

اللَّهُمَّ صَلِّ وَسَلِّمْ وَبَارِكْ عَلَى سَيِّدِنَا مُحَمَّدٍ





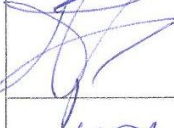

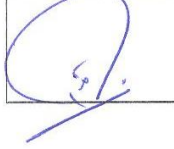
تاییدیه اعضای هیات داوران حاضر در جلسه دفاع از پایان نامه کارشناسی ارشد

خانم فاطمه ولایتی پایان نامه ۶ واحدی خود را با عنوان خوشه بندی سبک نگارش

دستنوشته برون خط فارسی در تاریخ ۱۳۹۰/۱۰/۲۵ ارائه کردند.

اعضای هیات داوران نسخه نهایی این پایان نامه را از نظر فرم و محتوا تایید کرده، پذیرش آنرا

برای اخذ درجه کارشناسی ارشد الکترونیک پیشنهاد می کنند.

امضا	رتبه علمی	نام و نام خانوادگی	عضو هیات داوران
	استاد	دکتر احسان اله کبیر	استاد راهنما
	دانشیار	دکتر نصراله مقدم چرکری	استاد ناظر
	استادیار	دکتر محمد صنیعی آباده	استاد ناظر
	دانشیار	دکتر کامبیز بدیع	استاد ناظر
	دانشیار	دکتر نصراله مقدم چرکری	مدیر گروه (پا نماینده گروه تخصصی)

دستورالعمل حق مالکیت مادی و معنوی در مورد نتایج پژوهشهای علمی دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاستهای پژوهشی دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیات علمی، دانشجویان، دانش آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهشهای علمی که تحت عناوین پایان نامه، رساله و طرحهای تحقیقاتی که با هماهنگی دانشگاه انجام شده است، موارد ذیل را رعایت نمایند:

ماده ۱- حقوق مادی و معنوی پایان نامهها / رسالههای مصوب دانشگاه متعلق به دانشگاه است و هرگونه بهره‌برداری از آن باید با ذکر نام دانشگاه و رعایت آیین نامه‌ها و دستورالعمل‌های مصوب دانشگاه باشد.

ماده ۲- انتشار مقاله یا مقالات مستخرج از پایان نامه / رساله به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی باید به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشد.

تبصره: در مقالاتی که پس از دانش آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان نامه / رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳- انتشار کتاب حاصل از نتایج پایان نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آئین نامه‌های مصوب انجام می‌شود.

ماده ۴- ثبت اختراع و تدوین دانش فنی و یا ارائه در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق حوزه پژوهشی دانشگاه انجام گیرد.

ماده ۵- این دستورالعمل در ۵ ماده و یک تبصره در تاریخ ۱۳۸۴/۴/۲۵ در شورای پژوهشی دانشگاه به تصویب رسیده و از تاریخ تصویب لازم الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری می‌شود.

نام و نام خانوادگی *فاطمه صالحی*
امضاء *[Signature]*
۱۳۹۲/۲

آیین نامه چاپ پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله) ی خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند:

«کتاب حاضر، حاصل پایان نامه کارشناسی ارشد/ رساله دکتری نگارنده در رشته است که در سال در دانشکده در دانشگاه تربیت مدرس به راهنمایی سرکار خانم/جناب آقای دکتر مشاوره سرکار خانم/جناب آقای دکتر و مشاوره سرکار خانم/جناب آقای دکتر از آن

دفاع شده است.»

ماده ۳: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه اهدا کند. دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر در معرض فروش قرار دهد.

ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده رابه عنوان خسارت به دانشگاه تربیت مدرس، تأدیه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقیف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب دانشجوی رشته

مقطع

تعهد فوق و ضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

نام و نام خانوادگی: فاطمه رسانی

تاریخ و امضا: ۱۳۹۱/۲/۲۰



دانشگاه تربیت مدرس
دانشکده مهندسی برق و کامپیوتر

پایان نامه دوره کارشناسی ارشد مهندسی برق-الکترونیک

خوشه‌بندی سبک نگارش دست‌نوشته برون خط فارسی

فاطمه ولایتی

استاد راهنما:

جناب آقای دکتر کبیر

دی ۱۳۹۰

تقدیم

تقدیم به مهربانترین‌های زندگی‌م

پدرم، مادرم و دیگر اعضای خانواده‌ام

تشکر و قدردانی

حمد خدایی را که ذات لایزالش ازلی است و پایان نپذیرفته جاویدان، آنکه دیده کوتاه بین بینندگان یارای دیدارش را ندارد و فکر توصیف کنندگان از عهده وصف او برنیاید. سپاس خدای را که درس خدانشناسی به ما آموخت و درهای علم به ربوبیت خویش را بر ما گشود، حمد او را که ابزار گستردن را در ساختمان ما قرار داد و ابزار جمع کردن را در اختیارمان گذاشت و از نعمت زندگی ما را بهره مند ساخت و اعضای ما را برای فعالیت کامل نمود.

در این مجال بر خود لازم می‌دانم، از زحمات و راهنمایی‌های ارزنده جناب آقای دکتر کبیر که در طول مدت تحصیل، همواره مشوق و راهنمای بنده بودند، تشکر کنم.

از اساتید هیئت محترم داوران، جناب آقای دکتر بدیع، جناب آقای دکتر مقدم و جناب آقای دکتر صنیعی آبا، تشکر می‌کنم.

از جناب آقای دکتر ولی‌زاده و خانم هما داوودی و سایر دوستان در آزمایشگاه سیستم‌های الکترونیک، بابت همکاری و مساعدت هایشان، سپاسگزارم.

چکیده

هدف این پایان‌نامه، یافتن و استخراج ویژگی‌هایی است که بر مبنای آن بتوان دست‌خط فارسی را خوشه‌بندی کرد. در این کار، در ابتدا بر روی ویژگی‌های مبتنی بر بافت، تمرکز شده است. این ویژگی‌ها شامل دو دسته ویژگی آماری ماتریس باهم‌آیی و ویژگی مبتنی بر تبدیل گابور است. برای استخراج این ویژگی‌ها، یک بافت مناسب در ابعاد 1024×1024 مستقل از محتوای سند، از تصویر دستنوشته ایجاد می‌شود. از ویژگی‌های دیگری که در این کار از آن‌ها استفاده شده است، تعدادی ویژگی ساختاری مبتنی بر منحنی پیرامونی است. این ویژگی‌ها را از هر یک از تصاویر موجود در یک مجموعه داده از ۹۷ دستنوشته فارسی که دارای متون متفاوتی بودند، استخراج کردیم و از الگوریتم k میانگین و شبکه عصبی نگاشت ویژگی خود سامان، برای خوشه‌بندی این ویژگی‌ها استفاده شده است.

برای ارزیابی این ویژگی‌ها، یک روش ارزیابی بر مبنای الگوریتم خوشه‌بندی k میانگین، طراحی کرده‌ایم. در این الگوریتم از معیار مقایسه باینری ژاکارد استفاده کرده‌ایم، هم‌چنین برای محاسبه مراکز خوشه در هر دوره تکرار از الگوریتم k میانگین، از روش انتخاب داده چرخ رولت، بهره گرفته‌ایم. نتایج بدست آمده، نشان می‌دهد با ترکیب دو نوع از ویژگی‌های مبتنی بر منحنی پیرامونی، نرخ خوشه‌بندی، ۷۵ درصد است که نسبت به سایر روش‌های مورد استفاده در این کار، نرخ بهتری را در برداشته است.

کلید واژه: سبک نگارش، خوشه‌بندی، بافت، فیلتر گابور، ماتریس باهم‌آیی، منحنی پیرامونی، ژاکارد،

چرخ رولت

Abstract

In this thesis, we proposed some features which are suitable for clustering of off-line Farsi handwriting styles. Some of these features are texture-based, such as co-occurrence and Gabor based features. For extract these features, we create a texture of size 1024*1024 from each handwritten image that is independent of text content of image. Other features are contour based. We extract all of the above feature from the database that collected from 97 different people in different texts and then apply k-means and Self organization feature map algorithms for clustering these feature and thus for clustering of handwriting styles.

