

تشخیص محدوده‌ی گفتاری در سیگنال صوتی مبتنی بر انتخاب مثبت و سیستم‌های ایمنی مصنوعی

رضا وفاشعار^۱، محمد مهدی همایون‌پور^۲

^۱ دانشجوی دکتری، دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، vafashoar@aut.ac.ir

^۲ دانشیار، دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات، دانشگاه صنعتی امیرکبیر، تهران، homayoun@aut.ac.ir

چکیده

در این مقاله روش جدیدی برای تشخیص محدوده‌ی گفتار در سیگنال صوتی بر اساس دسته‌بندی مبتنی بر انتخاب مثبت ارائه شده است. با توجه به مرزهای پیچیده‌ای که بین داده‌های مربوط به نویز و گفتار وجود دارد، استفاده از روش‌های دسته‌بندی معمول برای جداسازی آنها دشوار می‌باشد. در روش ارائه شده با استفاده از انتخاب مثبت گیرنده‌هایی برای تشخیص داده‌های گفتاری و داده‌های نویزی ایجاد می‌شود. روش دسته‌بندی ارائه شده شبیه یادگیری با نمونه می‌باشد و قادر به جداسازی کارآمد گفتار از نویز می‌باشد. در فاز آموزش روش ارائه شده ابتدا گیرنده‌هایی مبتنی بر انتخاب مثبت ایجاد می‌شوند. این گیرنده‌ها با استفاده از روش انتخاب کلونی بهبود می‌یابند تا بتوانند فضاهای گفتار و نویز را به طور مناسبی پوشش دهند. روش ارائه شده با چند روش متداول برای تشخیص محدوده‌ی کلمات، در محیط‌های مختلف با SNR متفاوت، مقایسه شده است. برای مقایسات از پیکره‌ی گفتاری TIMIT استفاده شده و عملکرد روش‌های مورد مقایسه بر اساس دو معیار نرخ تشخیص موفق و نرخ آلازم اشتباه بررسی شده است. نتایج مقایسات نشان می‌دهند که روش ارائه شده توانسته است با نرخ آلازم اشتباه پایین به نرخ تشخیص بالایی دست یابد به عنوان مثال در حضور نویز bobble، نرخ آلازم اشتباه بیشتر از ۰/۲ از روش‌های مقایسه شده بهتر بوده، در حالیکه نرخ تشخیص موفق مناسب و بیشتر از ۰/۹ می‌باشد. در حضور نویز factory نیز در بعضی موارد بهبودی در حدود ۰/۲ در نرخ آلازم اشتباه ایجاد شده است. روش پیشنهادی در حضور نویز سفید با نرخ آلازم اشتباه بسیار مناسب و پایین به نرخ تشخیص بالایی دست یافته که در SNR پایین ۰/۱۵ بهبود در نرخ تشخیص نیز مشاهده می‌شود.

کلمات کلیدی

تشخیص محدوده‌ی گفتاری سیگنال صوتی، دسته‌بندی مبتنی بر انتخاب مثبت، انتخاب کلونی.

روش‌های خوشه بندی [۶] و بسیاری روش‌های دیگر نیز تاکنون مطرح شده‌اند. در روش ارائه شده در [۴] ابتدا یک حدود تقریبی برای گفتار در سیگنال صوتی با استفاده از روشی مبتنی بر انرژی بدست می‌آید، سپس این حدود بدست آمده با استفاده از مدل‌های مخلوط گوسی که بر روی نویز و گفتار آموزش یافته‌اند بهبود می‌یابد. Vlahj و همکارانش روشی مبتنی بر وزن‌های غیرخطی اسپکترا برای تشخیص محدوده‌ی گفتار ارائه نموده‌اند که امکان تشخیص حروف صدا دار و شبه صدا دار را، که مشکل می‌باشد، افزایش می‌دهد [۷].

در این مقاله از دسته‌بندی مبتنی بر انتخاب مثبت [۸] برای جداسازی گفتار از نویز ضمیمه استفاده خواهد شد. در این روش ابتدا گیرنده‌هایی برای تشخیص داده‌های گفتاری ایجاد می‌شوند. این

۱- مقدمه

تشخیص محدوده‌ی گفتاری سیگنال صوتی برای جدا سازی گفتار موجود در سیگنال از نویز ضمیمه استفاده می‌شود، که تقریباً در اکثر کاربردهای مربوط به پردازش گفتار مانند تشخیص و کدگذاری گفتار کاربرد فراوان دارد. روش‌های فراوانی برای این کار تاکنون بررسی و مطرح شده‌اند روش‌های ساده‌ای مبتنی بر معیارهای انرژی، نرخ عبور از صفر و انرژی در فرکانس‌های پایین می‌توانند برای جداسازی گفتار از نویز ضمیمه مناسب باشند [۱ و ۲]. روش‌های پیچیده‌تری بر پایه‌ی مدل‌های مخفی مارکوف [۳]، مدل‌های مخلوط گوسی [۴ و ۵]،