگی‌های آموزش دیده بر مبنای فاصله مقایسه می‌شود و پیش‌بینی کلاس بر مبنای کمترین فاصله اقلیدسی بدست می‌آید. در واقع الگوریتم KNN نمونه‌های نزدیک را به عنوان مقدارهای پیش‌بینی برای مجموعه تست می‌گیرد. از پایگاه داده عربی IFN/ENIT برای ارزیابی این سیستم استفاده شده است و نرخ بازشناسی 76% گزارش شده است.

آقای دهقان و همکارانش در [7] ابتدا کدهای زنجیره‌ای را روی کانتور کلمه بدست می‌آورند. این روش تصویر کلمه را در راستای افقی به پنج قسمت مساوی تقسیم می‌کند سپس یک نوار باریک به اندازه دو برابر عرض قلم در تصویر با همپوشانی 50% روی تصویر حرکت می‌دهد. بردار هیستوگرام‌های جهتی کدهای زنجیره‌ای در هر کدام از این نوارها به عنوان بردار ویژگی انتخاب می‌شود. برای کم کردن ابعاد بردار مشاهده در آموزش مدل مخفی مارکوف گسسته، فضای ویژگی باید به یک مجموعه از بردارهای کدکلمه[2] کوانتیزه شود. از شبکه عصبی خودسامان کوهنن[3] برای تولید بردارهای کتاب‌رمز[4] استفاده می‌شود. با استفاده از الگوریتم بام-ولش[5] برای هر کلمه یک HMM گسسته آموزش داده می‌شود. در مرحله تست کلمه‌ها برحسب بیشترین احتمال دنباله مشاهده تصویر تست با هرکدام از کلاس‌ها مرتب می‌شوند. حدود 17000 تصویر از اسامی دست‌نویس 198 شهر ایران به عنوان پایگاه داده برای این سیستم استفاده شده است و موفقیت این روش در انتخاب اول 65% و در 20 انتخاب اول 95% گزارش شده است.

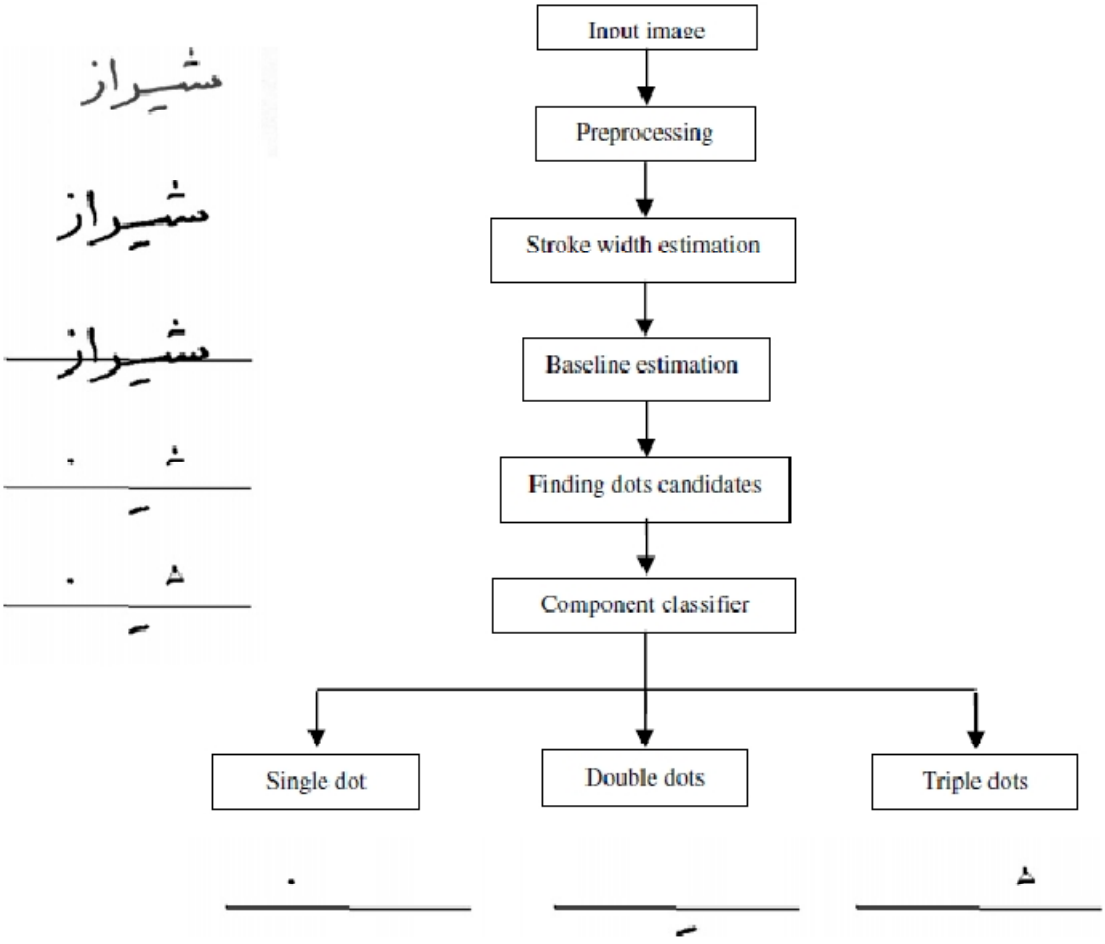
---

[1] K-Nearest Neighbor
[2] Codeword
[3] Kohonen self organizing
[4] Codebook
[5] Baum Welch Algorithm

در مرجع[8] برای تولید کتاب‌رمز فازی ، از خوشه‌بندیFuzzy c-means استفاده شده است. نرخ بازشناسی در این روش 67/18% بدست آمده است.

آقای مظفری و همکارانش در [2] یک مفهوم جدید از وقوع، تکرار و موقعیت نقاط نسبت به خط کرسی، محل اتصال حروف در کلمه، برای کاهش واژه‌نامه دست‌نوشته فارسی/ عربی معرفی کرده اند. شکل 2-1 بلوک دیاگرام الگوریتم استخراج نقطه پیشنهاد شده در این تحقیق را نشان می‌دهد.

ابتدا خط کرسی بر مبنای رگرسیون خطی دو مرحله‌ای روی نقاط مینیمم محلی کانتور کلمه تعیین می‌شود. به منظور متمایز کردن نقاط از بقیه کلمه، اندازه عناصر کلمه تعیین می‌شود. نقاط معمولا نسبت به حروف اندازه کوچکتری دارند بنابراین عناصر بزرگتر از اندازه میانگین نامزد نقطه در نظر گرفته نمی‌شوند. جدول(2-1) تنوع نوشتن نقاط را نشان می‌دهد.



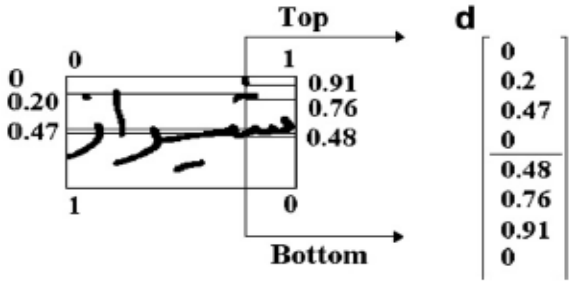
شکل 2-1: خلاصه الگوریتم استخراج نقطه پیشنهاد شده در[2]

جدول2-1: تنوع نقاط در نوشتار فارسی[2]

	Printed	Handwritten
Single dot	•	•
Double dot	••	
Triple dot	∴ ∵	

پس از استخراج و طبقه بندی نقاط در هر کلمه با در نظر گرفتن تعداد و موقعیت آن‌ها نسبت به خط کرسی یک کد توصیفگر برای هر کلمه ایجاد می‌شود . با استفاده از روش تطبیق دمرا-

لونشتاین[1] کد توصیفگر کلمه ورودی با تمام کدهای توصیفگر کلمات موجود در فرهنگ لغت مقایسه می‌شود و k تا از بهترین کلمات نامزد انتخاب می‌شود. در مرحله آموزش برای هر کلمه موجود در فرهنگ لغت یک مدل گسسته HMM با استفاده از ویژگی مکان و تعداد انتقال‌ها از پیکسل سیاه به سفید را در هر ستون، از بالا به پایین واز پایین به بالا محاسبه می‌کند(شکل2-2) .در مرحله بازشناسی ویژگی‌های استخراج شده از تصویر تست تنها با مدل HMM کلماتی مقایسه می‌شود که در مرحله کاهش فرهنگ لغت انتخاب شده‌اند. در این تحقیق از 17000 تصویر از اسامی 200 شهر ایران برای ارزیابی استفاده شده است و نرخ بازشناسی در انتخاب اول بدون کاهش فرهنگ لغت 60/2% و با کاهش فرهنگ لغت 73/61% گزارش شده است.

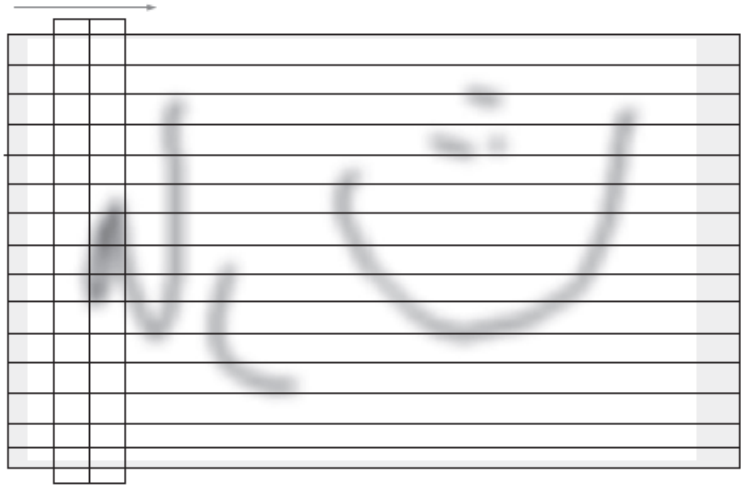


شکل2-2: نحوه استخراج ویژگی انتقال[2]

آقای Alkhateeb و همکارانش در [9] یک سیستم بازشناسی کلمات دست‌نویس عربی را با استفاده از HMM معرفی می‌کنند. در مرحله پیش پردازش تصاویرکلمات در راستای عمودی و افقی نرمالیزه شده است. ابتدا تصویر را به 15 نوار افقی تقسیم می‌کند. از حرکت نوار باریک عمودی با عرض 3پیکسل و همپوشانی 1پیکسل بر روی تصویر آینه‌ای کلمه برای استخراج بردارهای ویژگی استفاده کرده است (شکل2-3). میانگین روشنایی سلول های ایجاد شده در تصویر آینه‌ای شده به عنوان ویژگی در این سیستم استفاده شده است. علاوه بر این چندین

[1] Demerau levenshtein

ویژگی شبه ساختاری شامل تعداد ناحیه‌های متصل ، تعداد ناحیه های متصل زیر خط کرسی و تعداد ناحیه های متصل بالای خط کرسی استخراج شده است.



شکل2-3: ناحیه های مورد استفاده برای استخراج ویژگی و نوار متحرک[9]

با استفاده از ویژگی‌های استخراج شده یک روش ترکیبی برای بازشناسی با استفاده از HMM به عنوان کلاسه بند پایه و پیرو آن یک روش re-ranking با استفاده از ویژگی‌های ساختاری معرفی شده است. بعد از آموزش HMM ها در مرحله تست برای هر تصویر نسبت به تمام کلاس‌ها احتمال‌هایی بدست می‌آید بجای در نظر گرفتن بزرگترین احتمال به عنوان بهترین نتیجه، با استفاده از ویژگی‌های ساختاری تصویر تست، (n_a, n_b, n_c)، و میانگین و انحراف معیار ویژگی‌های ساختاری در هر کلاس کلمه، ($n_{t,c}$ و $\sigma_{t,c}$) ، احتمال‌ها به صورت زیر تصحیح می‌شود:

$$p'_c = p_c \prod_{t=a,b,r} R_t(n_t, c) \quad (1\text{-}2)$$

$$R_t(n_t, c) = \exp\left(-\left(\frac{n_t - \bar{n}_{t,c}}{\sigma_{t,c}}\right)^2\right) \quad (2\text{-}2)$$

در این مقاله برای ارزیابی عملکرد سیستم پیشنهاد شده از پایگاه داده IFN/ENIT استفاده شده است. نرخ بازشناسی با استفاده از re-ranking در انتخاب اول89/24% و بدون استفاده از آن 86/73% گزارش شده است.

در مرجع [10] یک سیستم بازشناسی دست‌نوشته با استفاده از روش تحلیلی معرفی شده است. در این تحقیق از کلاسه‌بند HMM برای بازشناسی استفاده شده است. برای تولید دنباله بردارهای ویژگی، تصویر کلمه به نوارهای باریک عمودی با عرض 8 پیکسل، دارای همپوشانی تقیسم شده و هر نوار به سلول هایی با ارتفاع 4 پیکسل تقسیم بندی می‌شود (شکل2-4)، از هر فریم 24 ویژگی استخراج شده است.

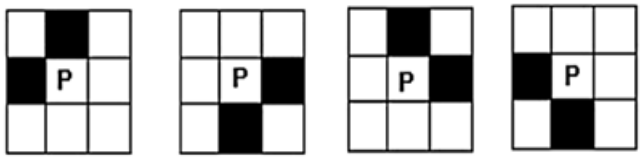
شکل2-4: فریم بندی تصویر برای استخراج ویژگی [11]

این ویژگی‌ها دو نوع هستند :

1. ویژگی‌های توزیعی[1]: این ویژگی‌ها بیشتر بر مبنای چگالی پیکسل‌های پیش‌زمینه در هر فریم، تعداد پیکسل‌های پیش‌زمینه در هر ستون از فریم، موقعیت عمودی مرکز ثقل پیکسل‌های پیش‌زمینه در هر فریم نسبت به خط‌کرسی پایینی وچگالی پیکسل‌های پیش‌زمینه، بالا و پایین خط‌کرسی پایینی است.

[1]Distribution

2. ویژگی‌های تقعر[1]: ویژگی‌های تقعر، اطلاعات تقعر محلی و جهت حرکت در هر فریم را مشخص می‌کند. هرکدام از ویژگی‌های تقعر نشان‌دهنده تعداد پیکسل‌های پس‌زمینه است که متعلق به یکی از انواع پیکربندی تقعر است که با استفاده یک پنجره 3×3 ، همانطور که در شکل 2-5 نشان داده شده است، بدست می‌آید.



شکل2-5: چهارنوع پیکربندی تقعر برای پیکسل پس‌زمینه P [12]

همچنین به طور جداگانه این چهار ویژگی تقعر در فضای بین خط کرسی پایینی و بالایی بدست آمده است.

با استفاده از ویژگی‌های استخراج شده در مرحله استخراج ویژگی، برای هر حرف در واژه نامه یک HMM پیوسته آموزش داده می‌شود. از آن‌جایی که شکل حروف با موقعیت قرارگیری آن‌ها در کلمه تغییر می‌کند، برای هر کدام از شکل‌های حروف مدل‌های مجزایی آموزش داده شده و مدل HMM هر کلمه با به هم پیوستن مدل HMM حروف تشکیل دهنده‌اش تولید می‌شود. در این تحقیق برای ارزیابی روش، سیستم دوبار بر روی پایگاه داده IFN/ENIT آزمایش شده است، بار اول فقط از ویژگی‌هایی که مستقل از خط کرسی هستند و نرخ بازشناسی حدود 75% گزارش شده است و باردوم از تمام ویژگی‌های استخراج شده استفاده شده است و نرخ بازشناسی حدود 86% بدست آمده است.

[1] Concavity

در[11] یک سیستم بازشناسی کلمه دست‌نوشته فارسی معرفی شده است. در این سیستم از ویژگی‌های مشابه ویژگی‌های استخراج شده در [10] استفاده می‌کند. به منظور آموزش مدل گسسته HMM برای هر کلمه در پایگاه داده، ابتدا بردارهای ویژگی استخراج شده با استفاده از الگوریتم خوشه‌بندی K-means، به دنباله مشاهدات تبدیل می‌شود. برای ارزیابی این سیستم از پایگاه داده CENPARMI ، شامل70 کلمه که از هر کلمه 516 تصویر دست‌نوشته در پایگاه داده موجود است، استفاده شده است و نرخ بازشناسی96% گزارش شده است.

فصل سوم:

مباحث تئوری

فصل سوم

در این فصل مباحث تئوری که در فصل‌های بعد استفاده شده است، ارائه می‌شود. روش های پیش پردازش[1]، استخراج ویژگی[2]، کوانتیزه سازی بردارها[3] و کلاسه‌بند HMM در این فصل بررسی می‌شوند.

3-1 پیش پردازش

پیش پردازش یکی از قسمت‌های مهم و اساسی در هر سیستم پردازش الگو است، مهم از این منظر که عدم اعمال روش‌های مناسب در این مرحله باعث انتشار خطا در کل سیستم شده و کارایی سیستم را در نهایت تحت تاثیر قرار می‌دهد.

هدف اصلی پیش پردازش ارتقا کیفیت تصویر است و روش های مختلفی برای این کار وجود دارد که متناسب با کاربرد استفاده می‌شود. در زیر مهم‌ترین کارهایی که در مرحله پیش پردازش مورد نظر است، آورده شده است.

3-1-1 استخراج خط کرسی

به دلیل نگارش پیوسته زبان فارسی تمام حروف کلمه روی یک خط فرضی به یکدیگر متصل می‌شوند. این خط مجازی که کلمه روی آن نگارش می‌شود خط کرسی نامیده می‌شود. استخراج خط کرسی می‌تواند برای اصلاح چرخش[4]، جداکردن نقاط و علائم و ... به کار برده شود.

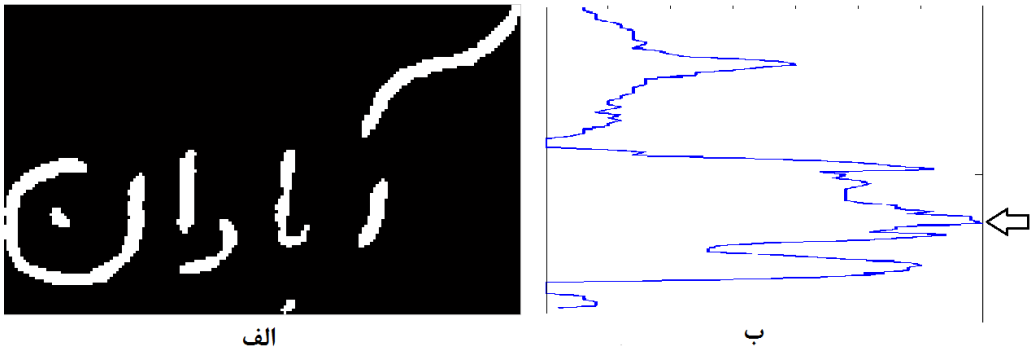
برای استخراج خط‌کرسی، پروجکشن پیکسل‌های پیش زمینه به صورت افقی محاسبه می‌شود (در هر سطر تعداد پیکسل‌های پیش زمینه بدست می‌آید) و در نهایت موقعیت خط کرسی با یافتن بیشترین مقدار این پروجکشن بدست می‌آید. در شکل 3-1 موقعیت خط کرسی با استفاده از این روش بدست

---

[1] Preprocessing
[2] Feature extraction
[3] Vector quantization
[4] Rotation

آمده است. این روش برای کلمه‌های با طول زیاد تخمین خوبی از موقعیت خط کرسی می‌دهد و به دلیل پیچیدگی محاسباتی کم و سرعت بالا کاربرد فراوانی دارد[12].

از آنجایی که در بیشتر موارد خط کرسی در نیمه پایینی تصویر قرار دارد، در بعضی موارد برای استخراج خط کرسی با روشی که بیان شد بیشترین مقدار هیستوگرام را در نیمه پایینی تصویر در نظر می‌گیرند[9].