For evaluation of this work, we designed k-means algorithm based on Jaccard binary matching criterion and Roulette wheel function, respectively in matching step and cluster center updating step in each iteration of k-means algorithm.

Obtained results show the combination of two types contour- based features, have better handwritten clustering rate (75%) than other types of features that we used in this work.

Keywords: Style, Clustering, texture, Gabor Filter, Co- occurrence Matrix, Contour, Jaccard, Roulette Wheel.



Clustering of Off-line Farsi Handwriting Styles

Thesis

Submitted in Partial Fulfillment of the
Requirements for the Degree of Master of Science (M.Sc.)
in Electrical Engineering,

Department of Electrical and Computer Engineering
School of Engineering
Tarbiat Modares University

By:

Fatemeh Velayati

Supervisor:

Ehsanollah Kabir

January 2011

فهرست مطالب

عنوان	صفحه
فهرست جدول‌ها.....	ج
فهرست شکل‌ها.....	د
فصل ۱- بازشناسی متون دستنویس فارسی.....	۱
۱-۱- مقدمه.....	۱
۱-۱-۱- بازشناسی.....	۱
۱-۱-۲- تأیید امضا.....	۲
۱-۱-۳- تأیید و تشخیص نویسنده.....	۲
۱-۱-۴- شخصیت‌شناسی (گرافولوژی).....	۳
۱-۱-۵- هدف پایان‌نامه.....	۳
۱-۲- جمع‌بندی فصل و ساختار پایان‌نامه.....	۵
فصل ۲- مروری بر کارهای انجام شده.....	۶
۱-۲- جمع‌بندی فصل.....	۱۵
فصل ۳- خوشه‌بندی خودکار دست‌خط فارسی.....	۱۶
۱-۳- مقدمه.....	۱۶
۲-۳- پیش‌پردازش.....	۱۶
۳-۳- استخراج ویژگی.....	۲۱
۱-۳-۳- ویژگی‌های مبتنی بر بافت.....	۲۱
۲-۳-۳- ویژگی‌های ساختاری.....	۳۰
۴-۳- خوشه‌بندی.....	۳۷

۳۷	الگوریتم K-means [۲۶]
۳۸	نگاشت خودسامان کوهونن [۲۷]
۴۰	جمع‌بندی فصل
۴۲	فصل ۴ - ارزیابی نتایج
۴۲	۱-۴ - مقدمه
۴۲	۲-۴ - ایجاد خوشه‌های معتبر
۴۳	۱-۲-۴ - خوشه‌بندی دستنوشته‌های موجود در مجموعه داده، توسط هشت نفر
۴۳	۲-۲-۴ - ماتریس باینری
۴۵	۳-۲-۴ - تعیین میزان شباهت بین ۸ خوشه‌بندی
۵۳	۳-۴ - گزارش نتایج
۵۷	۴-۴ - جمع‌بندی فصل
۵۸	فصل ۵ - جمع‌بندی و پیشنهاد
۶۰	ضمیمه أ - خوشه‌بندی ۹۷ دستنوشته توسط یکی از افراد
۷۷	فهرست مراجع
۷۹	واژه نامه فارسی به انگلیسی
۸۰	واژه نامه انگلیسی به فارسی

فهرست جدول‌ها

صفحه	عنوان
۳۰	جدول ۱-۳: پارامترهای فیلتر گابور در این کار.....
۴۴	جدول ۱-۴: جدول باینری بر اساس نمونه‌های موجود در هر خوشه.....
۴۴	جدول ۲-۴: جدول باینری بر اساس نمونه‌های موجود در هر خوشه.....
۴۷	جدول ۳-۴: روش‌های مقایسه داده‌های باینری.....
۵۳	جدول ۴-۴: ویژگی‌های استفاده شده برای خوشه‌بندی.....

