سومین کنفرانس فناوری اطلاعات و دانش
۶ تا ۸ آذرماه ۱۳۸۶
IKT 2007 کنفرانس بین المللی فناوری اطلاعات و دانش
دانشگاه فردوسی مشهد

بهبود مدل مخلوط گوسی با استفاده از خوشه‌بندیِ C-میانگینهٔ فازیِ وزن‌دارِ مقاوم

امیرحسین حاج‌احمدی، محمد مهدی همایون‌پور و سید محمد احدی

دانشگاه صنعتی امیرکبیر، دانشکده مهندسی کامپیوتر و فناوری اطلاعات، آزمایشگاه سیستمهای هوشمند صوتی-گفتاری

homayoun@aut.ac.ir

چکیده – مدل مخلوط گوسی یکی رایج‌ترین روش‌های طبقه‌بندی در سیستم‌های خودکار بازشناسی گفتار و گوینده محسوب می‌شود. متأسفانه دقت بازشناسی مدلهای مخلوط گوسی، وابستگی شدیدی به مراکز اولیهٔ مخلوط‌های گوسی دارند که این وابستگی در مواردی کارایی سیستم‌های بازشناسی گفتار و گوینده را تحت تأثیر قرار می‌دهد. امروزه اغلب برای رفع این وابستگی از روش‌خوشه‌بندی C-میانگینهٔ سخت استفاده می‌شود. روش خوشه‌بندی C-میانگینهٔ سخت یا همان الگوریتم *K-means* حساسیت شدیدی به داده‌های دورافتاده دارد که می‌توانند کارایی آن را کاهش دهند. با توجه به این حقیقت که در مجموعهٔ بردارهای ویژگی گفتاری نیز همواره تعدادی دادهٔ دورافتاده وجود دارند، استفاده از روش خوشه‌بندی دیگری که نسبت به دادگان دورافتاده از حساسیت کمتری برخوردار باشد، می‌تواند بر دقت سیستم‌های بازشناسی گفتار و گوینده بیفزاید. به همین دلیل در این مقاله سعی شده است تا تأثیر استفاده از روش خوشه‌بندی C-میانگینهٔ فازی و تعدادی از روش‌های خوشه‌بندی مقاوم در مقابل دادگان دورافتاده مبتنی بر آن مانند خوشه‌بندی C-میانگینهٔ فازی-اعتباری، خوشه‌بندی C-میانگینهٔ فازی وزندار مبتنی بر چگالی و روش جدید خوشه‌بندی C-میانگینهٔ فازیِ وزن‌دارِ مقاوم، بجای روش خوشه‌بندی C-میانگینهٔ سخت در سیستم‌های بازشناسی گوینده مورد بررسی قرار بگیرد. نتایج آزمایش‌های انجام شده، نشان دهندهٔ افزایش دقت سیستمهای بازشناسی گوینده با استفاده از روش‌های خوشه‌بندی مقاوم در مقابل دادگان دورافتاده بخصوص روش خوشه‌بندی C-میانگینهٔ فازیِ وزندار بجای روش خوشه‌بندی C-میانگینهٔ سخت در تولید مدلهای مخلوط گوسی است.

کلید واژه-بازشناسی خودکار گوینده، خوشه‌بندی C-میانگینهٔ فازی، دادگان دورافتاده، مدل مخلوط گوسی.

۱- مقدمه

سیگنال گفتار حاوی چندین سطح اطلاعاتی است. در سطح نخست حاوی اطلاعات گفتاری شامل واژه‌ها و پیغام‌های بیان شده است. ولی در سطح دوم حاوی اطلاعات مربوط به خصوصیات گویندهٔ گفتار شامل ویژگیهای مجرای گفتاری، احساسات و ... است. با توجه به هزینه کم و سهولت انتقال صوت از طریق تلفن، امروزه استفاده از آن در تشخیص هویت افراد مورد توجه قرار گرفته است. اولین سیستم بازشناسی گوینده در دههٔ ۱۹۶۰ در حدود ۱۰ سال پس از ایجاد اولین سیستم بازشناسی گفتار تولید شد. علی‌رغم بهبود بسیار ایجاد شده در کارایی سیستم‌های خودکار بازشناسی گوینده این موضوع همچنان یکی از زمینه‌های فعال تحقیقاتی محسوب می‌شود که همچنان تلاش‌های بسیاری در این زمینه انجام می‌گیرد [۱].

به منظور افزایش دقت سیستم‌های بازشناسی گفتار و گوینده تلاش‌های بسیاری انجام شده است. که در گروهی از آنها سعی شده، مستقیماً از گفتار ویژگی‌های کاراتری استخراج شوند. در دسته‌ای دیگر تلاش شده تا از میان ویژگیهای استخراج شده، ویژگیهای موثرتری که حاوی اطلاعاتِ مربوط به گویندهٔ بیشتری هستند، انتخاب شوند. همچنین روشهایی نیز پیشنهاد شده‌اند که قادرند تا حذف تأثیرات نویز را در مرحلهٔ مدل‌کردن ویژگی‌ها انجام دهند[۱ و ۹].

امروزه یکی از رایج‌ترین روشهای مدل کردن گویندگان،