شکل 3-1: نحوه تخمین مکان خط کرسی، (الف) تصویر کلمه، (ب) هیستوگرام افقی کلمه[12]

3-1-2 یکسان سازی عرض حرکت[1]

در مرحله استخراج ویژگی نیاز است که عرض پیکسل‌های پیش زمینه یکسان باشد لذا در این مرحله دو روش برای اصلاح عرض قلم معرفی شده است:

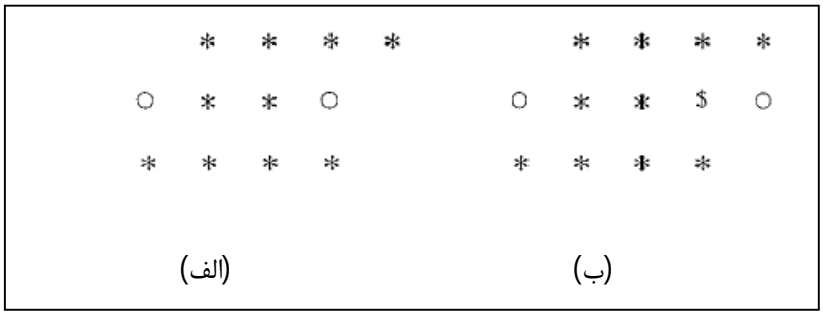
3-1-2-1 استفاده از همسایگی پیکسل های پیش‌زمینه[13]

هدف از یکنواخت سازی عرض قلم این است که در جهت افقی و عمودی طول حرکت پیکسل‌های پیش زمینه در تصویر بزرگتر از 3 باشد. اگر امتداد طول پیکسل‌های پیش‌زمینه کمتر از چهار پیکسل باشد، دو پیکسل مجاور به امتداد پیکسل‌های پیش‌زمینه، نقطه‌های نامزد برای پر شدن هستند. برای انتخاب اینکه کدام نقطه پرشود، هشت همسایگی هر نامزد بررسی می‌شود. اگر یک نقطه تعداد

[1]Stroke width

همسایگی پیش‌زمینه بیشتری داشته باشد آن نقطه پر خواهد شد. برای مثال در شکل3-2 نقطه‌ای که سمت راست امتداد سیاه قرار دارد 6 همسایه پیش‌زمینه و نقطه سمت چپ 4 همسایه دارد. بنابراین نقطه سمت راست پر می‌شود.

برای ساده سازی مسئله، از یک وزن برای اندازه گیری اهمیت هر نقطه نامزد استفاده می‌شود و نقطه با وزن بزرگتر پر می‌شود.



شکل3-2: (الف) شکل اصلی، (ب) شکل اصلی بعد از یک مرحله تصحیح عرض حرکت [13]

وزن یک نقطه نامزد با افزودن تاثیر پیکسل‌های پیش‌زمینه در 8 همسایه آن بدست می‌آید. این مقدار وزن برای همسایه‌ها در چهار جهت اصلی بیشتر از همسایه‌های قطری است. و برای پیکسل‌های پر شده نصف پیکسل‌های پیش زمینه است.

3-1-2-2 استفاده از اسکلت‌بندی[1] و انبساط[2][11]

هدف اصلی اسکلت بندی، استخراج ویژگی برمبنای شکل کلی یک شی، بدون تغییر ساختار آن شی است. برای اجتناب از سروکار داشتن با عرض قلم غیر یکنواخت در تصویر کلمه، اسکلت تصویر استخراج می‌شود. در شکل3-3(ب) اسکلت استخراج شده از تصویر کلمه نمایش داده شده است.

[1] Skeletonization
[2] Dilation

اسکلت تصویر کلمه برای یکسان سازی عرض حرکت انبساط داده می‌شود. عملگر مرفولوژی[1] انبساط، دو داده را به عنوان ورودی می‌گیرد، اولین داده تصویری است که باید انبساط داده شود و دومی یک مجموعه از نقاط مختصات است که به عنوان کرنل[2] شناخته می‌شود. این کرنل تعیین‌کننده مقدار و جهت انبساط در تصویر اصلی است.

ماتریس K به عنوان کرنل برای انبساط تصویر کلمه به کار می‌رود. با این اقدام عرض قلم در تصویر کلمه به اندازه 4 پیکسل می‌شود. شکل3-3(ج) تصویر کلمه بعد از تصحیح عرض قلم را نشان می‌دهد.

$$K = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (1\text{-}3)$$

(الف) (ب) (ج)

شکل3-3: (الف) تصویر اصلی، (ب) اسکلت تصویر، (ج) اسکلت منبسط شده

3-2 استخراج ویژگی

هدف از استخراج ویژگی بدست آوردن توصیفگری مناسب برای توصیف تصویر است. این توصیفگر به عنوان ورودی برای کلاسه‌بند استفاده می‌شود. کارایی کلاسه‌بندها تا حدود زیادی به کیفیت و نوع ویژگی‌های استخراج شده مرتبط است[11]، از طرفی در مقایسه با زمانیکه کل پیکسل‌های تصویر به عنوان ورودی استفاده شود، استخراج ویژگی‌های مناسب سبب کاهش بار محاسباتی کلاسه‌بند نیز می‌شود.

[1]Morphology
[2]Kernel

استخراج ویژگی برای حذف زوائد از داده و بدست آوردن ضروریترین اطلاعات از تصویر خام است[9].

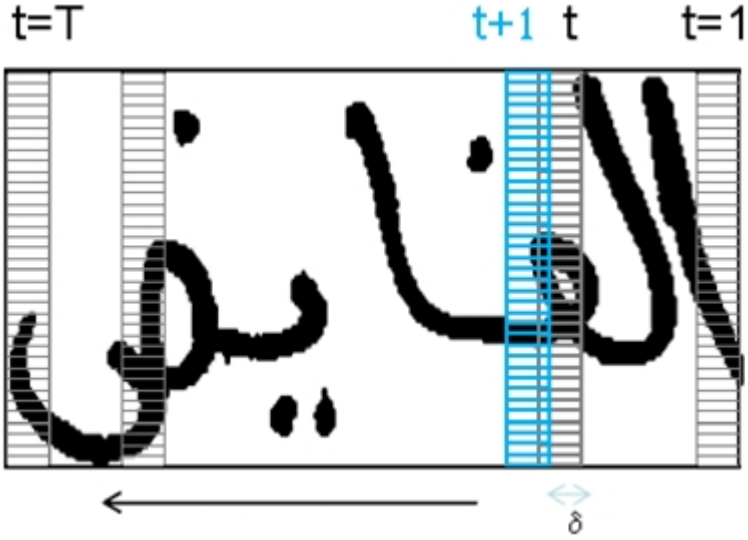
نحوه بدست آوردن این ویژگی‌ها به کلاسه‌بند مورد نظر بستگی دارد. برای بعضی از کلاسه‌بندها مانند HMM ، لازم است یک دنباله از بردارهای ویژگی بدست آید. در حالی که برای بعضی دیگر از کلاسه‌بند ها ممکن است یک بردار ویژگی مورد نیاز باشد.

3-2-1 سیستم پنجره لغزان[1]

اساس سیستم‌های پنجره لغزان بر مبنای استخراج یک دنباله از بردارهای ویژگی مشاهده شده با حرکت یک پنجره باریک از راست به چپ ، روی تصویرکلمه است. با توجه به شکل3-4 پنجره از اولین مکان در کلمه (در موقعیت t=1) شروع می‌کند و به طور منظم به اندازه δ انتقال داده می‌شود تا زمانی که t=T شود و در هر مکان یک بردار ویژگی استخراج می‌شود. همپنین می‌توان پنجره را به ناحیه های کوچکتر تقسیم کرد [11]. همانطور که در شکل دیده می‌شود در هر انتقال پنجره لغزان نسبت به مکان قبلی آن می‌تواند همپوشانی[2] وجود داشته باشد. عرض این پنجره بسته به نوع ویژگی که از آن استخراج می‌شود می‌تواند متفاوت باشد.

در ادامه چند نمونه از ویژگی‌هایی که با روش پنجره لغزان از تصویر کلمه دست‌نوشته استخراج می‌شود معرفی شده است.

[1] Sliding-window
[2] Overlap

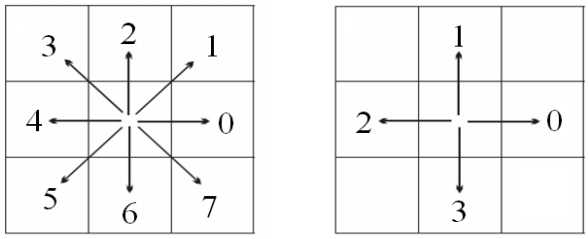


شکل3-4: حرکت نوار لغزان روی تصویر کلمه [11].

3-2-1-1کدهای زنجیره‌ای[1]

کدهای زنجیره‌ای یک روش رایج برای نمایش تصویر باینری با ترسیم خطوط یا کانتورها است. کد زنجیره‌ای یک شی را به وسیله یک دنباله از پاره‌خط‌هایی با اندازه واحد، با توجه به جهت توصیف می‌کند. فریمن[2] برای اولین بار یک کد زنجیره‌ای را معرفی کرد که حرکت در امتداد یک منحنی دیجیتال یا یک دنباله از پیکسل‌های مرز را، با استفاده از همسایگی4 تایی یا همسایگی 8 تایی، توصیف می‌کند. جهت هر حرکت با توجه به زاویه i×45 یا i×90 در جهت عکس عقربه‌های ساعت به صورت $\{i|i = 0, 1, 2, \dots, 7\}$یا $\{i|i = 0, 1, 2, 3\}$ شماره گذاری می‌شود. جهت‌های کد زنجیره‌ای در شکل3-5 نشان داده شده‌است.

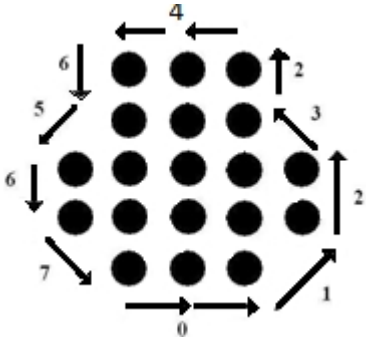
[1]Chain-code
[2]Freeman



(ب) (الف)

شکل3-5 : جهت های کد زنجیره ای پایه، (الف) برای چهار جهت، (ب) برای 8 جهت [14]

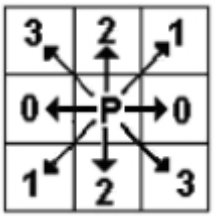
رمز زنجیره‌ای برای مرز یک تصویر باینری یک دنباله از اعداد صحیح$C = \{C_0, C_1, \dots, C_{n-1}\}$ است کهC_i، برای همسایگی 8 تایی، عددی بین 0 تا 7 است. تعداد عناصر دنباله C طول کد زنجیره‌ای را مشخص می‌کند. شکل3-6 مرز شکل را با استفاده از کد زنجیره‌ای 8 تایی نشان می‌دهد[15].



شکل3-6 : مرز با کد زنجیره‌ای 8 تایی[16]

کد زنجیره‌ای مرز، به نقطه شروع بستگی دارد. اما می‌تواند نسبت به نقطه شروع نرمال سازی شود. در واقع نقطه شروع جایی در نظر گرفته می‌شود که دنباله حاصل از اعداد، یک مقدار مینیمم صحیح را ایجاد کند. کد زنجیره‌ای نسبت به چرخش تصویر حساس است. برای رفع این مشکل می‌توان از تفاضل اول کد زنجیره‌ای به جای خود کد استفاده کرد. این تفاضل، با شمارش تعداد تغییرات جهت دو عنصر همجوار از رمز در جهت پادراستگرد، به دست می‌آید. به عنوان مثال تفاضل اول رمز زنجیره-ای 4 جهتی 1010322 برابر با 3133030 است. از سمت چپ جابجایی 1 به 0 معادل 3 تغییر، 0 به 1 یک تغییر و به همین ترتیب بقیه دنباله تفاضل می‌شوند.

هیستوگرام کدهای زنجیره‌ای به عنوان بردار ویژگی در سیستم پنجره لغزان استخراج می‌شود. ابتدا هر پنجره به پنج ناحیه تقسیم می‌شود. برای کاهش ابعاد بردار ویژگی، 4 جهت اصلی عمودی، افقی، قطری و غیر قطری طبق شکل3-7 در نظر گرفته می‌شود و از هر ناحیه 4 ویژگی هیستوگرام کد زنجیره‌ای در 4 جهت در نظر گرفته می‌شود .



شکل3-7: 4 جهت بدست آمده از 8 جهت [17]

3-2-1-2 میانگین بلوکی

برای استخراج این ویژگی در سیستم پنجره لغزان هر پنجره به چند ناحیه تقسیم می‌شود و در هر ناحیه تراکم نقاط پیش‌زمینه محاسبه می‌شود و به عنوان ویژگی میانگین بلوکی آن ناحیه در نظر گرفته می‌شود. این ویژگی می‌تواند از تصویر اصلی کلمه و یا از تصاویر گرادیان افقی و عمودی کلمه استخراج شود[18].

3-2-1-3 ویژگی‌های انتقال[1][2]

ویژگی‌های انتقال، موقعیت و تعداد انتقال‌های سفید- سیاه (پس‌زمینه به پیش‌زمینه) را محاسبه می‌کنند. در هر ستون از تصویر تعداد انتقال‌ها برای جهت‌های از بالا به پایین و از پایین به بالا محاسبه می‌شود. موقعیت هر انتقال به صورت فاصله نرمال شده نقطه انتقال نسبت به بالاترین و یا پایین‌ترین مکان تصویر، با توجه به جهت در نظر گرفته شده، معرفی می‌شود. برای داشتن بردارهای ویژگی با اندازه ثابت، یک بیشینه برای تعداد انتقال در هر ستون در نظر گرفته می‌شود، به عنوان نمونه 4 و اگر در یک ستون بیش از آن انتقال رخ داده باشد، فقط چهار انتقال اول در نظر گرفته

[1]Transition features

می‌شود و اگر کمتر از 4 انتقال در یک ستون رخ دهد مقدار صفر جایگزین انتقال‌های که وجود ندارد می‌شود.

3-2-2 ویژگی‌های ساختار- مانند[1]

برای استخراج این نوع ویژگی‌ها از تصویر کل کلمه به عنوان ورودی استفاده می‌شود. تعداد مولفه‌های متصل تصویر کلمه، تعداد مولفه‌های متصل بالای خط‌کرسی و تعداد مولفه‌های متصل پایین خط کرسی (نقاط) به عنوان ویژگی های ساختار مانند استخراج می‌شوند[9].

3-3شبکه عصبی خودسازمانده کوهنن[2] (SOM) [19]

SOM یک شبکه عصبی مصنوعی[3] با یادگیری بدون سرپرست[4] است.این شبکه شامل لایه‌های ورودی و خروجی است. تعداد نرون‌های لایه ورودی برابر با ابعاد بردار ورودی (ویژگی‌ها) است و تعداد نرون‌های لایه خروجی برابر با تعداد دسته‌هایی (تعداد خوشه هایی[5]) است که فضا به این نواحی کوانتیزه می شود. این شبکه اصطلاحا یک شبکه کاملا متصل[6] نامیده می شود، به این معنی که از هر نرون لایه ورودی به تمام نرون‌های لایه خروجی با یک وزن اتصال دارد. بنابراین برای هر نرون لایه خروجی یک بردار وزن به ابعاد تعداد نرون‌های لایه ورودی وجود دارد (شکل3-8).

[1] Structure-like

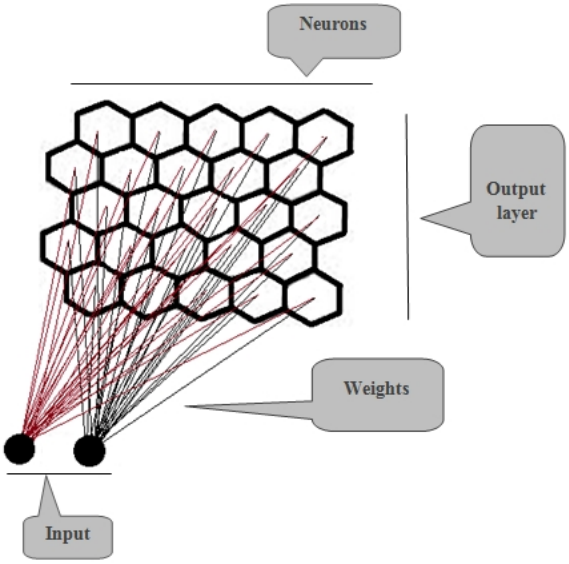
[2] Self-Organization Map (SOM kohonen)

[3] Artificial Neural Network

[4] Unsupervised

[5] Clusters

[6] Fully connected Artificial Neural Networks

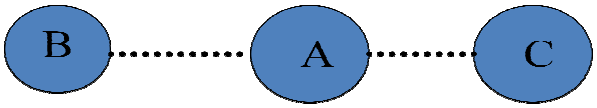


شکل3-8: ساختارشبکهSOM

این شبکه شبیه شبکه های عصبی ساده است با این تفاوت که در SOM بر خلاف شبکه های دیگر، توپولوژی[1] لایه خروجی می‌تواند یک بعدی یا دو بعدی باشد. بر اساس نتایج مطالعاتی که در زمینه نوروبیولوژی[2] انجام شده است، زمانی که در مغز یک نرون برنده[3] می‌شود نرون های همسایه این نرون تمایل دارند که بروز شوند و این تمایل با فاصله نرون از نرون برنده نسبت معکوس دارد. در شبکه SOM این نگرش شبیه سازی شده است[20].

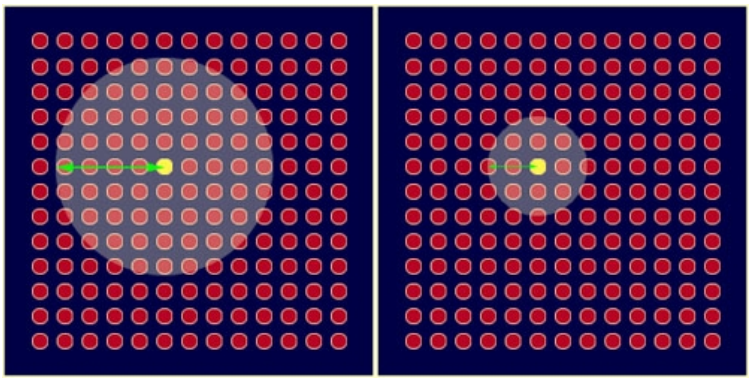
در شبکه SOM زمانی که یک نرون در تکرار[4] t ام برنده می‌شود، یک تابع $D(t)$ تعریف شده، مشخص می کند که به چه شعاع همسایگی، وزن‌های نرون‌های همسایه نرون برنده بروز رسانی می-شود. شکل 3-9 نرون های لایه خروجی یک بعدی را برای یک مثال نشان می‌دهد.

[1] Topology
[2] Neurobiology
[3] Fire
[4] Iteration



شکل3-9: نرون‌های لایه خروجی یک بعدی

که اگر $D(t) = 1$ باشد زمانی که نرونB برنده باشد وزن های ورودی به A وB بروز رسانی می‌شوند. در حالتیکه که توپولوژی خروجی دو بعدی باشد، تابع $D(t)$ شعاع همسایگی نرون خروجی را به عنوانی تابعی از زمان(تکرار) می‌دهد. معمولا شعاع ابتدا گسترده است که این باعث می شود شبکه به صورت کلی و عمومی تر [1] رفتار کند، در ادامه به تدریج شعاع همسایگی کوچکتر می‌شود که باعث می‌شود شبکه رفتار محلی تر [2] داشته باشد(شکل3-10).



شکل3-10: تغییر تدریجی شعاع همسایگی[19]

3-3-1 الگوریتم SOM :

الگوریتم آموزش نگاشت‌های خودسازمانده از نوع بدون ناظر می‌باشد. الگوریتم یادگیری نگاشت‌های خودسازمانده در حالت کلی مبتنی بر انتخاب نرون برنده و حرکت نرون مذکور و برخی از همسایگانش، بسوی داده ورودی مورد نظر است. الگوریتم یادگیری نگاشت‌های خودسازمانده را می‌توان به مراحل زیر خلاصه نمود.