فهرست شکل‌ها

عنوان	صفحه
شکل ۱-۱: طبقه‌بندی مباحث در تحلیل متون دستنویس	۱
شکل ۲-۱: طرح‌واره‌ای از سیستم‌های (الف): تأیید نویسنده، (ب): تشخیص نویسنده.....	۲
شکل ۳-۱: تفاوت در دست‌خط افراد؛ شکل (۱) و (۴) دست‌خط مربوط به نفر اول. شکل (۲) و (۵) دست‌خط مربوط به نفر دوم. شکل (۳) و (۶) دست‌خط مربوط به نفر سوم.....	۴
شکل ۱-۱: مراحل استخراج ویژگی برای نویسه عبری "الف". (الف): (۱): استخراج نواحی پیش زمینه، (۲) استخراج پوسته محدب (۱). (۳): نواحی پیش زمینه شکل (۲) که در شکل (۱) پیش زمینه نیستند. (ب) استخراج چهار ناحیه (R1-R4).....	۷
شکل ۲-۱: چهار شاخک استخراج شده از کلمات.....	۱۰
شکل ۳-۱: (الف): تصویر یک شاخک. (ب) بردار ویژگی مربوط به (الف)	۱۱
شکل ۴-۱: نتیجه تبدیل همگر به کار گرفته شده برای نویسه a.....	۱۲
شکل ۵-۱: ایجاد ۵ زیرتصویر از نویسه a.....	۱۲
شکل ۶-۱: (الف) سه سبک نگارشی مختلف. (ب) رخداد نمونه‌های مختلف از سه نویسه برای هر سبک. (خانه‌های که پررنگ تر هستند، رخداد بیش تر را نشان می‌دهد).....	۱۳
شکل ۱-۱: (الف) یک تصویر به همراه هیستوگرام افقی مربوط به آن. (ب) خطوط متن استخراج شده از تصویر فوق.....	۱۷
شکل ۲-۱: منحنی تصویر و خط برازش شده متناظر با آن.....	۱۹
شکل ۳-۱: اصلاح کجی خط زمینه.....	۲۰
شکل ۴-۱: (الف) تصویر اصلی. (ب) بافت حاصل از تصویر (الف).....	۲۰

- شکل ۱-۵: ماتریس چپ: تصویر باینری. ماتریس راست: ماتریس باهم‌آیی مربوط به تصویر..... ۲۳
- شکل ۱-۶: دو دست‌خط متفاوت از کلمه "آب"..... ۲۳
- شکل ۱-۷: (الف): تصویر اصلی. (ب)، (پ) و (د): تصویر (الف) پس از دو، سه و چهار برابر کردن عرض قلم..... ۲۴
- شکل ۱-۸: دو تصویر با بافت‌های مختلف در نیمه بالائی و پائینی در هر کدام ولی با آمارگان مرتبه ۲ یکسان در کل تصویر..... ۲۸
- شکل ۱-۹: نمایش همسایگی‌های هشت و چهار..... ۳۱
- شکل ۱-۱۰: توصیف نموداری از ویژگی تابع توزیع احتمال جهت منحنی پیرامونی..... ۳۳
- شکل ۱-۱۱: تصاویر دست‌نوشته‌های مربوط به سه فرد مختلف..... ۳۳
- شکل ۱-۱۲: مقایسه مقدار ویژگی تابع توزیع احتمال جهت منحنی پیرامونی برای دست‌نوشته‌های شکل (۱۱-۳). (الف): شکل قرمز رنگ برای دست‌نوشته اول و شکل سیاه رنگ برای دست‌نوشته دوم. (الف): شکل قرمز رنگ برای دست‌نوشته سوم و شکل سیاه رنگ برای دست‌نوشته دوم..... ۳۴
- شکل ۱-۱۳: توصیف نموداری از ویژگی تابع توزیع احتمال لولای منحنی پیرامونی..... ۳۵
- شکل ۱-۱۴: مقایسه مقدار ویژگی تابع توزیع احتمال لولای منحنی پیرامونی برای دست‌نوشته‌های شکل (۱۱-۳). (الف): مقدار ویژگی مربوط به شکل (۱۱-۳). (ب): مقدار ویژگی مربوط به شکل (۱۱-۳). (ج) مقدار ویژگی مربوط به شکل (۱۱-۳)..... ۳۵
- شکل ۱-۱۵: توصیف نموداری از تابع توزیع احتمال باهم‌آیی در جهت افقی..... ۳۶
- شکل ۱-۱۶: ساختار یک شبکه کوهونن ۴×۴..... ۳۹
- شکل ۱-۱۷: کاهش طول همسایگی با گذر زمان..... ۴۰
- شکل ۱-۱۸: محاسبه اوزان در هر دوره تکرار از الگوریتم..... ۴۰
- شکل ۱-۴: درصد خوشه‌بندی برای ویژگی‌های گوناگون..... ۵۴
- شکل ۴-۲: سه عدد از نمونه‌های قرار گرفته در یک خوشه..... ۵۵

