

چکیده

در این تحقیق شیوه‌ای از تشخیص الگوی دست‌نویس فارسی به روش برخط معرفی شده است که در آن واحدهای تشخیصی حروف مجزای فارسی هستند. در ادامه عملکرد سیستم در قالب دو فاز اصلی آموزش و تشخیص توضیح داده شده است.

فاز آموزش، شیوه اجرا در این فاز بدین صورت است که ابتدا حروف بر اساس شکل بدنه اصلی آنها به چند دسته مختلف تقسیم‌بندی می‌شود به عنوان نمونه حروف «ب»، «پ»، «ت» و «ث» به خاطر یکسانی بدنه اصلی در یک دسته قرار می‌گیرند، سپس هر یک از این دسته‌ها با استفاده از روش مدل مخفی مارکوف و یک مجموعه از نمونه‌ها آماده شده، به سیستم آموزش داده می‌شود و سیستم پارامترهای لازم برای مدل‌سازی هریک از این دسته‌ها را ذخیره می‌کند.

در فاز تشخیص ابتدا بدنه اصلی داده ورودی به هریک از این مدل‌ها داده می‌شود تا مشخص شود بیشترین تطابق با کدام دسته برقرار است، سپس با استفاده از سایر قسمت‌های حروف (ریزحرکات) مثل نقطه‌ها، سرکش‌ها و ... در قالب یک درخت تصمیم دقیقاً مشخص می‌شود که حرف مورد نظر کدام عضو از اعضای آن دسته می‌باشد.

نتایج این پژوهش بر اساس یک داده‌های پایگاه با عنوان Online-TMU نشان می‌دهد تشخیص بدنه اصلی بر اساس این دسته‌بندی تا ۹۳ درصد درست صورت می‌گیرد. اما این تشخیص برای ریز حرکات به ۸۰ درصد کاهش یافته و علت آن نیز عدم انحصار در شکل آنهاست، از این رو برای تشخیص حروف یک گروه از یکدیگر، اولویت در محل ریز حرکات و تعداد آنهاست و نه تشخیص شکل آنها. به عنوان نمونه برای تشخیص «ک» و «گ» از یکدیگر تنها تعداد حرکات قلم که در «ک» ۲ و در «گ» ۳ است معیار قرار گرفته.

کلمات کلیدی: بازشناسی کلمات، دست‌نویس، برخط، مدل مخفی مارکوف