[1] Global
[2] Local

- مرحله آغازین، در این مرحله وزن هر نرون بصورت یک عدد تصادفی قرار داده می‌شود، سپس یک الگوی ورودی $X = (x_1, x_2, \cdots, x_d)$ به شبکه اعمال می‌گردد.
- تعیین نرون برنده، در این مرحله بر اساس معیار تشابه شبکه، نرون برنده مشخص می‌گردد. معیارهای تشابه مختلفی را می‌توان در نگاشت‌های خودسازمانده بکارگرفت، امّا معمول‌ترین معیاری که در این‌گونه از شبکه‌ها بکارگرفته می‌شود، فاصله اقلیدسی است. رابطه معیار تشابه اقلیدسی مطابق رابطه زیر می‌باشد:

$$\|X - W\| = \left(\left(\sum_{i=1}^{d}(x_i - w_i)^2\right)\right)^{\frac{1}{2}} \qquad (3\text{-}3)$$

حال بصورت همزمان ورودی $X = (x_1, x_2, \cdots, x_d)$ با تمامی عناصر موجود در شبکه مقایسه می‌گردد. نرون برنده، نرونی با فاصله کمینه در میان تمامی الگوهای مرجع از داده ورودی می‌باشد.

- تعیین نرون‌های همسایه، بعد از مشخص‌شدن نرون برنده مجموعه‌ای از نرون‌های همسایه نرون برنده که می‌بایست مقادیرشان تغییر نمایند، مشخص می‌گردند. تغییر مقادیر مربوط به نرون‌های همسایه در حالت کلی به دو صورت انجام می‌پذیرد. در حالت اول یک شعاع همسایگی معین در اطراف سلول برنده انتخاب می‌گردد. در این روش تمامی نرون‌هایی از شبکه که در فاصله معین از نرون برنده می‌باشند با یک ضریب ثابت به سمت ورودی حرکت می‌نمایند. در روش دوم تمامی نرون‌های موجود در شبکه با ضریبی نابرابر به سمت ورودی حرکت می‌نمایند. این ضریب نابرابر بصورتی است که در نرون برنده حداکثر مقدار را داشته و با دورشدن از نرون برنده مقدارش کاهش پیدا می‌کند.

- اصلاح اوزان، در انتها اوزان مربوط به نرون برنده و همسایگانش می‌بایست بر اساس ورودی شبکه اصلاح گردند. این تغییرات بر اساس رابطه زیر صورت می‌گیرد.

$$w_j(t+1) = w_j(t) + \eta(t)(i_l - w_j(t)) \quad (4\text{-}3)$$

نرخ یادگیری، $\eta(t)$، معمولا با گذشت زمان به طور خطی با تعداد ارائه بردارهای ویژگی کاهش می‌یابد:

$$\eta(t) = e_d \eta_0 \left(1 - \frac{t}{T}\right) \qquad 1 \le t \le T \quad (5\text{-}3)$$

$$0 < \eta(t) \le \eta(t-1) \le 1 \quad (6\text{-}3)$$

که η_0 نرخ یادگیری اولیه است، T تعداد کل ارائه همه بردارهای ویژگی، t نشان‌دهنده t امین ارائه همه بردارهای ویژگی و e_d تحریک جانبی وابسته به موقعیت بردار وزن انتخاب شده و بردار وزنی که باید بروزرسانی شود، است [21].

بعد از اینکه آموزش فرایند خودسازمانده کامل شد، وزن‌های شبکه آموزش داده شده، مراکز خوشه‌های داده‌های ورودی، به عنوان‌بردارهای اولیه[1]در یک کتاب‌رمز برای کوانتیزه سازی بردارهای ویژگی به کار می‌رود.

کوانتیزه کننده بردار[2] که با یک فرایند خودسازمانده کوهنن طراحی می‌شود می‌تواند اطلاعات همسایگی بردارهای اولیه در کتاب‌رمز را به همراه داشته باشد. این اطلاعات می‌تواند در بهبود کارایی سیستم بازشناسی موثر باشد[21].

[1] Prototype vectors
[2]Vector Quantizer

3-4 کلاسه‌بند

آخرین مرحله در بازشناسی، کلاسه‌بندی است. روش‌های مختلفی برای کلاسه‌بندی وجود دارد که همه آنها مبتنی بر کاربرد، ماهیت داده ها و همچنین تعداد داده ها می‌باشد. برای مثال زمانی‌که تعداد داده های ورودی مناسب باشد می‌توان از شبکه عصبی برای کلاسه‌بندی استفاده کرد در حالی که در مواردی که تعداد داده های ورودی مناسب نباشد ماشین بردار پشتیبان پیشنهاد می‌شود[22]. در این پایان نامه تمرکز اصلی روی شکل کلی کلمه است از طرفی نوشتار فارسی دارای ماهیت پیوسته است و همچنین موقعیت مکانی یک حرف در کلمه، نوشتار حرف را مشخص می‌کند. با نگاهی به این ویژگی‌ها می توان استنباط کرد که کلاسه‌بند HMM سازگاری مناسبی با این داده‌ها دارد لذا این کلاسه‌بند پیشنهاد می‌شود.

3-4-1 مدل مخفی مارکوف[11]

یک سیستم مارکوف، یک زنجیره است که حالت‌های مختلفی با شرایط تصادفی دارد. در این زنجیره حالت‌هایی در زمان t داریم که از حالت‌های در زمان $t-1$ تاثیر می‌پذیرد. HMM ها بهترین راه حل برای این نوع از مسائل در بازشناسی گفتار، بازشناسی دست‌نوشته، بازشناسی حرکات موقع سخن‌گفتن، بیوانفورماتیک[1] و تحلیل مالی هستند. در یک مدل مارکوف همه حالت‌ها آشکار است بنابراین تنها پارامترها، احتمالات انتقال حالت هستند.

در HMM ها حالت‌ها برای ناظر مخفی است اما خروجی‌ها مشخص است. این نوع از سیستم‌های مارکوف ، مدل مخفی مارکوف نامیده می‌شود زیرا دنباله‌ای که منجر به یک خروجی خاص می‌شود مخفی است حتی اگر پارامترها همه به دقت تعیین شده باشد.

[1]Bioinformatics

3-4-1-1پارامترهای HMM

یک HMM سه پارامتر دارد که به صورت $\lambda = (\pi, A, B)$ نشان داده می‌شود.که λ یک HMM است که با π،A وB تعریف شده است. π توزیع اولیه حالت‌ها است. A ماتریس انتقال حالت وB ماتریس احتمال مشاهدات است. M وN به ترتیب به عنوان تعداد حالت‌ها در یک مدل و تعداد نمونه‌های مشاهده مجزا در هر حالت در نظر گرفته می‌شود، که درHMM گسسته M تعداد خوشه‌ها است. ماتریس انتقال حالت به صورت زیر مشخص می‌شود:

$$A = \{a_{ij}\} \tag{7-3}$$

$$a_{ij} = P\left[q(t+1) = S_j | q_t = S_i\right]; \quad 1 \le i, j \le N \tag{8-3}$$

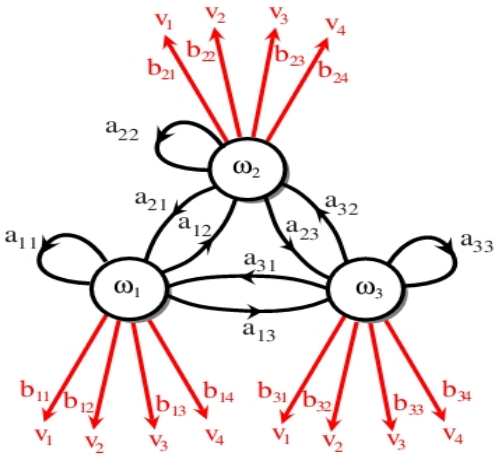
S به صورت $S = \{S_1, S_2, S_3, \dots, S_n\}$ مشخص می شود که حالت ها را نشان می‌دهد و حالت در زمان t، q_t است.

اگر ساختار HMM به گونه‌ایی در نظر گرفته شود که بتوان از هر حالت به همه حالت‌ها رسید در آن‌صورت احتمال انتقال بین تمامی حالت‌ها $a_{ij} > 0$ خواهد بود. احتمال مشاهده به ازای هر یک از حالت‌های مخفی توسط عناصرماتریس B نشان داده شده است که به صورت زیر تعریف می‌شود.

$$B = \{b_j(k)\} \tag{9-3}$$

$$b_j(k) = P\left[V_k \; at \; t | q_t = S_j\right]; \; 1 \le j \le N \; and \; 1 \le k \le N \tag{10-3}$$

عنصر $b_j(k)$ احتمال مشاهدات برای مشاهده k ام در حالت مخفی j را نشان می‌دهد. شکل3-11 یک مدل HMM با سه حالت و 4 بردار مشاهده را نشان می‌دهد. w_3, w_2, w_1 حالت‌های مخفی و v_4, v_3, v_2, v_1 بردارهای مشاهدات می‌باشند.



شکل3-11: یک مدل HMM با سه حالت و چهار بردار مشاهده[22]

یک پارامتر دیگر که باید برای کامل کردن مدل تعریف شود، بردار احتمال حالت‌های اولیه است که به صورت زیر تعریف می‌شود:

$$\pi = \{\pi_i\} \tag{11-3}$$

$$\pi_i = P[q_1 = S_i]; \qquad 1 \le i \le N \tag{12-3}$$

π یک بردار است که احتمال بودن در هر حالت در زمان t=1 را نشان می‌دهد.

3-4-1-2 HMM گسسته یا پیوسته

در بازشناسی گفتار HMM پیوسته استفاده می‌شود در حالیکه در بازشناسی دست‌نوشته معمولا یک چالش برای انتخاب بین HMM گسسته و پیوسته وجود دارد. در سال 1996 مطالعه‌ای برای نشان‌دادن مقایسه بین سیستم‌های گسسته و پیوسته انجام شد. این تحقیق نشان داد که HMM گسسته منجر به نتایج بهتری در بازشناسی دست‌نوشته می‌شود[23].

دریک مدل مخفی مارکوف گسسته[1] (DHMM) خروجی فرآیند مورد بررسی به عنوان دنباله‌ایی از مشاهدات، که متعلق به یک مجموعه محدود است، در نظر گرفته می‌شود. هر مشاهده یک کد در کتاب‌رمز می‌تواند باشد. همانطور که در توضیحات بخش های قبلی اشاره شد کتاب‌رمز با یک روش کوانتیزه‌سازی بردار ویژگی مشاهدات را به یک کد ترجمه می‌کند. درDHMM بردار مشاهده به یک کد در مجموعه محدود کوانتیزه می‌شود که البته باعث از دست رفتن بخشی از اطلاعات می‌شود. در یک مدل مخفی مارکوف پیوسته[2](CHMM)، به دلیل کوانتیزه نکردن بردار مشاهدات، اطلاعاتی از دست نمی‌رود اما در مقابل CHMM نیاز به پارامترهای بیشتری دارد ولذا نیاز به حافظه بیشتری دارد. با توجه به اینکه مدل مخفی مارکوف گسسته نیاز به منابع کمتری جهت طراحی و آموزش دارد ما DHMM را برای طراحی و پیاده‌سازی سیستم بازشناسی کلمه دست‌نوشته خود انتخاب می‌کنیم.

3-1-4-3 سه مسئله اساسی در HMM

برای اینکه HMM در کاربردهای واقعی مفید باشد سه مسئله اساسی باید حل شود.

ارزیابی[3]:فرض می‌کنیم ما به ازای هر فرآیند (در مسئله مورد بررسی هر کلمه) یک HMM را آموزش داده ایم، هرHMM با یک مجموعه سه گانه $l=(p,A,B)$ و ویژگی‌های استخراج شده (از کلمه) به عنوان یک دنباله از مشاهدات می‌باشد. مسئله، یافتن بهترین مدل HMM است که بتواند مشاهدات (ویژگی کلمه) را تولید کند. الگوریتم پیشرو[4] می‌تواند با محاسبه احتمال اینکه دنباله مشاهدات توسط هر یک از HMM ها تولید گردد ، مسئله را حل کند. معمولا در فرآیندهای بازشناسی خط یا بازشناسی گفتار با چنین مسئله‌ایی روبه رو هستیم. برای هر کلمه در فرهنگ لغت یک مدل HMM آموزش می‌دهیم، پس از مشاهده ویژگی‌های یک کلمه بهترین مدل HMM تعیین کننده هویت کلمه مورد پرسش است.

---

[1]Discrete Hidden Markov Model

[2]Continuous Hidden Markov Model

[3]Evaluation

[4]Forwards algorithm

رمزگشایی[1]: مسئله، پیدا کردن حالت‌های مخفی است که منجر به دنباله مشاهدات می‌شود. الگوریتم ویتربی[2] معمولا برای حل این مسئله استفاده می‌شود. هیچ دنباله صحیحی برای رمز گشایی وجود ندارد بنابراین از معیارهای بهینه برای حل این مسئله استفاده می‌شود. این مسئله به طور گسترده در پردازش زبان طبیعی[3]نیاز به حل شدن دارد که در آن لازم است کلمات با کلاس نحوی به عنوان اسم‌ها و فعل‌ها و ... برچسب زده شود. بنابراین کلمات در جمله به عنوان مشاهده و کلاس نحوی به عنوان حالت‌های مخفی در نظر گرفته می‌شود و هدف یافتن بهترین کلاس نحوی برای یک کلمه داده شده در متن است.

یادگیری[4]: در این مسئله، ما دنباله مشاهده را داریم و حالت‌های مخفی که منجر به مشاهدات می‌شود را می‌دانیم و تلاش می‌کنیم بهترین (محتمل ترین) HMM که دنباله مشاهده شده را توصیف می‌کند، پیدا کنیم. الگوریتم پیشرو-پسرو[5] معمولا برای حل این مسئله استفاده می‌شود.

در بازشناسی خط ما از حل مسئله3 (یادگیری) برای مدل کردن خط، از حل مسئله 2(رمز گشایی) برای بهبود مدل و از حل مسئله1(ارزیابی) برای یافتن بهترین تطبیق برای تصویر آزمون داده شده استفاده می‌کنیم.

3-4-1-4 آموزشHMM برای سیستم بازشناسی کلمه دست‌نوشته

روش‌های مختلفی برای آموزش HMM وجود دارد. در این مرحله از معیار بیشترین شباهت[6](ML) برای آموزش مدل‌ها استفاده می‌شود. در این معیار در طول مرحله آموزش ابتدا پارامترهایHMM مقداردهی اولیه می‌شوند و سپس در یک فرایند تکراری دوباره تخمین زده می‌شوند به گونه‌ای که

[1]Decoding
[2]Viterbi algorithm
[3]Natural Language Processing
[4]Learning
[5]forward-backward algorithm
[6]Maximum likelihood

شباهت مدل تولید شده به دنباله‌های آموزش افزایش یابد. فرایند آموزش زمانی متوقف می‌شود که شباهت به یک مقدار بیشینه برسد. حداکثر اطلاعات متقابل[1] و حداقل اطلاعات تبعیض[2]، معیارهای دیگر آموزش HMM هستند.

از الگوریتم بام-ولش[3] برای آموزش مدل‌های کلاس کلمه استفاده می‌شود. اولین گام محاسبه $P(O|\lambda)$، احتمال دنباله مشاهده O با توجه به مدل λ است.

از الگوریتم پیشرو[4] برای محاسبه $P(O|\lambda)$ استفاده می‌شود. متغیر پیشرو، $\alpha_t(i)$ به صورت زیر تعریف می‌شود:

$$\alpha_t(i) = P(O_1, O_2, O_3, \dots, O_t, q_t = S_i|\lambda) \quad (13\text{-}3)$$

این متغییر احتمال دنباله مشاهدات جزئی $O_1, O_2, O_3, \dots, O_t$ و حالت S_i در زمان t با توجه به مدل λ مشخص می‌کند . الگوریتم استقرای $\alpha_t(i)$ به صورت زیر است:

مقداردهی اولیه:

$$\alpha_t(i) = \pi_i b_i(O_i), \quad 1 \le i \le N \quad (14\text{-}3)$$

استقرا:

[1]Maximum mutual information
[2]Minimum discrimination information
[3]Baum–Welch
[4]Forward algorithm

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_i(O_{t+1}), \quad 1 \le t \le T-1, 1 \le j \le N \tag{15-3}$$

خاتمه:

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{16-3}$$

هدف الگوریتم بام- ولش، تنظیم پارامترهای مدل $\lambda = (\pi, A, B)$ برای بیشینه ساختن $P(O|\lambda)$ است. این سخت‌ترین مسئله در حوزه HMM است. بام-ولش یک الگوریتم تکرارشونده بر مبنای الگوریتم‌های پیشرو و پسرو است. متغیر $\beta_t(i)$ به صورت زیر تعریف می‌شود:

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, O_{t+3} \cdots, O_T, q_t = S_i, \lambda) \tag{17-3}$$

که احتمال دنباله مشاهده از $t+1$ تا آخر در حالت S_i در زمان t با توجه به مدل λ است. الگوریتم استقرای $\beta_t(i)$ به صورت زیر است:

مقداردهی اولیه:

$$\beta_T(i) = 1, \quad 1 \le i \le N \tag{18-3}$$

استقرا:

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \cdots, 1, \qquad 1 \le i \le N \tag{19-3}$$

اکنون می‌توان احتمال بودن در حالت S_i در زمان t و حالت S_j در زمان $t+1$ با توجه به مدل و دنباله مشاهده به صورت زیر محاسبه کرد:

$$\xi(i,j) = P\left(q_t = S_i, q_{t+1} = S_j \middle| O, \lambda\right) \quad (20\text{-}3)$$

$$\xi(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{P(O|\lambda)} \quad (21\text{-}3)$$

$\gamma_t(i)$ احتمال بودن در حالت S_i در زمان t با توجه به دنباله مشاهده O و مدل λ به صورت زیر محاسبه می‌شود:

$$\gamma(i) = P(q_t = S_i|O, \lambda) \quad (22\text{-}3)$$

یا با استفاده از متغییرهای پیشرو و پسرو به صورت زیر محاسبه می‌شود:

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{P(O|\lambda)} \quad (23\text{-}3)$$

تعداد انتقال مورد انتظار از S_i به صورت زیر تعریف می‌شود:

$$\sum_{t=1}^{T-1} \gamma_t(i) \quad (24\text{-}3)$$

و تعداد انتقال مورد انتظار از S_i به S_j به صورت زیر تعریف می‌شود:

$$\sum_{t=1}^{T-1} \xi(i,j) \quad (25\text{-}3)$$

برای برآورد مجدد پارامترهای π ، A،B یک HMM می‌توانیم از فرمول‌های زیر استفاده کنیم:

1-تکرار مورد انتظار در حالت S_i در زمان t :

$$\bar{\pi}_t = \gamma_l(i) \quad (26\text{-}3)$$

2-ضریب انتقال: تعداد انتقال مورد انتظار از حالت S_i به حالت S_j تقسیم بر تعداد انتقال مورد انتظار از حالت S_i.

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (27\text{-}3)$$

3-احتمال مشاهده نمونه: تعداد دفعات مورد انتظار در حالت j و مشاهده نمونه v_k، تقسیم بر تعداد کل دفعات مورد انتظار در حالت j.

$$\bar{b}_j(k) = \frac{\sum_{t=1,S.t.O_t=v_k}^{T-1} \gamma_t(i)}{\sum_{t=1}^{T} \gamma_t(j)} \quad (28\text{-}3)$$

3-4-1-5 آزمایش سیستم HMMبرای بازشناسی کلمه دست‌نوشته

در مرحله آزمایش، مسئله ما یافتن بهترین مدل است که بتواند داده را تولید کند. الگوریتم ویتربی می‌تواند یک مدل را به دنباله مشاهده شده از نمونه‌ها تطبیق دهد.

متغیرهای زیر را درنظر بگیرید:

$\delta_t(i)$:

میزان احتمال دنباله مشاهده $O_1, O_2, \cdots, O_t$ که توسط محتمل‌ترین دنباله حالات مدل ایجاد شده است، به طوریکه در زمان t در حالت i باشد.