شکل ۱-۳: سه عدد از نمونه‌های قرار گرفته در یک خوشه. ۵۵.....

شکل ۱-۴: دو عدد از نمونه‌های قرار گرفته در یک خوشه. ۵۵.....

شکل ۱-۵: دو عدد از نمونه‌های قرار گرفته در یک خوشه. ۵۵.....

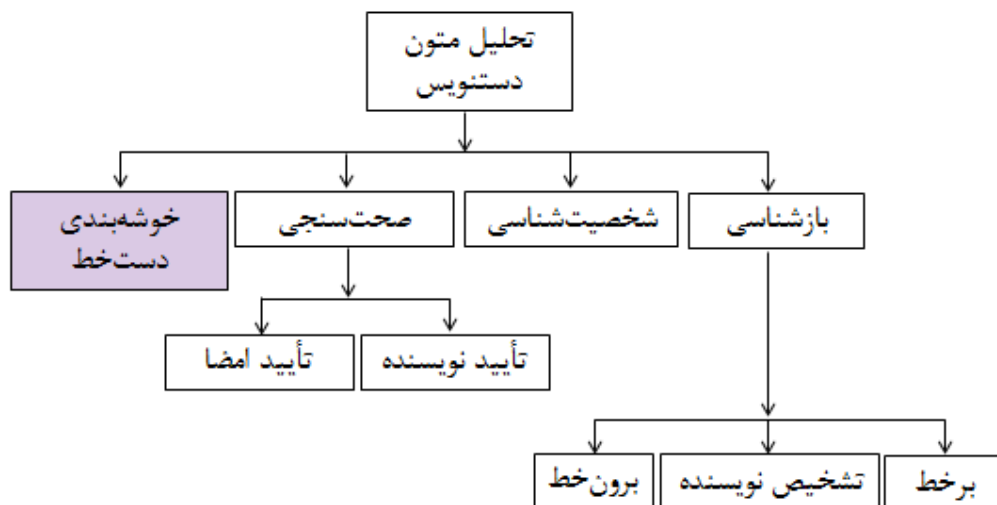
شکل ۱-۶: نمونه‌های قرار گرفته در یک خوشه. ۵۶.....

شکل ۱-۷: دو عدد از نمونه‌های قرار گرفته در یک خوشه. ۵۷.....

فصل ۱ - بازشناسی متون دستنویس فارسی

۱-۱-۱ مقدمه

پردازش نوشتار یکی از شاخه‌های بازشناسی الگو است که در آن، با در اختیار داشتن تصویر یک سند، به بازشناسی محتوای آن می‌پردازیم. در شکل (۱-۱)، نموداری از زمینه‌های مورد بحث در پردازش نوشتار، آورده شده است.