$: \psi_t(i)$

متغیری است که برای ردیابی مسیر ML استفاده می‌شود که حالت‌های که شباهت را از زمان 1 تا t بیشینه می‌کنند، نگه می‌دارد.

الگوریتم ویتربی به صورت زیر توصیف می‌شود:

مقداردهی اولیه:

$$\delta_1(i) = \pi_i b_i(O_l), 1 \leq i \leq N \quad (29\text{-}3)$$

$$\psi_1(i) = 0 \quad (30\text{-}3)$$

استقرا:

$$\delta_t(i) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(O_t), \quad 2 \leq t \leq T, 1 \leq j \leq N \quad (31\text{-}3)$$

$$\gamma_t(i) = \operatorname{argmax}_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, 1 \leq j \leq N \quad (32\text{-}3)$$

خاتمه:

$$P^* = \max_{1 < i < N} [\delta_T(i)] \quad (33\text{-}3)$$

$$q_T^* = \operatorname{argmax}_{1 < i < N} [\delta_T(i)] \quad (34\text{-}3)$$

پیمایش معکوس[1] مسیر دنباله حالت:

$$q_T^* = \psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \cdots, 1 \quad (35\text{-}3)$$

[1]Backtracking

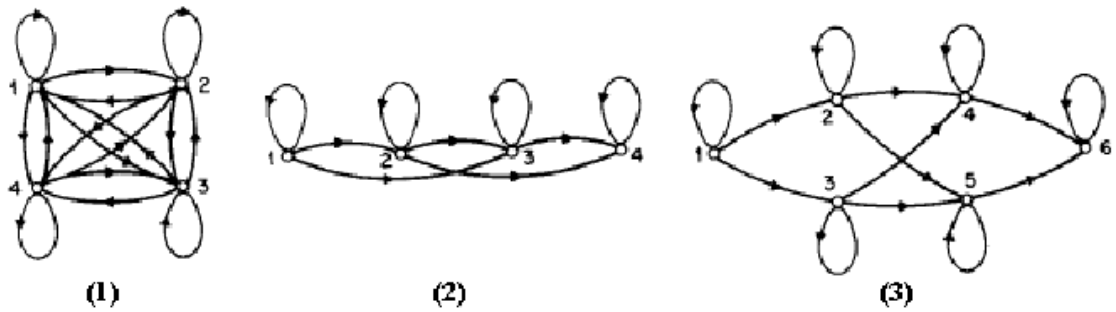
اینجا P^*، احتمال تولید دنباله توسط هریک از مدل‌ها است. مدلی که دارای بیشترین احتمال تولید این دنباله مشاهدات باشد، به عنوان کلاس کلمه شناخته می‌شود.

3-4-1-6 انواع مختلف HMM [24]

ساختارهای مختلفی می‌توان برای توپولوژی شبکه اتصال حالت‌ها در یک HMM در نظر گرفت که تفاوتشان تنها در مقادیر ماتریس A می باشد. شکل3-12 سه مدلHMM را نشان می‌دهد. برای مثال در یک شبکه ی انتقال برای چهار حالت ماتریس A به شکل زیر خواهد بود :

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \quad (36\text{-}3)$$

حال اگر بخواهیم ماتریس فوق مربوط به یک شبکه‌ی ارگودیک[1] باشد، که در آن قابلیت انتقال از هر حالت به همه‌ی حالت‌ها وجود دارد، شرط غیر صفر بودن را باید به آن اضافه کرد. برای بعضی کاربردهای خاص گاهی لازم می شود محدودیت‌های دیگری روی مقدار درایه‌های این ماتریس در نظر گرفت مانند شکل زیر :



شکل 3-12 : ساختارهای متفاوت شبکه ی انتقالی HMM : (1) ارگودیک، (2) چپ به راست محدود، (3) دلخواه.

[24]

[1]Ergodic

یکی از این مدل‌ها شکل 3-12(2) است. چنین شبکه‌هایی را چپ به راست می‌نامند زیرا همه‌ی انتقال‌ها از حالت‌های سمت چپ به حالت‌های سمت راست و یا روی خود آن‌ها (به صورت طوقه) صورت گرفته و هیچ انتقالی به حالت‌های قبل (سمت چپ) وجود ندارد. چنین شبکه‌هایی کاربرد زیادی در مدل کردن سیگنال‌هایی که از توالی وقایع زمانی تشکیل می‌شوند به ویژه سیگنال‌های گفتار دارد. برای این شبکه‌ها داریم :

$$a_{ij} = 0, \quad j < i \tag{37-3}$$

که بیان می‌کند هیچ انتقالی از یک حالت به حالت‌های قبل (با اندیس کمتر) از آن وجود ندارد. علاوه بر این، برای توزیع اولیه حالت‌ها خواهیم داشت :

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{38-3}$$

زیرا دنباله حالت باید از حالت1 شروع شود. اغلب برای مدل‌های چپ به راست، محدودیت‌های دیگری روی ضرایب انتقال حالت قرار می گیرد، مانند:

$$a_{ij} = 0, \quad j > i + \Delta \tag{39-3}$$

به طور خاص برای شکل 3-12(2)، مقدارΔ برابر 2 است یعنی هیچ پرشی بیشتر از 2 مجاز نیست. بنابراین شکل ماتریس A برای این مثال به صورت زیر است:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & 0 \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix} \tag{40-3}$$

پس برای آخرین حالت در مدل چپ به راست، ضرایب انتقال حالت به صورت زیر تعریف می‌شود:

$$a_{NN} = 1 \quad (41\text{-}3)$$

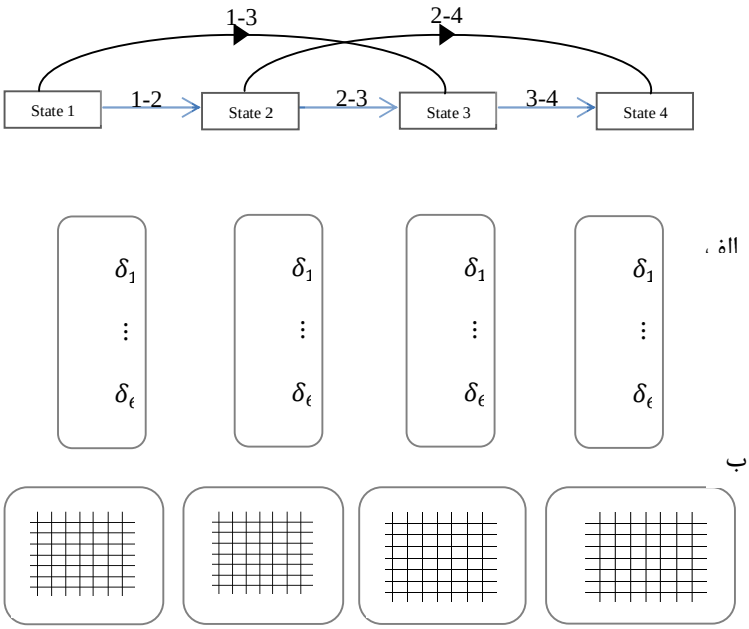
$$a_{Ni} = 0, \quad i < N \quad (42\text{-}3)$$

البته بدیهی است که این محدودیت‌ها تاثیری در تئوری تخمین پارامترهای مدل ندارد و تنها اعمال این محدودیت‌ها سبب می‌شود تعداد پارامترهای مدل که باید تخمین زده شوند کاهش یابد. زیرا اگر مقدار پارامتری را در مدل اولیه برابر صفر قرار دهیم تا آخر فرآیند صفر باقی می‌ماند. همانطور که ملاحضه شد با تغییر محدودیت‌های اعمال شده بر روی درایه‌های ماتریس A می‌توان انواع متعددی از مدل‌ها را ساخت.

3-4-2 هموارسازی[1] پارامترهای HMM با استفاده از شبکه خودسازمانده[22]

همانطور که قبلا بیان شد، برای کوانتیزه‌سازی بردارهای ویژگی از SOM استفاده شد، که خروجی آن می‌تواند به صورت ماتریس دوبعدی نمایش داده شود در این بازنمایی بجای شماره نرون خروجی، مختصات مکان در نظر گرفته می‌شود. هر حالت مدل HMM برای هر کدام از این مشاهدات یک احتمال انتشار محاسبه می‌کند، می‌توان هر بردار احتمال انتشار را به صورت یک ماتریس نمایش داد (شکل3-13).

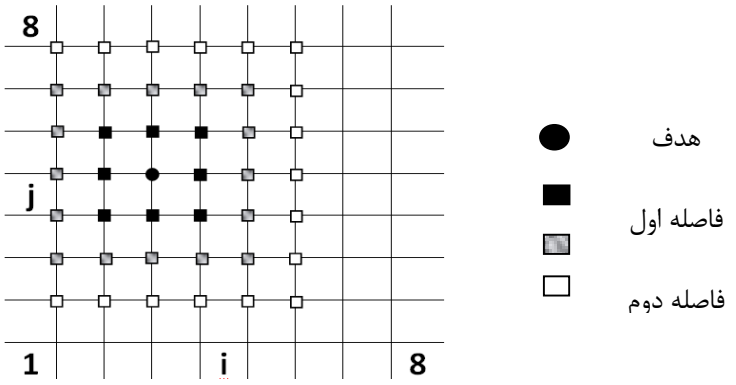
[1]Smoothing



شکل3-13:(الف) بردار احتمال انتشار، (ب) نمایش بردار احتمال انتشار به صورت ماتریسی.

اگر داده‌های آموزشی برای یک مدلHMM کلمه کافی نباشد مقدارهای صفر در بردار احتمال انتشار نمونه (بردار احتمال مشاهده برای هر حالت) رخ می‌دهد.

زمانی که آموزش یک HMM کامل می‌شود مقدار هر احتمال انتشار نمونه می‌تواند با اضافه کردن یک مجموع وزن داده شده از احتمال‌های نمونه‌های همسایه، که با الگوریتم خودسازمانده تعریف شده، افزایش یابد. در شکل 3-14 همسایگی‌های اول، دوم و سوم برای یک نقطه هدف نشان داده شده‌است.



شکل3-14: همسایگی‌ها برای یک هدف

$$P_{ij}^{new} = P_{ij}^{old} + \sum_{kl \in N_{ij}} wt_{ij,kl} P_{kl}^{old} \quad (43\text{-}3)$$

در اینجا p_{ij} احتمال مشاهده نمونه مشخص شده با موقعیت نرون (i,j) روی شبکه SOM ، N_{ij} همسایگی نرون (i,j) و $wt_{ij,kl}$ ضریب وزن است که توسط یک تابع از فاصله بین دو نرون (i,j) و (k,l) به صورت زیر تعریف می‌شود:

$$wt_{ij,kl} = c^{(d_{ij,kl}-1)}.SF \quad (i.j) \neq (k,l), \quad 0 < c < 1 \quad (44\text{-}3)$$

که c یک مقدار ثابت است ، که مقدار آن برای کار ما 0/5 انتخاب شده است، SF فاکتور کنترل درجه هموارسازی است و $d_{ij,kl}$ فاصله نگاشت بین دو نرون است که به صورت زیر تعریف می‌شود:

$$d_{ij,kl} = \max(|i-k|, |j-l|) \quad (45\text{-}3)$$

که (i,j) و (k,l) نشان‌دهنده موقعیت دو نرون در شبکه SOM است. بعد از اینکه تمام متغیرهای جدید احتمال انتشار نمونه‌ها بدست آمد، برای اینکه در بردار احتمال انتشار شرایط احتمال، مجموع احتمال‌ها برابر 1 گردد ، برقرار شود نرمالیزه سازی مقادیر بدست آمده نیاز است.

فصل چهارم:

پایگاه داده

فصل چهارم

4-1 مقدمه

در هر سیستم بازشناسی الگو یکی از مهم‌ترین موضوعات، فراهم کردن پایگاه داده مناسب برای آموزش و ارزیابی سیستم است.

خط فارسی به صورت پیوسته از راست به چپ نوشته می‌شود و شباهت‌های زیادی از لحاظ نوشتاری با زبان عربی دارد. وجود چهار حرف اضافه در زبان فارسی(گ،چ،پ ،ژ) باعث می‌شود نتوان از پایگاه داده‌های استاندارد موجود در زبان عربی برای تحقیقات در زمینه بازشناسی دست‌نوشته‌های فارسی استفاده کرد. در سال‌های اخیر کارهایی در زمینه ایجاد پایگاه داده برای بازشناسی دست‌نوشته زبان فارسی انجام شده است.

در سال 2008 پایگاه داده IfN/Farsi اسم نام شهرها ارائه شد. این پایگاه داده شامل 7271 تصویر باینری است و از 1080 نام شهر و استان که توسط 600 نویسنده جمع آوری شده است[25].

IAUT/PHCN یک پایگاه داده دیگر است که در سال 2008 توسط دانشگاه آزاد اسلامی واحد تهران شمال، ارائه شده است این مجموعه داده شامل 32400 تصویر باینری از اسامی شهرهای ایران است و توسط 380 نویسنده جمع‌آوری شده است[26].

در سال 2009 پایگاه داده چند منظوره CENPARMI شامل432537 تصویر باینری و خاکستری از تاریخ‌ها، کلمات، حروف جدا، ارقام جدا، رشته های عددی، نماد های خاص و اسناد است، ارائه شد. این داده ها توسط 400 نویسنده فارسی زبان جمع آوری شده است. انتخاب کلمات فارسی بر مبنای اسناد مالی بوده است[27].در این پایگاه داده 70 کلمه وجود دارد که ازهر کلمه تقریبا 516 تصویر وجود دارد[11].

پایگاه داده ایرانشهر پایگاه دیگری از شهر های ایران شامل16000 تصویر کلمه دست نوشته از 503 نام شهر های ایران که برای هرشهر حداقل 27 و حداکثر 35 تصویر دست نوشته وجود دارد.

اکثر پایگاه داده‌های موجود در این زمینه شامل اسامی شهرهای ایران است. در پایگاه‌های موجود مثال‌های تهیه شده برای هر کلمه نسبتا کم و یا تعداد کلمات آن‌ها کم است. از آنرو که سیستم‌های بازشناسی مبتنی بر یادگیری نیاز به مثال‌های فراوانی برای یادگیری تغییرات کلمه دارند پایگاه‌های داده موجود مناسب برای آموزش این سیستم‌ها نمی‌باشند. از طرف دیگر به منظور توسعه سیستم‌های بازشناسی کلمه، نیاز به وجود پایگاهی از کلمات پرکاربرد در زبان فارسی احساس می‌شود. با ایجاد پایگاه داده فارسا برای برطرف شدن کاستی‌های پایگاه داده‌های موجود در زمینه بازشناسی دست‌نوشته‌ها تلاش شده است.

4-2 پایگاه داده فارسا[28]

4-2-1 جمع‌آوری و مرتب‌سازی کلمات پرکاربرد

در ایجاد پایگاه داده فارسا یک نسخه از روزنامه جام‌جم، جهت استخراج کلمات پرکاربرد مورد ارزیابی قرار گرفت. تعداد تکرار هر کلمه در متن این روزنامه شمرده شد و کلماتی که حداقل چهار بار تکرار شده بود انتخاب شد. علاوه بر این چهار گروه دیگر از کلمات شامل:

- 15 کلمه از اعداد
- سه فعل پر کاربرد: است، بود، شد
- حروف اضافه شامل: را، از، در، برای و با
- ضمایر فاعلی: من، تو، او، ما، شما، ایشان

انتخاب شد و در کل یک مجموعه 300 تایی از کلمات لیست برداری شد.

4-2-2 طراحی فرم و جمع‌آوری دست‌نوشته از افراد

به منظور تهیه پایگاه داده دست‌نوشته، کلمات استخراج شده در 13 فرم تایپ شد . هر فرم شامل 6 ستون و 12 سطر می‌باشد، در هر فرم 24 کلمه تایپ شده است. از نویسنده خواسته شد تا از هر کلمه دو بار بنویسد. اندازه پنجره‌های موجود در جدول به صورتی طراحی شده است که برای نوشتن طولانی‌ترین کلمه موجود، فضای کافی وجود داشته باشد. به هر نویسنده 3 عدد فرم داده شده است. در نتیجه از هر نویسنده144 کلمه دست‌نویس در پایگاه داده موجود است. در مجموع 13 فرم مختلف 50 بار تکرار شده و توسط بیش از 250 نفر نویسنده پر شده است. به منظور همسان سازی نوع قلم به همه افراد یک نوع خودکار مشکی برای نوشتن داده شد.

4-2-3 اسکن فرم‌ها

همه فرم‌های دست نوشته با دقت بدون کج شدن برگه اسکن شد. و برای کاهش میزان خرابی داده‌ها در زمان اسکن، تمامی فرم‌ها با دقت 300dpi به صورت خاکستری توسط اسکنر اسکن شد. شکل 4-1 نمونه اسکن شده از فرم شماره 1 را نشان می‌دهد. درپایین تصویر، شماره فرم و شماره تکرارفرم و نام ونام خانوادگی نویسنده وجود دارد. از شماره فرم و شماره تکرارفرم برای برچسب‌گذاری کلمه دست‌نوشته استفاده می‌شود.

کشور	کشور	کشور	افزایش	افزایش	افزایش
قیمت	قیمت	قیمت	پلیس	پلیس	پلیس
کاهش	کاهش	کاهش	بررسی	بررسی	بررسی
ارز	ارز	ارز	بازار	بازار	بازار
سازمان	سازمان	سازمان	بانک	بانک	بانک
مناقصه	مناقصه	مناقصه	اسلامی	اسلامی	اسلامی
اداری	اداری	اداری	مدیریت	مدیریت	مدیریت
شرکت	شرکت	شرکت	ایران	ایران	ایران
برنامه	برنامه	برنامه	حوزه	حوزه	حوزه
کالا	کالا	کالا	دولت	دولت	دولت
مردم	مردم	مردم	تومان	تومان	تومان
سال	سال	سال	تغییر	تغییر	تغییر

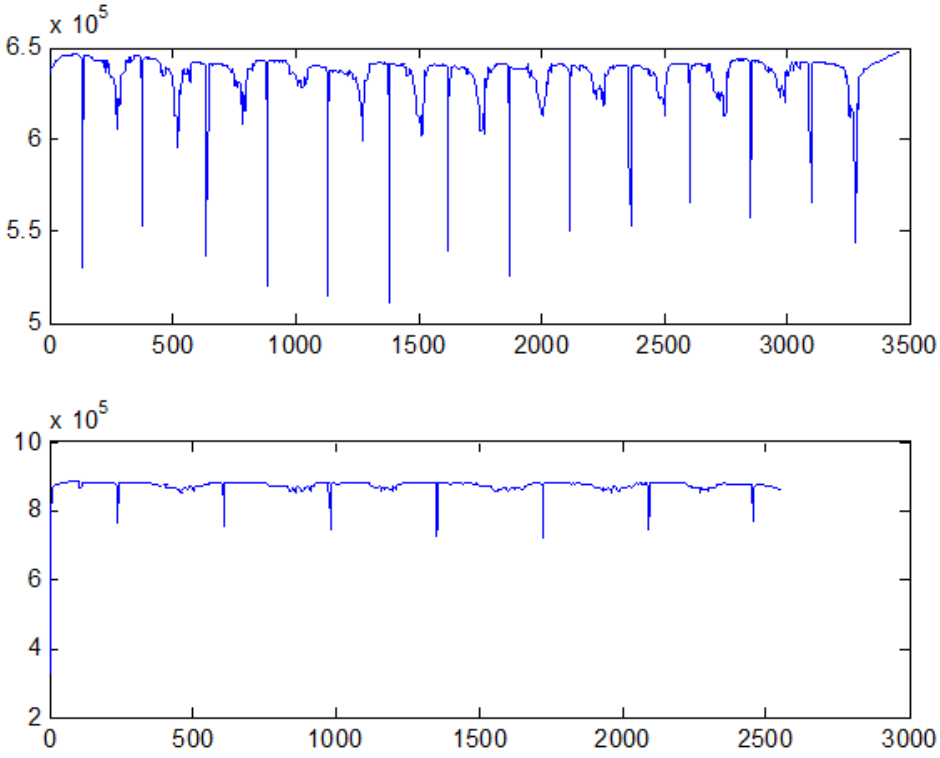
شماره فرم: 1 نام و نام خانوادگی: شماره تکرار: ۴ سن: ۲۳

شکل4-1: یک نمونه پر شده از فرم شماره1

4-2-4 جداسازی کلمات از فرم‌ها و برچسب گذاری

ابتدا تمامی تصاویر فرم‌ها برای حذف خط خوردگی‌های احتمالی مورد بازبینی قرار گرفته است. سپس با اعمال تصویر سازی افقی و عمودی و با دانستن محل تقریبی کادرها، مختصات سلول های مستطیل شکل حاوی کلمه دست‌نوشته شناسایی شده و هر سلول به صورت یک فایل تصویری مستقل با فرمت tif[1] ذخیره می‌شود.

همانطور که در شکل 4-2 نشان داده شده با تصویرسازی افقی و عمودی مکان سطرها و ستون‌ها در فرم مشخص می‌شود و کادرهای مستطیلی که کلمه‌ها داخل آن‌ها نوشته شده است، استخراج می‌شود.



الف

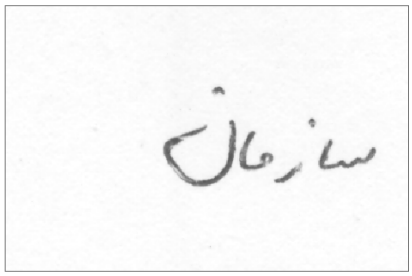
ب

شکل4-2: تصویر سازی افقی و عمودی برای آشکارسازی مکان سطرها و ستون‌ها در فرم، (الف) تصویر سازی افقی، (ب) تصویر سازی عمودی.

[1] Tagged Image File Format

هر تصویر کلمه با یک رشته نامگذاری می‌شود به صورتی که از چپ عددی بین 0 تا 299 شماره کلاس کلمه را مشخص می‌کند، است. پس از خط زیرین[1] دو رقم اول، عددی بین 01 تا 13، شماره فرم و دو رقم دوم، عددی بین 01 تا 50، شماره تکرار و رقم آخر مکان کلمه را در فرم از سمت راست نشان می‌دهد که 1 یا 2 است.

شکل 4-3 تصویر کلمه ای که از فرم شکل 4-1 استخراج شده را نشان می‌دهد این تصویر با اسم "4_01042" ذخیره شده است.

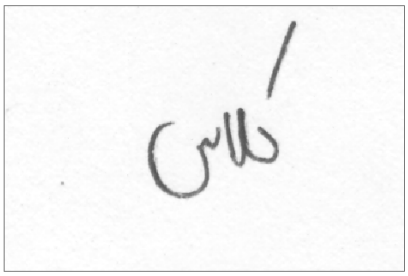


شکل4-3: تصویر استخراج شده از فرم شماره 1

4-2-5 حذف نویز و بهبود کیفیت تصویر

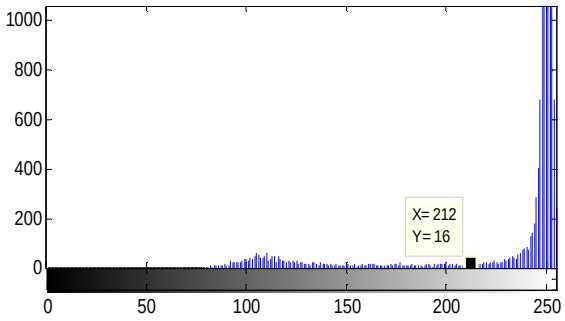
برای کاهش حجم حافظه مورد نیاز، تصویر هر کلمه را از بالا، پایین، چپ و راست به پیکسل‌های کلمه محدود می‌کنیم، پایگاه داده تهیه شده در این مرحله تصاویر سطح خاکستری کلمات دست نوشته می‌باشد. برای دوسطحی کردن[2] تصاویر، هیستوگرام روشنایی تصویر استخراج شده را رسم کرده‌ایم. آستانه باینری کردن هر تصویر را دره بین دو قله در هیستوگرام روشنایی انتخاب نمودیم. از آنجایی که در حضور نویز برخی نواحی با مساحت کم با رنگ قلم در پس زمینه حضور دارند. اجزاء مستقل تصویر که مساحت آنها کمتر از آستانه انتخابی می‌باشد را حذف نمودیم(شکل4-4).

[1]Underline
[2] Binary



(الف)

(ب)

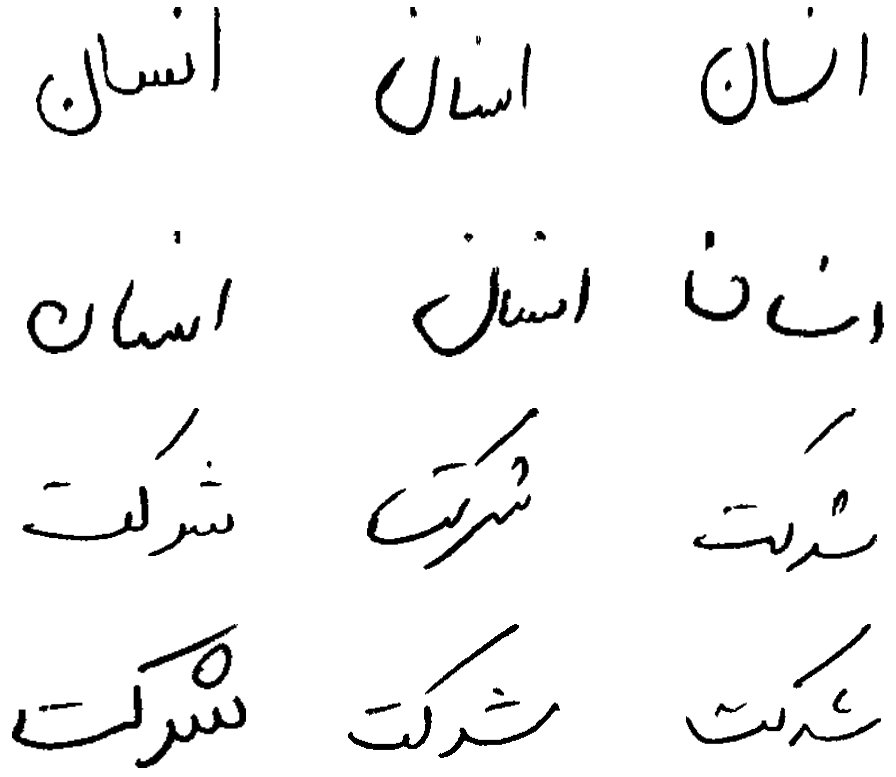


(ج)

(د)

شکل4-4: مراحل دو سطحی سازی تصویر. (الف) تصویر اصلی، (ب)تصویر محدود شده به پیکسل کلمه از چهار جهت، (ج)هیستوگرام روشنایی تصویر، (د) تصویر دوسطحی شده

در نهایت از هر کلمه دست‌نوشته، 100 نمونه موجود است. در مجموع پایگاه داده فارسا شامل 30000 تصویر از 300 کلمه دست‌نوشته است. این تصاویر از 650 فرم مخصوص، که توسط بیش از 250 نویسنده نوشته شده، استخراج شده است. شکل 4-5 چند نمونه دوسطحی از کلمه‌های "انسان" و "شرکت" را نشان می‌دهد.



شکل4-5: شش نمونه از دو کلمه در پایگاه داده فارسا

فصل پنجم:

روش پیشنهادی

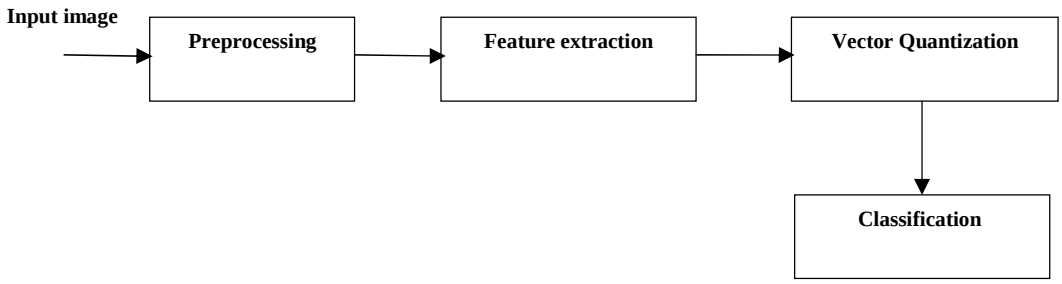
فصل پنجم

5-1مقدمه

در این پایان نامه روشی برای بازشناسی کلمات دست‌نوشته فارسی با استفاده از شکل کلی کلمات ارائه شده است. برای ارزیابی روش پیشنهادی از پایگاه داده فارسا و پایگاه داده ایرانشهر استفاده شده است. برای مقایسه با سایر روش‌های موجود در این زمینه، نتایج ارائه شده با استفاده از 198 کلاس از پایگاه داده فارسا، شامل 19800 تصویر دوسطحی از کلمات دست‌نوشته، است. 198 کلاس از پایگاه داده ایرانشهر شامل6375 تصویر دو سطحی از کلمات اسامی شهرهای ایران است. در هر پایگاه داده، 70% تصاویر برای آموزش[1] سیستم و 30% برای آزمایش[2] به طور تصادفی انتخاب شده‌اند.

الگوی کلی سیستم پیشنهاد شده برای بازشناسی کلمات دست‌نویس به صورت بلوک دیاگرام شکل 5-1 است.

در ادامه هر کدام از مراحل به طور مختصر بیان می‌شوند.



شکل 5-1: بلوک دیاگرام سیستم کلی بازشناسی کلمات دست‌نویس فارسی

[1] Train
[2] Test

5-1-1پیش پردازش

با توجه به کیفیت تصاویر مورد مطالعه و همچنین استفاده از روش کد زنجیره‌ای در مرحله استخراج ویژگی، پیش پردازش لازم برای سیستم پیشنهادی شامل یکسان‌سازی عرض قلم و تخمین خط کرسی است. عرض قلم در تصاویر با استفاده از روش اسکلت‌بندی و انبساط به عرض 4 پیکسل یکسان‌سازی و تصاویر را از ارتفاع به اندازه 45 پیکسل نرمالیزه می‌شود.

5-1-2 استخراج ویژگی

به منظور مدل کردن هر کلمه از یک HMM استفاده می‌شود، برای این منظور دنباله‌ای از بردارهای ویژگی، به کمک پنجره لغزان، از تصویر کلمه استخراج می‌شود.

5-1-3 کوانتیزه سازی بردارهای ویژگی

در این مرحله، بردارهای ویژگی که از تصاویر مجموعه آموزشی استخراج شده وارد شبکه SOM می‌شود و پس از آموزش شبکه(تعیین وزن‌های شبکه عصبی)، به ازای هر بردار ویژگی یک کد عددی بین 1 تا اندازه کتاب‌رمز استخراج می‌شود. این کد نماینده بخشی از فضا است که بردار ویژگی داده شده در آن قسمت قرار می‌گیرد. اندازه کتاب‌رمز توسط تعداد نرون‌های خروجی در شبکه عصبی تعیین می‌گردد. هرچه اندازه کتاب‌رمز بزرگتر باشد خوشه بندی بردارهای ویژگی دقیق‌تر است. از طرفی افزایش اندازه کتاب‌رمز، تعداد بردارهای وزن شبکه را افزایش می‌دهد و این باعث می‌شود زمان آموزش شبکه به نحو چشم‌گیری افزایش یابد، همچنین زمانی که تعداد خوشه‌ها خیلی زیاد شود بردارهای مشابه با تفاوتهای ناچیز در خوشه‌های مجزا قرار می‌گیرند که این عامل باعث انتشارخطا در تشخیص تصویر کلمه در مراحل بعدی می‌شود.

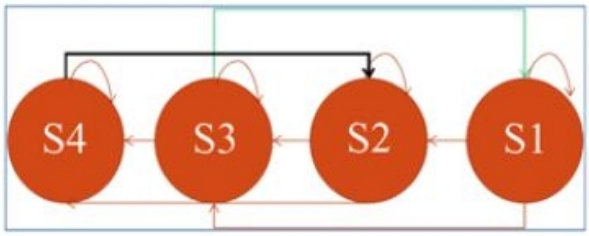
5-1-4 کلاسه‌بندی

به دلیل نگارش پیوسته زبان فارسی و تاثیرگذاری توالی حروف در شکل نوشتاری آن‌ها از کلاسه‌بند HMM در این پایان نامه استفاده می‌شود.

5-1-4-1ساختار HMM و مقداردهی اولیه پارامترها

زمانی که می‌خواهیم سیستم بازشناسی کلمه را با یک HMM طراحی و مدل کنیم، حفظ اندازه تصویر اصلی بسیار مهم است، زیرا تعداد پنجره‌های لغزان که باید تصویر کلمه را پوشش دهد خود، یک ویژگی است که می‌تواند برای متمایز نمودن کلمات از یکدیگر مورد استفاده قرار بگیرد.

متناسب با مشخصات دست‌خط فارسی، برای ساختار مدل HMM یک مدل راست به چپ طراحی شده است(شکل5-2). همانطور که در شکل 5-2 برای یک مدل 4 حالته نشان داده شده است، در هر حالت می‌تواند یک انتقال به خود، یک انتقال به حالت بعدی و یا انتقال به دو حالت بعدی وجود داشته باشد.



شکل5-2: یک مدل مخفی مارکوف با چهار حالت[11]

برای شروع کار لازم است مقادیر اولیه پارامترهای HMM شامل ماتریس‌های A ، B و بردار π معین شود. راهکارهای متفاوتی برای مقداردهی اولیه وجود دارد. یکی از ساده ترین این راهکارها تعیین مقادیر اولیه به صورت تصادفی است. یک راهکار دیگر می‌تواند تعیین مقادیر برابر و ثابت برای پارامترهای تمام HMM های مدل کلمات مختلف باشد[11].

در این پایان‌نامه از ترکیب دو راهکار بالا استفاده شده است به این صورت که، از مقادیر اولیه برابر و ثابت برای ماتریس‌های A و B استفاده می‌شود. ماتریس احتمال اولیه انتقال(A)، برای مدلی که در شکل 5-2 نشان داده شده است، به صورت زیر در نظر گرفته می‌شود:

	S_1	S_2	S_3	S_4
S_1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
S_2	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
S_3	$\frac{1}{3}$	0	$\frac{1}{3}$	$\frac{1}{3}$
S_4	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$

این ماتریس انتقال یک ماتریس با اندازه N×N است، که N تعداد حالت‌ها است. برای محاسبه مقادیر اولیه برای ماتریس احتمال مشاهدات، احتمال انتشار برابر و ثابت برای همه‌ی نمونه‌ها در نظر گرفته می‌شود. بنابراین، اگر M نشان‌دهنده تعداد نمونه‌ها باشد، احتمال خارج شدن هر نمونه در هر حالت $\frac{1}{M}$ است. برای مثال، برای یک مدل HMM با چهار حالت و سه نمونه، ماتریس احتمال مشاهدات اولیه به صورت زیر است:

	O_1	O_2	O_3
S_1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
S_2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
S_3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
S_4	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

که O_i نشان‌دهنده مشاهده یا نمونه است.

درنهایت، مقادیر تصادفی برای احتمال اولیه حالت‌ها (π) در نظر گرفته می‌شود. سپس برای اینکه مجموع آرایه‌ها در π برابر 1 شود، مقادیر نرمالیزه می‌شود. برای پیاده‌سازی HMM از جعبه ابزار Murphy استفاده می‌شود[29].

5-1-4-2 هموارسازی پارامترهای HMM

بعد از اینکه مدل‌های HMM برای کلمات موجود در پایگاه داده آموزش داده شد. در مواردی که تعداد داده‌ها برای آموزش HMM کافی نباشد، نیاز است مقادیر ماتریس احتمال مشاهدات با استفاده از روش هموارسازی که در فصل 3 شرح داده شد هموار شود.

5-2 روش‌های پیشنهادی

روش پیشنهادی اول: در این روش از هیستوگرام کدهای زنجیره‌ای برای استخراج ویژگی استفاده می‌شود.

روش پیشنهادی دوم: در این روش از ترکیب هیستوگرام کدهای زنجیره‌ای و میانگین بلوکی برای استخراج ویژگی استفاده می‌شود.

روش پیشنهادی سوم : در این بخش یک مفهوم جدید به نام معیار اطمینان[1] برای HMM معرفی می‌شود و با این نگاه جدید سعی در بهبود نتایج روش پیشنهادی دوم می‌شود.

5-3 تعاریف اولیه:

در این قسمت چند اصطلاح که در ادامه این فصل مورد استفاده قرار می‌گیرد تعریف می‌شود:

- **نرخ بازشناسی(RR)**[2]: (تعداد تصاویری که کلاس کلمه آن‌ها به درستی تشخیص داده شده تقسیم بر تعداد کل تصاویر آزمون).

$$RR = \frac{Correct\ recognized\ image}{Total\ test\ image}$$

- **Top-n**: درصد کلماتی که کلاس درست، جزء n اولویت اول توسط کلاسه‌بند HMM می‌باشد.

[1] Confidence Measur

[2] Recognition Rate

- **ماتریس سردرگمی**[1]: در این ماتریس درایه سطر i و ستونj ، تعداد تصویر آزمون است که برچسب کلاس آن‌ها i است و به عنوان تصویر کلاس j بازشناسی شده است.
- **معیارهای فاصله** [30]:

1- فاصله اقلیدسی[2]: فاصله اقلیدسی بین دو نقطه a وb با ابعاد K به صورت زیر بدست می‌آید:

$$D_E(a,b) = \sqrt{\sum_{j=1}^{k}(a_j - b_j)^2} \qquad (1\text{-}5)$$

این فاصله همیشه بزرگتر یا مساوی صفر است.

2- فاصله بلوک شهری[3] : فاصله بلوک شهری بین دو نقطه a وb با ابعاد K به صورت زیر بدست می‌آید:

$$D_C(a,b) = \sum_{j=2}^{k}|a_j - b_j| \qquad (2\text{-}5)$$

3- فاصله cosine : یکی دیگر از معیارهای شباهت است که در آن فاصله بین دو نقطه a وb با ابعاد K به صورت زیر محاسبه می‌شود:

$$D_{cos}(a,b) = \frac{\sum_{j=1}^{k} a_j \times b_j}{norm(a) \times norm(b)} \qquad (3\text{-}5)$$

[1] Confusion Matrix
[2] Euclidean Distance
[3] City Block Distance

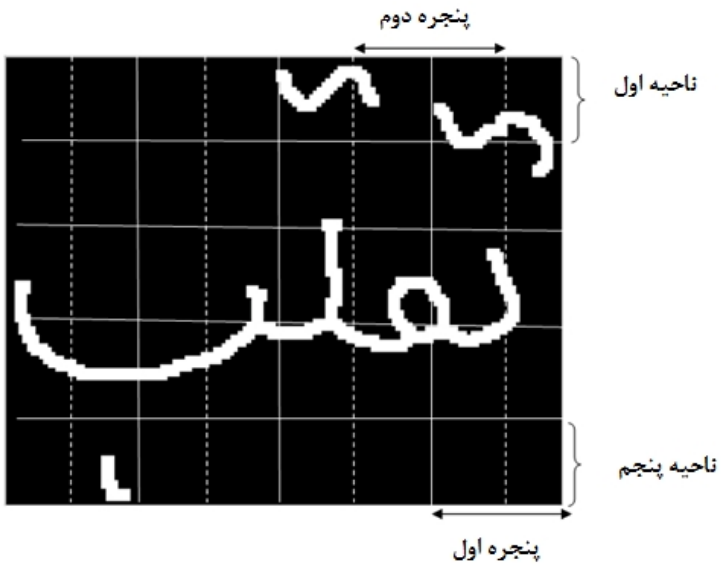
بطوریکه نرم بردارa برابر است با :

$$\text{norm}(a) = \sqrt{\sum_{j=1}^{k} a_j^2} \qquad (4\text{-}5)$$

4-5 روش پیشنهادی اول: استفاده از هیستوگرام کدهای زنجیره‌ای

در این روش، پنجره لغزان با عرض دوبرابر عرض قلم، 8 پیکسل، در نظر گرفته می‌شود و به اندازه نصف عرض پنجره یعنی 4 پیکسل همپوشانی برای پنجره‌های متوالی در نظر گرفته می‌شود و هر پنجره به پنج ناحیه[1] افقی تقسیم می‌شود. شکل 5-3 پنجره‌های متوالی را روی تصویر کلمه دست‌نوشته نشان می‌دهد. برای بررسی تاثیر عرض پنجره در کارایی سیستم، نتایج کار با عرض کمتر از 8 پیکسل و همچنین بیشتر از 8 پیکسل در قسمت نتایج آمده است. همانطور که در فصل‌های گذشته توضیح داده شده است، کدهای زنجیره‌ای اطلاعات جهتی کانتور کلمه را شامل می‌شوند. از کد زنجیره‌ای برای هر یک از پنج ناحیه افقی در یک پنجره لغزان، هیستوگرامی بدست می‌آید شامل تعداد نقاطی از کانتور که در یکی از 4 جهت با زاویه (0،45،90،135°) قرار گرفته‌اند. بنابراین در هر پنجره یک بردار ویژگی 20 تایی (5(ناحیه)× 4(جهت)) استخراج می‌شود. شکل 5-4 بردار ویژگی را نشان می‌دهد.