شکل ۱-۱: طبقه‌بندی مباحث در تحلیل متون دستنویس [۱]

۱-۱-۱-۱ بازشناسی

هدف از بازشناسی، تبدیل یک متن از فضای تصویر به فضای نمادین است. همانطور که در شکل (۱-۱) نشان داده شد، بازشناسی سند به دو صورت بر خط^۱ و برون خط^۲ انجام می‌شود. ورودی سیستم در روش بازشناسی برخط به صورت پیوسته و همزمان با عمل نوشتن نویسنده بر روی یک صفحه رقمی

^۱ on-Line

^۲ off-Line

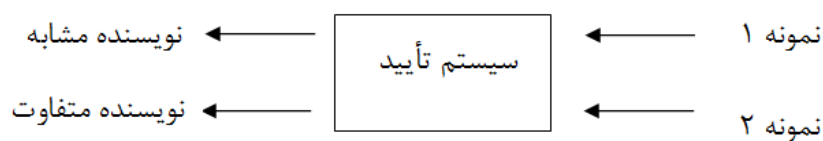
کننده^۱ به سیستم وارد می‌شود. در این روش، از اطلاعات زمانی و مکانی قلم، سرعت نوشتن و میزان فشار قلم بر صفحه، می‌توان برای بازشناسی استفاده کرد. ورودی سیستم در روش برون خط، تصویر روبش^۲ شده متن دستنویس است.

۱-۱-۲- تأیید امضا

امضا یک روش تصدیق است که مورد پذیرش جوامع مختلف است و بطور عمده به عنوان یک سند در زندگی روزمره ما قرار می‌گیرد. تأیید امضا^۳، یکی از موضوعات جالب در زمینه بازشناسی الگو است که تحقیقات زیادی پیرامون آن انجام شده است. جرم‌شناسی، یکی از کاربردهای مهم تأیید امضا است.

۱-۱-۳- تأیید و تشخیص نویسنده

تأیید نویسنده، یک مقایسه بین دو دستنوشته است که بر اساس آن مشخص می‌شود که نویسنده دو دستنوشته یک فرد بوده است یا خیر. اما تشخیص نویسنده^۴، یک جستجو در مجموعه‌ای از تعدادی دستنوشته با نویسنده‌های معین است که طی آن مشخص می‌شود که یک نمونه دستنوشته به کدام یک از نویسنده‌ها تعلق دارد. در شکل زیر، طرح‌واره‌ای از سیستم‌های تأیید و تشخیص نویسنده، آورده شده است. از کاربردهای مهم این حوزه، همانند تأیید امضا، جرم‌شناسی است.



(الف)

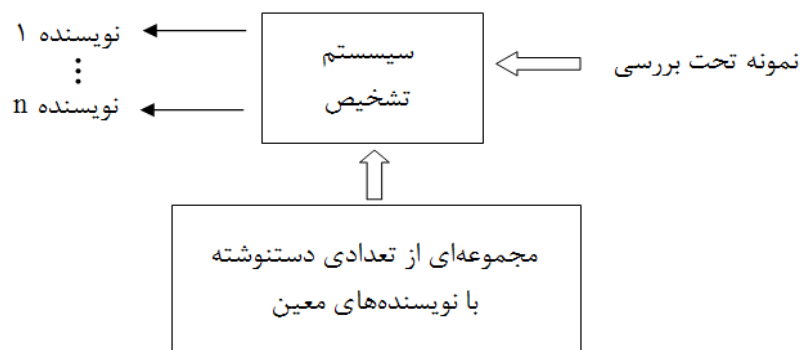
شکل ۱-۲: طرح‌واره‌ای از سیستم‌های (الف): تأیید نویسنده، (ب): تشخیص نویسنده [۱]

¹ Tablet Digitizer

² Scan

³ Signature Verification

⁴ Writer Identification



(ب)

ادامه شکل ۱-۲: طرح‌واره‌ای از سیستم‌های (الف): تأیید نویسنده، (ب): تشخیص نویسنده

۱-۱-۴- شخصیت‌شناسی (گرافولوژی)

گرافولوژی^۱، علم مطالعه و بررسی شخصیت و خصوصیات فردی بر اساس نحوه نوشتن است. بطور کلی، مهم‌ترین کاربرد گرافولوژی را می‌توان در استخدام متقاضیان در نهادهای مختلف دانست. در مبحث گرافولوژی، با دانستن قواعدی که در مسائل روانشناسی، برای شناخت شخصیت افراد استفاده می‌شود، برای شخصیت‌شناسی خودکار، ویژگی‌های دست‌خط افراد استخراج می‌شود. از جمله این ویژگی‌ها، می‌توان به درشت یا ریزنویسی، اشاره کرد، در روانشناسی افرادی که خط بسیار درشتی دارند، افراد اجتماعی تلقی می‌شوند، در حالی که، افرادی که خیلی ریز می‌نویسند، افرادی هستند که تمایل به تنهایی دارند. به همین ترتیب، ویژگی‌های دیگری نیز، می‌توان استخراج کرد.