[1]Zone



شکل5-3: پنجره بندی تصویر قبل از استخراج ویژگی

	۰°	۴۵°	۹۰°	۱۳۵°
ناحیه اول				
ناحیه دوم				
ناحیه سوم				
ناحیه چهارم				
ناحیه پنجم				

شکل 5-4: بردار ویژگی کد زنجیره‌ای برای یک پنجره

5-4-1 نتایج ارزیابی عملکرد روش اول با استفاده از پایگاه داده فارسا

ویژگی‌های مربوط به هیستوگرام کدهای زنجیره‌ای، به سیستم مبتنی بر مدل HMM جهت آموزش و ارزیابی داده می‌شود. سیستم پیشنهادی به دو صورت مورد بررسی و آزمایش قرار می‌گیرد. حالت اول اینکه برای تمام مدل کلمات HMM ، تعداد حالت‌ها یکسان می‌باشند و در حالت دوم با توجه به طول تصویر کلمه در مدل‌های HMM تعداد حالت‌ها متفاوت می‌باشند.

از آنجایی که مقدار دهی اولیه ماتریس وزن‌ها در شبکه SOM به صورت تصادفی است، هر کدام از آزمایش‌ها در این پایان نامه 3 بار انجام شده و بهترین نتیجه گزارش شده است.

5-4-1-1 HMM ها با تعداد حالت یکسان

برای بدست آوردن عرض پنجره بهینه در سیستم پنجره لغزان، ابتدا آزمایش و ارزیابی روی 10 کلاس اول از پایگاه داده فارسا انجام می‌شود. در این آزمایش اندازه کتاب‌رمز، 49 و تعداد حالت، 20 در نظر گرفته می‌شود. جدول 5-1 نشان می‌دهد که عملکرد سیستم به ازای بردار ویژگی استخراج شده از عرض پنجره برابر 8 بهتر از سایر حالت‌ها می‌باشد.

جدول5-1: درصد نرخ بازشناسی روش اول برای 10 کلاس از پایگاه داده فارسا برای عرض پنجره لغزان متفاوت

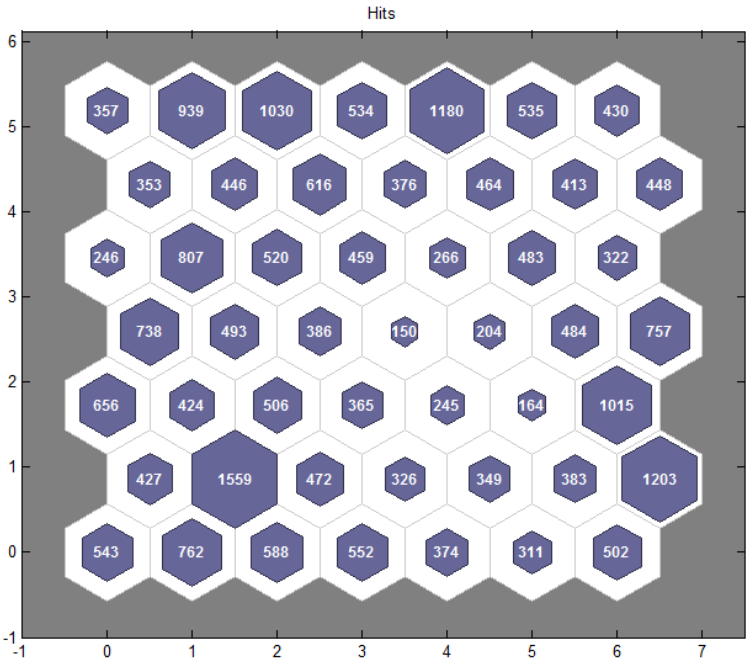
عرض پنجره	Top-1[1]	Top-5[2]
6	81/52	95/12
8	84/7	95/65
10	82/97	94/92

درتمام مراحل بعدی، عرض پنجره برای استخراج هیستوگرام کدهای زنجیره‌ای برابر با 8 در نظر گرفته می‌شود .

برای بدست آوردن تعداد حالتی که بهترین نرخ بازشناسی را می‌دهد، ابتدا آموزش و ارزیابی روی 10 کلاس اول از پایگاه داده فارسا با اندازه کتاب‌رمز 49 انجام می‌شود. شکل 5-5 و 5-6 به ترتیب هیستوگرام چینش بردارهای ویژگی داده‌های آموزش را در خروجی شبکه SOM و بردارهای وزن در شبکه SOM را نشان می‌دهد.

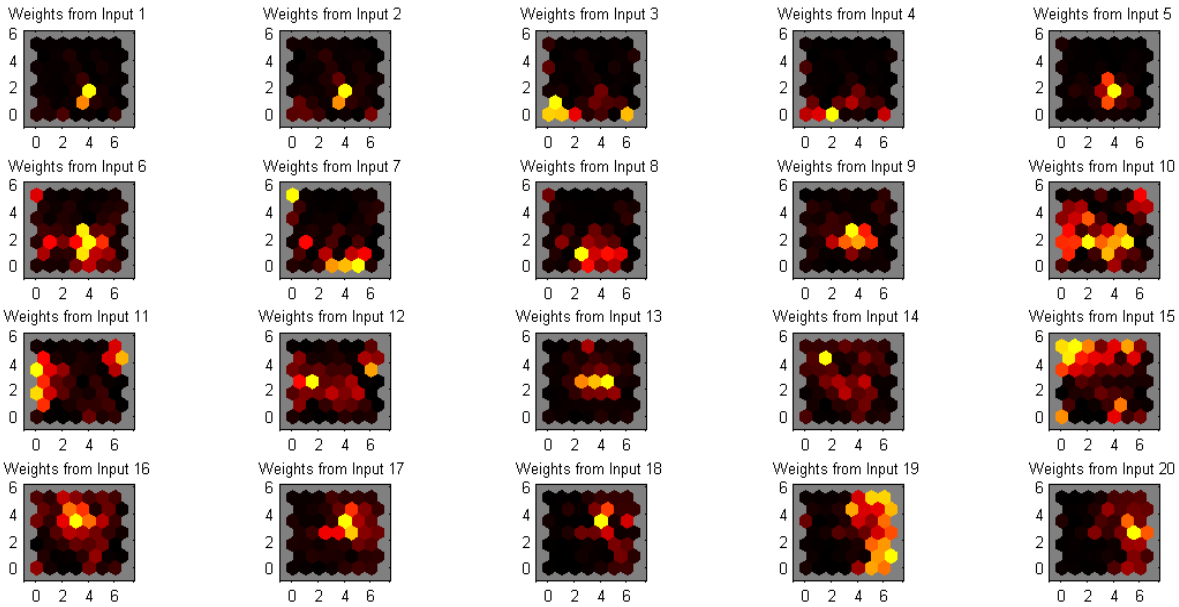
در شکل5-6 رنگ مشکی نشان دهنده مقدار صفر و رنگ‌های زرد تا قرمز مقادیر مثبت هستند و هرچه رنگ به قرمز نزدیک‌تر باشد(پر رنگ‌تر باشد) مقدار بزرگتر است.

[1] نرخ بازشناسی در انتخاب اول
[2] نرخ بازشناسی در5 انتخاب اول



شکل 5-5: هیستوگرام چینش بردارهای ورودی شبکه SOM 49تایی

در جدول 5-2 نتایج حاصل از این آزمایش، برای بردارهای ویژگی هیستوگرام کدهای زنجیره‌ای، نشان داده شده است.



شکل5-6: بردارهای وزن در شبکهSOM برای ورودی ها

جدول5-2: نتایج آزمایش روش اول بر روی 10 کلاس اول از پایگاه داده فارسا

به ازای تعداد حالت‌های مخفی متفاوت HMM

تعداد حالت‌های مخفی	Top-1	Top-5
5	81/8	96/7
10	82/6	96/37
15	84/42	96/73
20	84/7	95/65
25	84/42	95/65
30	84/42	96/06
35	85/86	95/65
40	85/5	95/65

همانطور که در جدول 5-2 مشاهده می‌شود، HMM در انتخاب اول با 35 حالت مخفی به بهترین نتیجه دست یافته است. جدول 5-3 نتایج هموار سازی ماتریس احتمال مشاهدات را روی نتایج آزمایش انجام شده در مرحله قبل در حالتی که تعداد حالت‌ها 35 در نظر گرفته شده است نشان می‌دهد. ضریب C در رابطه (3-44) برابر با 0/5 در نظر گرفته می‌شود و آزمایش روی مقادیر مختلف از ضریب SF در رابطه (3-44) انجام می‌شود. همانطور که در جدول نشان داده شده است در مقدار0/001 برای ضریبSF به بهترین نتیجه دست یافته است.

جدول5-3: نرخ بازشناسی با استفاده از هموارسازی ماتریس احتمال مشاهدات

به ازای ضریب SF متفاوت

Top-5	Top-1	SF
98/55	91/66	0/1
98/55	95/28	0/01
98/55	95/65	0/001
98/91	93/84	0/0001

ماتریس سردرگمی برای خروجی آزمایش قبل در جدول 5-4 نشان داده شده است.

جدول5-4: ماتریس سردرگمی سیستم برای 10 کلاس اول از پایگاه داده فارسا

		کشور	قیمت	کاهش	ارز	سازمان	مناقصه	اداری	شرکت	برنامه	کالا
		1	2	3	4	5	6	7	8	9	10
کشور	1	35	0	0	0	0	0	0	0	0	0
قیمت	2	1	24	1	0	0	0	0	2	0	0
کاهش	3	0	0	27	0	1	0	0	0	0	0
ارز	4	0	0	0	24	0	0	1	0	0	0
سازمان	5	0	0	0	0	21	0	0	0	0	0
مناقصه	6	0	0	0	0	0	28	0	0	0	0
اداری	7	0	0	0	0	1	0	30	0	0	0
شرکت	8	0	1	0	0	1	0	0	27	0	0
برنامه	9	0	0	0	0	0	0	1	0	20	0
کالا	10	0	0	0	0	0	0	2	0	0	29

در آزمایش بعدی تعداد حالت‌های مخفی برای مدل HMM تمام کلمات، 35 در نظر گرفته می‌شود و سیستم روی 198 کلاس کلمه از پایگاه داده فارسا ارزیابی می‌شود. جدول 5-5 نتایج آموزش و ارزیابی بر روی 198 کلاس از پایگاه داده فارسا برای کتاب‌رمز با اندازه‌های 36، 49 و64 قبل و بعد از هموار سازی پارامترهای HMM را نشان می‌دهد. همانطور که مشاهده می‌شود برای کتاب‌رمز با اندازه 49 جواب بهتری بدست می‌آید.

جدول5-5: نرخ بازشناسی با استفاده از ویژگی هیستوگرام کدهای زنجیره‌ای برای 198 کلاس

به ازای سه اندازه کتاب‌رمز مختلف ، بدون هموار سازی(خاکستری) با هموارسازی(سفید)

اندازه کتاب رمز	هموارسازی با SF=0.001	Top-1	Top-5	Top-10	Top-20
36		60/02	78/80	83/81	88/23
	ü	64/20	85/20	90/36	94/58
49		59/14	77/21	81/90	86/49
	ü	66/70	86/34	91/60	95/10
64		58/37	75/18	80/10	84/34
	ü	65/40	85/76	90/97	94/74

5-1-4-2 HMM ها با تعداد حالت‌های مخفی متفاوت

برای محاسبه تعداد حالت‌های مخفی برای مدل هایHMM ، ابتدا حداقل تعداد پنجره‌ها، در سیستم پنجره لغزان، در هر کلاس کلمه بدست می‌آید. سپس ضریبی از این مقدار به عنوان تعداد حالت مخفی برای آن کلاس در کلاسه‌بند انتخاب می‌شود.

برای تنظیم کردن مقدار این ضریب، ابتدا آزمایش بر روی 10 کلاس اول از پایگاه داده فارسا با مقادیر مختلف این ضریب و اندازه کتاب رمز 49 انجام شد. نتایج این آزمایش در جدول 5-6 نشان داده شده است.

به عنوان مثال حداقل تعداد پنجره‌ها در کلاس کلمه "کالا" با پنجره‌هایی به عرض 8 پیکسل ،13 است.

همانطور که در جدول 5-6 مشاهده می‌شود با ضریب 1/8 بهترین جواب بدست می‌آید، بنابراین کلمه "کالا" با 23=1/8×13 حالت مخفی مدل می‌شود.

جدول 5-6: نتایج بازشناسی برای HMM ها با تعداد حالت های مخفی متفاوت

مقدار ضریب	Top-1	Top-5
0/4	82/24	98/55
0/6	92/7	98/91
0/8	90/42	98/91
1	92/75	98/91
1/2	92/02	98/91
1/4	94/2	98/91
1/6	93/47	98/91
1/8	94/92	98/91
2	94/56	98/91

بنابراین اندازه ضریب در بدست آوردن تعداد حالت برای هرکدام از مدل‌های HMM برابر با 1/8 در نظر گرفته می‌شود و سیستم برای 198 کلاس از پایگاه داده فارسا آموزش و ارزیابی می‌شود. نتایج بدست آمده در این مرحله بعد از هموار سازی در جدول 5-7 مشاهده می‌شود.

همانطور که مشاهده می‌شود نرخ بازشاسی سیستم برای کتاب رمز های 36، 49 و 64 تفاوت زیادی ندارد. با توجه به این نتایج و کاهش پیچیدگی ساختار HMM و پیچیدگی محاسباتی، اندازه کتاب رمز 49 استفاده می‌شود.

جدول 5-7: ارزیابی سیستم برای HMM ها با تعداد حالت‌های مختلف بعد از هموار سازی

اندازه کتاب رمز	Top-1	Top-5	Top-10	Top-20
36	65/3	85/96	91/19	95/14
49	66/57	86/4	91/55	94/93
64	65/97	86/26	91/6	95/1

5- 5 روش پیشنهادی دوم: استفاده از بردار ویژگی 25 تایی

در این روش نیز مشابه پیشنهاد اول برای استخراج ویژگی از روش پنجره لغزان استفاده می‌شود و هر پنجره به 5 ناحیه تقسیم می‌شود. منتها در این روش از میانگین بلوکی در کنار ویژگی‌های هیستوگرام کدهای زنجیره‌ای (روش پیشنهادی اول) استفاده می‌شود.

به ازای هر پنجره لغزان پنج ناحیه افقی استخراج گردید که تعداد پیکسل‌های پیش‌زمینه(قلم) در هر ناحیه بعنوان ویژگی میانگین بلوکی آن ناحیه مورد استفاده قرار می‌گیرد. لذا برای هر پنجره لغزان 5 ویژگی میانگین بلوکی به 20 ویژگی هیستوگرام کدهای زنجیره‌ای اضافه می‌شود و در مجموع یک بردار 25 عنصری به عنوان ویژگی هر پنجره لغزان برای مدل کردن HMM مورد استفاده قرار می‌گیرد.

5-5-1نتایج ارزیابی عملکرد روش دوم با استفاده از پایگاه داده فارسا

در تمامی آزمایش‌ها اندازه کتاب رمز برابر با 49 انتخاب شده است. و برای تعداد حالت‌ها در مدل‌های HMM از ضریب 1/8 استفاده شده است.

جدول 5-8 نتایج ارزیابی روش را با استفاده از 198 کلاس از پایگاه داده فارسا نشان می‌دهد. و جدول 5-9 نتایج ارزیابی پس از هموار سازی پارامترهای HMM ، با مقادیر مختلف SF را نشان می‌دهد.

جدول 5-8: نرخ بازشناسی روش پیشنهادی دوم قبل از هموار سازی پارامترهای HMM به ازای ضرایب مختلف کنترل کننده تعداد حالت‌های مخفی

اندازه ضریب	Top-1	Top-5	Top-10	Top-20
1/6	60/39	76/89	81/8	85/98
1/8	60/73	77/35	82/1	86/27
2	60/66	76/38	81/22	85/66

جدول5-9: نرخ بازشناسی روش پیشنهادی دوم بعد از هموار سازی پارامترهای HMM به ازای مقادیر مختلف پارامتر SF

SF	Top-1	Top-5	Top-10	Top-20
0/1	61/00	82/22	87/96	92/6
0/01	68/10	86/64	91/57	95/05
0/001	68/88	87/54	91/75	95/63
0/0001	68/83	87/25	91/70	95/2

همانطور که مشاهده می‌شود روش دوم نرخ بازشناسی را 2 درصد افزایش می‌دهد.

5-6 روش پیشنهادی سوم: بازشناسی به کمک دو خبره

به دلیل بالا بودن تنوع و پیچیدگی نوشتار در زبان فارسی از یک طرف و گوناگونی دست‌نوشته‌ها از طرف دیگر نرخ بازشناسی مطلوب نیست. روش‌های مبتنی بر یک کلاسه‌بند نتوانسته‌اند نرخ بازشناسی را به حد قابل قبولی برسانند در اینجا با تحلیل خروجی‌های HMM در روش پیشنهادی دوم یک مفهوم جدید به نام معیار اطمینان معرفی می‌شود.

وقتی یک مدل HMM مناسب‌ترین مدل برای یک کلمه آزمون تشخیص داده می‌شود به این معنی است که احتمال تولید ویژگی‌های استخراج شده از کلمه توسط یک مدل HMM ، بیشتر از تولید آن توسط مدل‌های دیگر HMM در پایگاه داده است. ولی اگر این احتمال با احتمال سایر مدل‌های HMM تفاوت فاحشی نداشته باشد به معنی این است که در شناسایی کلمه اطمینان زیادی وجود ندارد. این واقعیت ما را برانگیخت تا معیار اطمینانی براساس احتمال تولید کلمه توسط یک مدل HMM تعریف کنیم و در مواردی که این معیار اطمینان پایین باشد شناسایی کلمه را به یک کلاسه‌بند دیگر واگذار کنیم.

در کلاسه‌بند HMM هر تصویر آزمون که وارد سیستم می‌شود، احتمال انتصاب تک تک مدل‌های HMM به کلمه ورودی محاسبه می‌شود. مدل HMM بزرگترین احتمال کلاس کلمه ورودی تست را مشخص می‌کند.

جهت ارزیابی احتمال مربوط به مدل‌های HMM به عنوان معیار مناسب برای اعتبارسنجی شناسایی کلمه، آزمونی را انجام دادیم. در این آزمون نتایج HMM برای تمامی تصاویر کلمه‌های تست بررسی شد و برای هر تصویر آزمون، اختلاف دو بزرگترین احتمال بدست آمد. با بررسی این اختلاف احتمال‌ها مشاهده شد که در مواردی که HMM با خطا کلمه را بازشناسی می‌کند نسبت به مواردی که کلمه بدرستی بازشناسی می‌شود، اختلاف احتمال در سطح پایین‌تری قرار دارد. شکل5-7 هیستوگرام اختلاف احتمال برای کلماتی که با مدل‌های HMM بدرستی بازشناسی شده‌اند و با خطا بازشناسی شده اند را نشان می‌دهد.

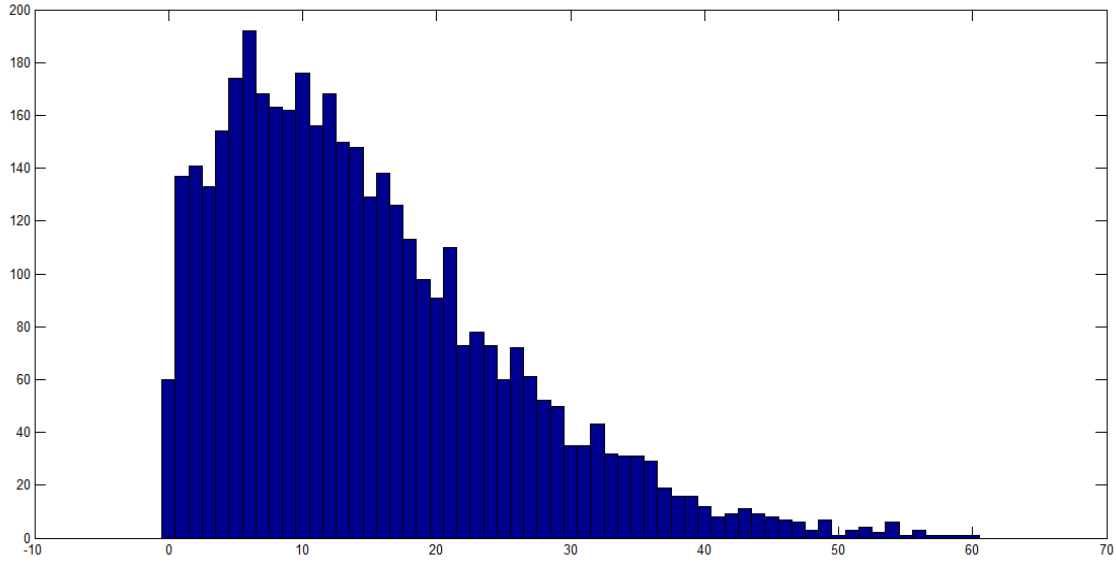
با استفاده از هیستوگرام‌های رسم شده برای HMM یک معیار اطمینان تعریف می‌نمائیم . با این معیار، HMM برای هر تصویرکلمه تعیین می‌کنیم آیا می‌تواند بازشناسی کلمه معتبر است یا اینکه تصمیم‌گیری به کلاسه‌بند دیگری واگذار شود. برای این منظور باید یک آستانه برای معیار اطمینان مشخص نمود. این آستانه تعیین می‌کند اگر اختلاف دو بزرگترین احتمال در خروجی

HMM بیشتر از حد تعیین شده باشد HMM می‌تواند بازشناسی را برای تصویر مورد نظر انجام دهد در غیر اینصورت نمی‌تواند کلاس تصویر را تعیین کند.

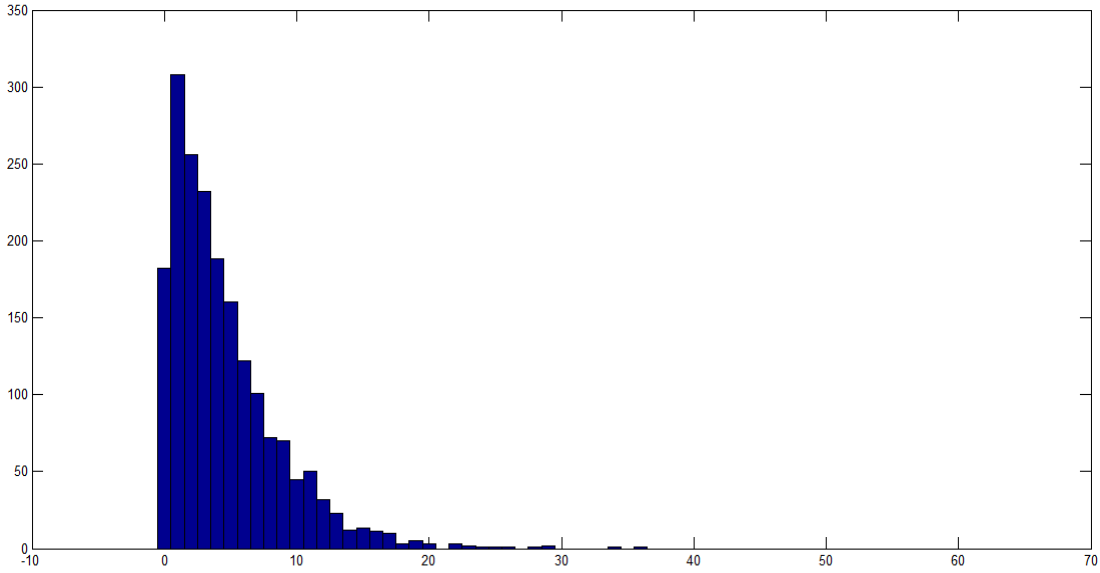
با توجه به شکل 5-7 مقدار آستانه نباید بیش از حد بزرگ یا بیش از حد کوچک باشد زیرا در حالتی که کلمات مشابهی وجود داشته باشد بعضی تغییرات در نوشتار باعث می‌شود اختلاف دو بهترین احتمال در خروجی HMM کوچک باشد و بازشناسی را دچار خطا کند. و زمانی که مقدار آستانه بزرگ انتخاب شود با توجه به شکل 5-7 (الف)، HMM برای بسیاری از تصاویری که به درستی بازشناسی می‌کند قادر به تصمیم گیری نخواهد بود.

برای مواردی که معیار اطمینان کوچکتر از مقدار آستانه باشد از روش دیگری برای بازشناسی استفاده می‌شود.

با بررسی نتایج روی تصاویر کلماتی که معیار اطمینان، برای بازشناسی HMM کوچکتر از حد آستانه است، مشخص شد کلاس درست در میان چند کلاس با حداکثر احتمال HMM می‌باشد. به عنوان مثال در 90 درصد موارد کلاس درست کلمه ورودی در میان 20 کلاسی است که بیشترین احتمالات را توسط HMM ایجاد کرده است.



الف

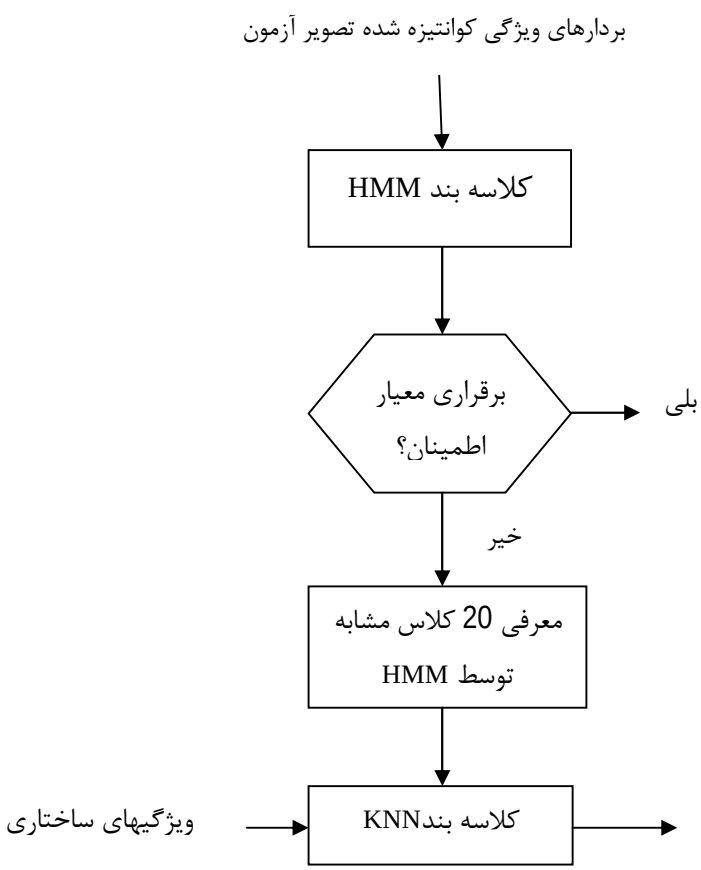


ب

شکل5-7: هیستوگرام اختلاف احتمال دو تا بهترین مدل، (الف) برای تصاویری که به درستی با HMM بازشناسی شده اند، (ب) برای تصاویری که اشتباه بازشناسی شده‌اند.

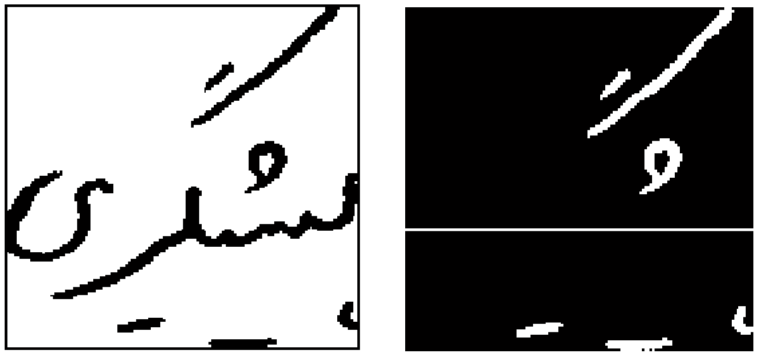
در این پایان نامه برای بازشناسی تصویر آزمونی که HMM قادر به بازشناسی آن نیست از کلاسه بند KNN، روی 20 کلاسی که HMM معرفی می‌کند، استفاده می‌شود. برای این منظور از ویژگی‌های ساختاری برای کلاسه‌بند KNN استفاده می‌شود.

شکل 5-7 بلوک دیاگرام مرحله کلاسه‌بندی در روش دوخبره‌ای پیشنهاد شده را نشان می‌دهد.



شکل 5-7: بلوک دیاگرام مرحله کلاسه بندی در روش پیشنهاد شده دو خبره‌ای

برای استخراج ویژگی‌های ساختاری از روش ارائه شد در [12] استفاده نموده‌ایم. به طور مشخص تعداد مولفه‌های متصل تصویر، تعداد مولفه‌های متصل بالای خط کرسی و تعداد مولفه‌های متصل پایین خط کرسی به عنوان ویژگی استفاده می‌شود. به این ترتیب که بعد از استخراج خط کرسی با در نظر گرفتن فاصله 10 پیکسل از بالا و پایین خط کرسی مولفه‌هایی که بالاتر از این مکان قرار می‌گیرند به عنوان مولفه‌های بالا و پایین خط کرسی در نظر گرفته می‌شوند. شکل 5-8 تصویر یک کلمه و مولفه‌های متصل بالا و پایین خط را نشان می‌دهد.

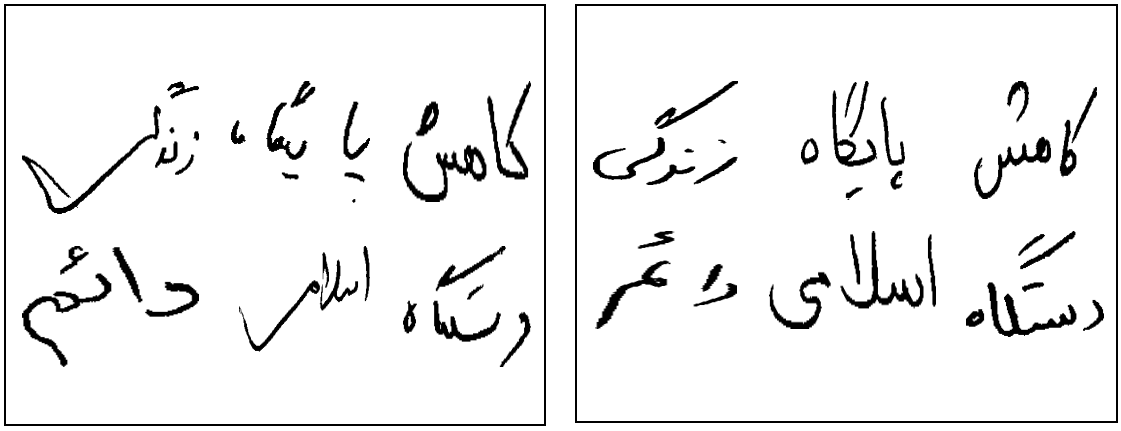


الف ب

شکل5-8: استخراج ویژگی ساختار-مانند (الف) تصویر اصلی، (ب) مولفه‌های متصل بالا و پایین خط کرسی

5-6-1 ارزیابی روش بازشناسی دوخبره‌ای با استفاده از پایگاه داده فارسا

با توجه به هیستوگرام های رسم شده در شکل 5-6 مقدار آستانه برای معیار اطمینان برای کلاسه‌بند HMM برابر 5 در نظر گرفته شد. با این مقدار آستانه از میان 5908 کلمه آزمون، اعتبار 2292 کلمه در بازشناسی توسط HMM تأیید نمی‌شود. شکل 5-9(الف و ب) به ترتیب تعدادی از کلمات که معیار اطمینان برای آن‌ها پایین‌تر از حد آستانه و بالاتر از حد آستانه می‌باشد را نشان می‌دهد.



الف ب

شکل5-9: تقسیم بندی نمونه‌هایی از تصاویر با معیار اطمینان (الف)تصاویری که شرایط اطمینان را ندارند، (ب) تصاویری که شرایط اطمینان را دارند.

برای کلماتی که اختلاف بزرگترین احتمال‌ها در خروجی در مدل HMM کوچکتر از مقدارآستانه باشد، نتیجه بازشناسیHMM نامعتبر در نظر گرفته می‌شود. با این وجود HMM لیستی از مدل کلمات مشابه با کلمه ورودی را فراهم می‌سازد. برای مثال 20 احتمال بزرگتر در خروجی HMM برای هر ورودی آزمون در نظر گرفته می‌شود.

کلاسه‌بندی که به کمکHMM می‌آید تا بازشناسی را در موارد نامطمئن تکمیل نماید می‌بایست پیچیدگی محاسباتی کمی داشته باشد. از طرفی چون لیست کاندیدها بعد از ارزیابی HMM حاصل می‌گردد و این لیست از قبل تعیین شده نیست، کلاسه‌بند مورد استفاده باید به آموزش نیاز نداشته باشد. در مباحث بازشناسی آماری الگو[1] یکی از ساده ترین کلاسه‌بندهایی که شرایط فوق را دارا می-باشد کلاسه‌بند KNN است، از این کلاسه‌بند در مواردی که HMM قادر به بازشناسی مطمئن کلمه ورودی نیست استفاده می‌شود. در این موارد از کلاسه‌بند KNN برای تعیین کلاس کلمه ورودی از میان 20 کلاس کاندید شده توسط HMM استفاده می‌شود. در جدول 5-10 نتایج عملکرد KNN را برای تعداد فرد همسایه، K از 1 تا 13 و معیارهای فاصله اقلیدسی، فاصله بلوک شهری و فاصله cosine نشان می‌دهد.

[1] Statistical pattern recognition

جدول 5-10: نتایج عملکردKNN برای تصاویری که شرایط معیار اطمینان درHMM ندارند

تعداد همسایه ها (K)	فاصله اقلیدسی	فاصله بلوک شهری	فاصله cosine
1	53/2	53/2	49
3	57	57	53/5
5	59	58/9	55/4
7	60/08	60/12	56/8
9	61/08	61/21	57/37
11	61/65	61/69	57/46
13	61/56	61/61	57/64

همانطور که مشاهده می‌شود KNN با 11 نزدیک‌ترین همسایه و معیار بلوک شهری به بهترین نتیجه می‌رسد. به دلیل شباهت نوشتار کلمات در 20 کلاس معرفی شده توسط HMM و پیچیده بودن تصاویری که شرایط معیار اطمینان درHMM را ندارند، KNN روی این تصاویر نرخ بازشناسی بسیار بالایی را ایجاد نمی‌کند.

برای تصاویری که شرایط معیار اطمینان را دارند HMM عملکرد خوبی دارد. جدول 5-11 نتایج نرخ بازشناسی HMM را نشان می‌دهد.

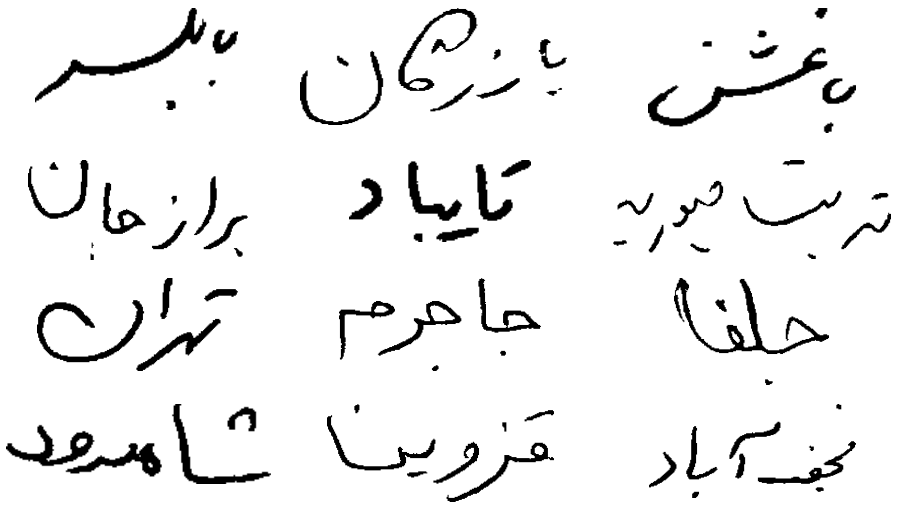
جدول5-11: نرخ بازشناسی HMM برای تصویرهایی که شرایط معیار اطمینان را دارند

	Top-1	Top-5	Top-10	Top-20
نرخ بازشناسی	85/87	94/22	96/24	97/73

بنابراین در حالت کلی با استفاده از سیستم پیشنهادی دو خبره‌ای نرخ بازشناسی به 76/49% ارتقا می‌یابد. این در مقایسه با نرخ بازشناسی حاصل شده از کلاسه‌بند HMM که 68/88% بود افزایش قابل توجهی می‌باشد.

5-7 ارزیابی روش‌های پیشنهادی با استفاده از پایگاه داده ایرانشهر

پایگاه داده ایرانشهر تصاویر اسامی شهرهای ایران است. بنابراین کلمات موجود در پایگاه داده ایرانشهر با کلمات در پایگاه داده فارسا متفاوت است به همین دلیل نیاز است برای ارزیابی روش‌ها با استفاده از پایگاه داده ایرانشهر، سیستم با این پایگاه داده آموزش ببیند و سپس ارزیابی شود. شکل 5-10 تعدادی از تصاویر موجود در پایگاه داده ایرانشهر را نشان می‌دهد.



شکل 5-10: نمونه‌های از تصاویر موجود در پایگاه داده ایرانشهر

برروی 198 کلاس از پایگاه داده ایرانشهر با اندازه کتاب رمز 49 و ضریب 1/8 برای تعداد حالت‌ها در HMM، سیستم آموزش داده می‌شود. SF برابر با 0/001 در نظر گرفته می‌شود. جدول 5- 12 نتایج این آزمایش را برای روش‌های پیشنهادی اول و دوم نشان می‌دهد.