۱-۱-۵- هدف پایان‌نامه

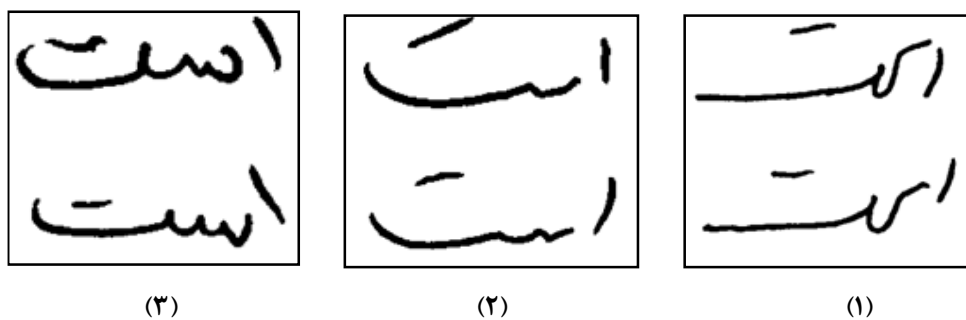
موضوع این پایان‌نامه خوشه‌بندی خودکار دست‌خط^۲ فارسی است، به عبارت دیگر می‌خواهیم میزان شباهت دست‌خط افراد را مشخص کنیم. ما می‌خواهیم با در دست داشتن دست‌خط افراد که از هیچ

¹ Graphology

² Style

قید خاصی پیروی نمی‌کنند و با استفاده از یک روش بدون ناظر^۱، دست‌خط‌های مشابه را خوشه‌بندی کنیم. بنابراین به دنبال روش‌هایی برای استخراج ویژگی‌ها هستیم که این خوشه‌بندی را با بالاترین کارایی انجام دهند. نکات قابل توجه در این مساله، علاوه بر انتخاب ویژگی‌های مناسب و روشی برای خوشه‌بندی، نحوه ارزیابی نتایج است. در شکل (۱-۳) دو نمونه از دست‌خط‌های سه فرد برای دو کلمه "است" و "زندگی" آورده شده است. در هر کادر دو نمونه از دست‌خط یک فرد آورده شده است، همانطور که مشاهده می‌شود، علاوه بر تفاوتی که بین دست‌خط افراد مختلف وجود دارد، نمونه‌های مربوط به یک فرد هم، کاملاً مشابه هم نیستند. از جمله ویژگی‌هایی که دست‌خط‌های آورده شده را از هم متمایز می‌کند، دندان‌دار نوشتن و یا بدون دندان‌دار نوشتن نویسه‌ها (مثل حرف س)، کشیده یا فشرده نوشتن آن‌ها (مثل حرف ت)، میزان انحنا^۲ی نویسه‌ها (مثل حرف ی در کلمه زندگی) است.

خوشه‌بندی خودکار دست‌خط می‌تواند به عنوان یک مرحله پیش پردازشی در بازشناسی نوشتار و با هدف افزایش نرخ بازشناسی استفاده شود. به عبارتی پیش از بازشناسی چنانچه بتوان، دست‌خط‌های مشابه را در یک خوشه قرار داد، برای هر خوشه می‌توان روش‌های بازشناسی متناسب با خصوصیات همان خوشه به کار برد. از کاربردهای دیگر این موضوع، شخصیت شناسی افراد است، هم‌چنین می‌توان در شناسایی نویسنده، روشی بر مبنای دست‌خط، ارائه کرد.



شکل ۱-۳: تفاوت در دست‌خط افراد؛

¹ Unsupervised

² Curvature