جدول 5-12: ارزیابی نرخ بازشناسی روش های پیشنهادی با استفاده از پایگاه داده ایرانشهر

	هموارسازی با SF=0.001	Top-1	Top-5	Top-10	Top-20
روش پیشنهادی اول	-	31/02	48/90	56/02	63/92
روش پیشنهادی اول	ü	62/4	86/1	90/62	94/28
روش پیشنهادی دوم	ü	63/82	86/17	91/51	94/76

همانطور که مشاهده می‌شود قبل از هموار سازی با این پایگاه داده نرخ بازشناسی بسیار پایین است دلیل، این است که در پایگاه داده ایرانشهر تعداد نمونه‌ها برای هر کلاس پایین، حدود 30 تا 35 تا در مقابل پایگاه داده فارسا که در هر کلاس 100 نمونه وجود دارد که تقریبا 3 برابر پایگاه داده ایرانشهر، است. با این تعداد تصویر مدل‌های HMM نمی‌توانند به خوبی آموزش ببینند. همانطور که مشاهده می‌شود در این حالت مرحله هموارسازی پارامترهای HMM بسیار موثر عمل کرده‌است، دلیل این بهبود ملموس این است که زمانیکه تعداد داده ها برای یک کلاس کم باشد، به نسبت آن احتمال مشاهده یک بردار ویژگی کاهش می‌یابد و این باعث می‌شود در ماتریس احتمال مشاهدات مقدار صفر بگیرد. با مشاهده خروجی های این آزمایش قبل از مرحله هموار سازی، تعداد درایه هایی از ماتریس احتمال مشاهدات که مقدار صفر داشتند بسیار زیاد بود که این دلیل پایین بودن نرخ بازشناسی قبل از مرحله هموار سازی است. هموار سازی با استفاده از روشی که در فصل 3 شرح داده شد باعث حذف این مقادیر صفر شده و آنها را با یک احتمال بسیار کوچک جایگزین می‌کند، این راهکار باعث افزایش نرخ بازشناسی می‌شود.

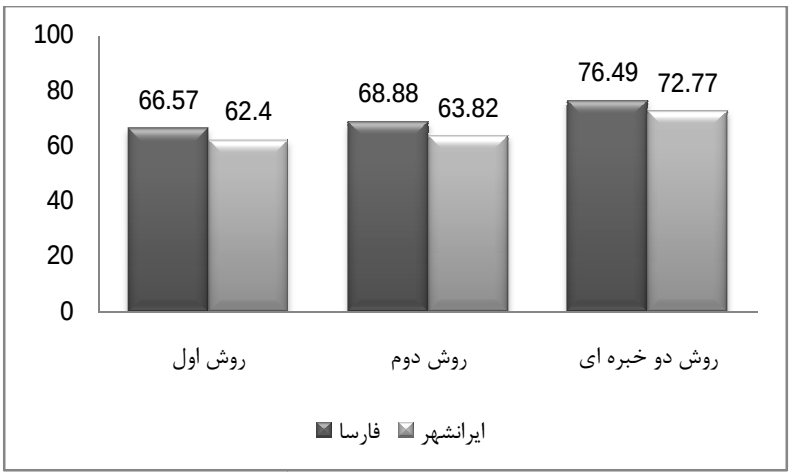
جدول 5-13 نتایج روش پیشنهادی دو خبره‌ای را نشان می‌دهد در این حالت آستانه برای معیار اطمینان 6 در نظر گرفته می‌شود. از میان کل تصاویر آزمون، 1910 ، تعداد تصویر900 شرایط معیار اطمینان را ندارند.

جدول5-13 نرخ بازشناسی روش دوخبره‌ای با استفاده از پایگاه داده ایرانشهر

کلاسه‌بند HMM	کلاسه‌بند KNN	کل سیستم
80/75	59/1	72/77

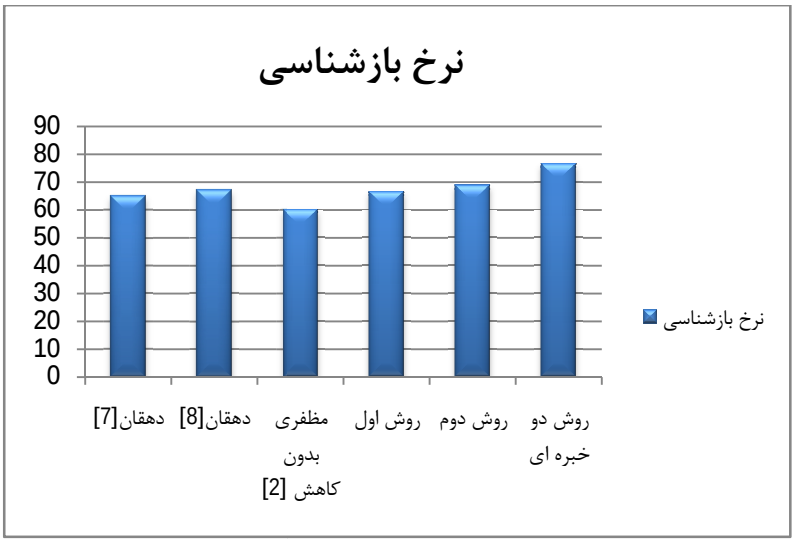
5-8 تحلیل نتایج و مقایسه با روش‌های موجود

در شکل 5-11 نمودار مقایسه نتایج بدست آمده از ارزیابی روش‌های پیشنهادی با استفاده از پایگاه‌های داده فارسا و ایرانشهر است.



شکل 5-11: مقایسه نتایج بدست آمده از روش‌های پیشنهادی در پایگاه داده فارسا و ایرانشهر

در شکل 5-12 نمودار مقایسه نتایج تحقیقات انجام شده در زمینه بازشناسی برون خط کلمات دست نویس فارسی با استفاده HMM را نشان می‌دهد. همانطور که مشاهده می‌شود روش‌های پیشنهادی در مقایسه با روش‌های موجود، عملکرد خوبی دارند.



شکل5-12: مقایسه نتایج نرخ بازشناسی روش‌های موجود و روش های پیشنهادی

فصل ششم:

نتیجه گیری و پیشنهادات

فصل ششم

6-1 نتیجه گیری

در این پایان‌نامه به مسئله بازشناسی برون‌خط کلمات دست نویس فارسی پرداخته شد. در حالت کلی یک سیستم بازشناسی کلمه دست نوشته شامل مراحل پیش پردازش، استخراج ویژگی و کلاسه‌بندی است. در این خصوص مروری بر کارهای گذشته انجام شد.

برای ارزیابی کارایی سیستم پیشنهادی به دلیل نبود یک پایگاه داده مناسب با تعداد تصویر کافی، یک پایگاه داده شامل 30000 تصویر از 300 کلمه متداول در زبان فارسی ایجاد شد. 198 کلاس از این پایگاه داده برای ارزیابی کارایی روش‌های پیشنهادی مورد استفاده قرار گرفت. البته آموزش و ارزیابی با استفاده از 198 کلاس از پایگاه داده ایرانشهر نیز، که تعداد تصاویر کمتری برای هر کلمه دارد، روی روش‌های پیشنهادی انجام شد.

با توجه به ساختار نوشتاری زبان فارسی و پیچیدگی‌های آن از کلاسه‌بند HMM استفاده شد . در حالت کلی سه روش پیشنهاد شد:

در روش اول از هیستوگرام کدهای زنجیره‌ای به عنوان بردار ویژگی استفاده شد. بردارهای ویژگی از پنجره لغزان تصویر استخراج می‌گردد، هر پنجره یک بردار 20 عنصری را تولید می‌نماید. از آنجایی که برای آموزش و ارزیابی HMM گسسته نیاز است دنباله ای از مشاهدات برای هر الگو بکار گرفته شود ، بردار ویژگی‌های استخراج شده از پنجره لغزان، با استفاده از شبکه SOM کوانتیزه می‌شود . پارامترهای که در این روش لازم است تنظیم شود شامل عرض پنجره در سیستم پنجره لغزان، تعداد حالت‌های مخفی برای مدل کردن HMM برای هرکدام از کلاس‌های کلمات، پارامتر هموار سازی ماتریس احتمال مشاهدات درHMM و اندازه کتاب رمز است. در مقایسه با سایر عرض‌های پنجره ، عرض پنجره برابر با 8 به نرخ بازشناسی بهتری دست می‌یابد. برای تعداد حالت‌های مخفی مدل

HMMدر هر کلاس کلمه ، کمترین تعداد پنجره برای تصاویر آموزش در هر کلاس بدست می‌آید. تعداد حالت‌ها ضریبی از این مقدار کمینه است. با آزمایش ضرایب مختلف، سیستم در مقدار 1/8 به بهترین نرخ بازشناسی می‌رسد. برای پارامتر هموارسازی روی مقادیر مختلف آزمایش شد، سیستم در مقدار 0/001 به جواب بهتری رسید. اندازه کتاب رمز 49 تنظیم می‌شود. در نهایت روش پیشنهادی اول با تعیین ضرایب بیان شده در بالا به نرخ بازشناسی 66/57% در198 کلاس از پایگاه داده فارسا می‌رسد.

در روش پیشنهادی دوم علاوه بر هیستوگرام کدهای زنجیره‌ای، از ویژگی میانگین بلوکی برای کلاسه بندی استفاده می‌شود و ابعاد بردارهای ویژگی به 25 افزایش می‌یابد. نرخ بازشناسی با استفاده از همان ضرایب تنظیم شده در روش اول، 68/88% بدست می‌آید. که افزایش بیش از 2% را نسبت به روش اول نشان می‌دهد.

روش پیشنهادی سوم در واقع بکارگیری یک سیستم دو خبره‌ای است. با استفاده از بررسی که روی نتایج ارزیابی روش دوم انجام شد یک مفهوم جدید به نام معیار اطمینان برای کلاسه‌بند HMM معرفی می‌شود. با مشاهده هیستوگرام اختلاف دو بزرگترین احتمال در خروجی HMM برای تصاویر آزمون یک مقدار آستانه به عنوان معیار اطمینان معرفی می‌شود. تصاویری که شرایط معیار اطمینان را دارند با کلاسه‌بند HMM بازشناسی می‌شوند و تصاویری که این شرایط را ندارند با استفاده از کلاسه‌بند KNN بازشناسی می‌شوند. برای کلاسه‌بند KNN از ویژگی‌های ساختار مانند استفاده می‌شود. کلاسه‌بند KNN با استفاده از این ویژگی‌ها و با 11 نزدیک‌ترین همسایه و معیار فاصله بلوک شهری به نرخ بازشناسی 61/69% دست می‌یابد. در این سیستم HMM برای تصاویری که معیار اطمینان را دارند به نرخ بازشناسی 85% دست می‌یابد. در کل نرخ بازشناسی این روش برای 198 کلاس از پایگاه داده فارسا 76/49% است. که نسبت به روش اول افزایش 7% را بدنبال دارد.

6-2 پیشنهادات

پیشنهاداتی که برای بهبود کارایی سیستم‌های برون خط بازشناسی کلمه دست‌نوشته و ادامه روش‌های ارائه شده در این پایان نامه ارائه می‌شود به شرح زیر می‌باشد:

- بالابردن تعداد کلمات موجود در پایگاه داده فارسا و همچنین تنوع دست نوشته در هر کلمه
- اعمال یک مرحله پیش پردازش دقیق برای حذف شکستگی‌های احتمالی در تصویر کلمه، زیرا وجود این شکستگی‌ها باعث انتشار خطا در مراحل بعدی می‌شود.
- ترکیب چند نوع ویژگی برای اینکه بتواند تمایز بیشتری بین نمونه‌های دو کلاس مختلف ایجاد نموده، سپس استفاده از روش‌های کاهش ویژگی مانند الگوریتم ژنتیک[1]، تجزیه و تحلیل مولفه‌های اصلی (PCA)[2] برای کاهش ابعاد بردار ویژگی پیشنهاد گردد.
- استفاده از نرم افزار HTK[3] برای مراحل کوانتیزه سازی بردار و آموزش و ارزیابی HMM.
- استفاده از یک مرحله کاهش واژه نامه قبل از کلاسه‌بند HMM زیرا چنانچه تعداد کلاس‌ها در پایگاه داده کم باشد، HMM کارایی بسیار بالایی دارد.
- برای بهبود عملکرد روش دو خبره‌ای در این پایان نامه، می‌توان قبل از استفاده از کلاسه بند KNN توسط ویژگی‌های ساختاری مانند نقاط و علائم در تصویر کلمه، کلمات نامشابه را از مرحله مقایسه در KNN حذف کرد.

[1] Genetic Algorithm
[2] Principle Component Analysis
[3] HMM ToolKit

مراجع

[1]شورای پژوهشی OCR کار گروه خط و زبان فارسی،1388، پژوهشنامه نویسه خوان نوری،تهران شورای عالی اطلاع رسانی دبیر خانه 1388

[2]Mozaffari S., Faez F., Margner V., El-Abed H.(2008)“Lexicon reduction using dots for off-line Farsi/Arabic handwritten word recognition”. Pattern Recognition Letters, Vol.29, PP.724–734.

[3]Haji, M. M. (2005), MSC thesis“Farsi Handwritten Word Recognition using Continous Hidden Markov Models and Structural Features”, Computer Engineering Shiraz University, Shiraz, Iran.

[4]Bahmani, Z., Alamdar, F., Azmi, R. (2010) “Off-Iine Arabic/Farsi Handwritten Word Recognition Using RBF Neural Network and Genetic algorithm”. IEEE Conference, 978-1-4244-6585-9

[5] Malik Waqas Sagheer, Chun Lei He, Nicola Nobile, Ching Y. Suen.(2010) “ Holistic Urdu Handwritten Word Recognition Using Support Vector Machine”, 1051-4651/10 © 2010 IEEE

[6]AlKhateeb, J., H., Khelifi, F., Jiang, J. (2009) “ A New Approach for Off-Line Handwritten Arabic Word Recognition Using KNN Classifier”, IEEE International Conference on Signal and Image Processing Applications.

[7]Dehghan M., Faez K., Ahmadi M. and Shridhar M. (2001) “Handwritten Farsi (Arabic) word recognition: a holistic approach using discrete HMM”, Pattern Recognition,vol. 34, no. 5, pp. 1057-1065

[8]Dehghan M., Faez K., Ahmadi M. and Shridhar M.,(2001b) “Unconstrained Farsi handwritten word recognition using fuzzy vector quantization and hidden Markov models”, Pattern Recognition Letter,Vol.2,PP. 209–214

[9]Alkhateeb, J.H., Ren, J., Jiang, J., Al-Muhtaseb, H. (2011) “Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking”. Pattern Recognition Letters 32 (2011) 1081–1088

[10]Ramy El-Hajj, Laurence Likforman-Sulem, Chafic Mokbel,(2005)“ Arabic Handwriting Recognition Using Baseline Dependant Features and Hidden Markov Modeling”, Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (*ICDAR’05*), 1520-5263/05 $20.00 © 2005 IEEE

[11]Volker Märgner , Haikal El Abed,(2012)“ Guide to OCR for Arabic Scripts”, DOI 10.1007/978-1-4471-4072-6, © Springer-Verlag London 2012

[12] علیپور سراجی م، (1391)، پایان نامه ارشد :" تشخیص برون خط کلمات دست نوشته فارسی به کمک بلوک بندی تطبیقی "، دانشکده برق و رباتیک، دانشگاه صنعتی شاهرود

[13]Jianming Hu, Donggang Yu, Hong Yan, (1996) “Algorithm for stroke width compensation of handwritten characters”, ELECTRONICS LETTERS 21st November 7996, Vol.32, No.2

[14]Yang Mingqiang, KpalmaKidiyo, Ronsin Joseph, (2008).“ A survey of shape feature extraction techniques”, Author manuscript, published in Pattern Recognition, Peng-Yeng Yin (Ed.) (2008) 43-90

[15]Www. users.utcluj.ro/~rdanescu/PI-L6e.pdf

[16]https://icpcarchive.ecs.baylor.edu/external/48/4884.pdf

[17]Alaei A., Nagabhushan P., Pal U., (2010). “A New Two-stage Scheme for the Recognition of Persian Handwritten Characters”, 12th International Conference on Frontiers in Handwriting Recognition

[18] Khorsheed M.S., (2007). “Offline recognition of omnifont Arabic text usingthe HMM ToolKit (HTK)”,Pattern Recognition Letters 28 (2007) 1563–1571
[19] http://www.ai-junkie.com/ann/som/som3.html
[20] Haykin, S. (1999). “9.Self-organizing maps”. Neural networks-A comprehensive foundation (2nd Ed.). Prentice-Hall. ISBN 0-13-908385-5
[21] Zhao Z., Rowden C.G., (1992). “Use of Kohonen self-organizing feature maps for HMM parameter smoothing in speech recognition”, IEE PROCEEDINCS-F, Vol. 139, No. 6
[22] Duda R. O., Hart P. E., Strok D. G., (2000) “Pattern Classification” 2nd Edition, Wiley, 2000, ISBN 978-0-471-05669-0
[23] Rigoll G., Kosmala A., Rottland J., Neukirchen, C.(1996) “A comparison between continuous anddiscrete density hidden Markov models for cursive handwriting recognition”, In: Proc. 13th International Conference on Pattern Recognition (ICPR’96),Vienna, Austria, pp. 39–48 (1996)
[24] Rabiner L. R., (1989)“A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Proceeding of the IEEE, Vol.77, No.2
[25] Mozaffari S., El Abed H., Margner V., Faez K., Amirshahi, A., (2008) “IfN/Farsi-Database: A Database of Farsi Handwritten City Names”. In: Proceedings of 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 11), Montreal, Canada, pp. 397–402
[26] Bidgoli A.M., Sarhadi, M., (2008) “IAUT/PHCN: Islamic Azad University of Tehran/ Persian Handwritten City Names, A very large database of handwritten Persian word”. In: Proceedings of 11th International Conference on Frontiers in Handwriting Recognition (ICFHR11), Montreal, Canada, pp. 192–197
[27] Haghighi P. J., Nobile N., He C.L., Suen C.Y., (2009) “ A New Large-Scale Multi-purpose Handwritten Farsi Database”. In: ICIAR 2009, LNCS 5627, pp. 278–286,
[28] ایمانی، ز، احمدی فرد، ع.،زهره وند، ع.(1391)، "معرفی پایگاه داده فارسا: نصاویر دیجیتال از کلمات دست‌نویس فارسی"، یازدهمین کنفرانس سیستم‌های هوشمند، تهران دانشگاه خوارزمی
[29] http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html
[30] http://stn.spotfire.com/spotfire_client_help/hc/hc_distance_measures_overview.htm

Abstract

In this thesis we address the problem of recognizing Farsi handwritten words. Three methods have been proposed to implement this system. In the first method, the image of a word is divided into overlapped vertical stripes and for each strip chain-code of word boundary are extracted and used as features. The extracted feature vectors are coded using Self Organizing Map vector quantization. The codes of quantized vectors are then used for training the model of each word in the database. We model each word in database using discrete Hidden Markov Models (HMM).

In the second method, we extract two types of features from each word in the training and evaluation phase, chain code of boundary and distribution of foreground density across the image word.

To improve the performance of classification system we utilize the provided confidence measure by HMM for each test data. For those test data (feature vector of input words) which the confidence measure are below a pre-define threshold, we use k-nearest neighbor classifier as tandem to decide to which class the data point belongs.

In order to evaluate the performance of the proposed system we conducted an experiment using a new prepared database FARSA. We tested the proposed method using 198 word classes in this database.

Result of experiment three proposed methods showed first method achieved 66.57% with considered codebook size 49 and smoothing factor 0.001, then with same parameters value second method evaluated that result of this experiment showed 2% improvement than first method. Finally in third proposed method, we experimentally set a proper threshold (threshold is equal to 5) for confidence measure, then we evaluated the system with same previous parameters and several value for k in k-nearest neighbor classifier. The best result obtained 61.69% with city-block distance and k equal to 11. So total recognition rate equal to 76.49% that showed 7% improvement than first method. This result showed the performance of proposed systems is considerably better than the performance of existing methods.

Keywords: Handwritten Farsi word recognition, FARSA database, chain-code histogram, Self Organization Map, Hidden Markov Models, Confidence measure, K-Nearest Neighbors

Shahrood University of Technology

Faculty of Electrical and Robatic

Offline Farsi Handwritten Word Recognition Using Hidden Markov Model

Zahra Imani

Supervisor:

Dr.Alireza Ahmadifard

Assistant:

Dr. Hossein Khosravi

FOR THE DEGREE OF

MASTER OF SCIENCE

September 2013