استفاده از مدل‌های وابسته به محتوا در واژه‌یاب گفتار متمایزساز

شیما طبیبیان، احمد اکبری و بابک ناصرشریف

چکیده: رویکردهای واژه‌یابی گفتار به دو گروه تقسیم می‌شوند: رویکردهای مبتنی بر مدل مخفی مارکف و رویکردهای متمایزساز. یکی از فواید رویکردهای مبتنی بر مدل مخفی مارکف، قابلیت استفاده از اطلاعات وابسته به محتوا (سه واج) در جهت بهبود کارایی سیستم واژه‌یاب گفتار می‌باشد. از طرفی، عدم امکان استفاده از اطلاعات وابسته به محتوا یکی از معایب رویکردهای واژه‌یابی گفتار متمایزساز محسوب می‌شود. در این مقاله، راهکاری برای رفع این عیب ارائه شده که به این منظور، بخش استخراج ویژگی یک سیستم واژه‌یاب گفتار متمایزساز مبتنی بر الگوریتم تکاملی (EDSTD)[1]- که در کارهای قبلی ما ارائه شده است- به گونه‌ای تغییر یافته که اطلاعات وابسته به محتوا را در نظر بگیرد. در مرحله نخست، یک رویکرد استخراج ویژگی مستقل از محتوا پیشنهاد شده و سپس رویکردی برای به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی ارائه شده است. نتایج ارزیابی‌ها روی دادگان TIMIT حاکی از آن است که نرخ بازشناسی سیستم EDSTD وابسته به محتوا (CD-EDSTD) در اخطار اشتباه بر کلمه کلیدی بر ساعت بزرگ‌تر از دو، حدود ۳٪ از نرخ بازشناسی درست سیستم EDSTD مستقل از محتوا (CI-EDSTD) بالاتر است. هزینه این بهبود دقت، حدود ۰٫۳۶ افت سرعت پاسخ‌گویی است که قابل چشم‌پوشی می‌باشد.

***کلید واژه:* استخراج ویژگی، بازشناس واج، مستقل از محتوا، وابسته به محتوا، ماشین بردار پشتیبان، واژه‌یابی گفتار متمایزساز.**

۱- مقدمه

بازشناسی گفتار، گستره وسیعی از کاربردها از قبیل فهم دستورات بسیار ساده تا دریافت تمام اطلاعات مستتر در عبارت گفتار اداشده توسط گوینده را پوشش می‌دهد. در بسیاری از کاربردهای ارتباط میان انسان و رایانه مانند سیستم‌های اپراتور خودکار، دسته‌بندی نامه‌های صوتی، شنود مکالمات تلفنی با اهداف امنیتی و کمک به افراد ناتوان جسمی، نیازی به بازشناسی تمام عبارات و کلمات اداشده توسط گوینده نبوده و تنها بازشناسی دقیق تعدادی از کلمات مهم‌تر کفایت می‌کند. به آشکارسازی مجموعه‌ای از کلمات کلیدی در گفتار پیوسته ورودی، واژه‌یابی گفتار[2] اطلاق می‌شود.

رویکردهایی که برای آموزش سیستم‌های واژه‌یاب گفتار مورد استفاده قرار می‌گیرند به دو دسته تقسیم‌بندی می‌شوند: رویکردهای مبتنی بر مدل مخفی مارکف و رویکردهای متمایزساز[3]. مدل‌های مخفی مارکف، روش‌ها و ابزارهای آشنایی برای آموزش سیستم‌های واژه‌یاب گفتار هستند. رویکردهای واژه‌یابی گفتار مبتنی بر مدل مخفی مارکف به دو دسته تقسیم می‌شوند: واژه‌یابی گفتار مبتنی بر بازشناسی گفتار پیوسته با دادگان بزرگ[4] [۱] تا [۴] و واژه‌یابی گفتار مبتنی بر جستجوی واجی[5] [۵] تا [۹].

در رویکرد اول، ابتدا با استفاده از یک بازشناس گفتار مبتنی بر دادگان بزرگ، آرشیوهای صوتی بزرگ به شبکه‌های واجی یا کلمه‌ای تبدیل می‌شوند و این قسمت از کار به صورت برون‌خط[6] انجام می‌شود. سپس با به کارگیری تکنیک‌های جستجو در شبکه‌های واجی یا کلمه‌ای، مجموعه کلمات کلیدی ورودی به صورت برخط[7] مورد جستجو قرار می‌گیرند. در این رویکرد که واژه‌یابی گفتار مبتنی بر بازشناس گفتار با واژگان بزرگ نامیده می‌شود، اگرچه استفاده از مدل‌های زبانی بهبود دقت واژه‌یابی گفتار را در پی دارد، نیاز به وجود حجم زیاد داده آموزشی برچسب زده شده و پیچیدگی محاسباتی بالا از چالش‌های اساسی آن محسوب می‌شود. از دیگر چالش‌های مهم این رویکرد، وجود دادگان خارج از واژگان است. برای برخورد با این چالش رویکردهای مختلفی پیشنهاد شده‌اند [۱۰] و [۱۱].

در رویکرد دوم (واژه‌یابی گفتار مبتنی بر جستجوی واجی)، مدل کلمات کلیدی و مدل پرکننده[8] از زیرمدل‌های مستقل از محتوا[9] (در سطح واج) یا وابسته به محتوا[10] (در سطح دو واج یا سه واج) ساخته می‌شوند. در واقع مدل کلمه کلیدی از اتصال زیرمدل‌های متناظر با دنباله واجی کلمه کلیدی حاصل می‌شود. مدل پرکننده غالباً از اتصال کامل تمام زیرمدل‌های واجی حاصل می‌شود. واژه‌یابی گفتار مبتنی بر جستجوی واجی یک ضعف اساسی دارد و آن بالابودن خطاهای درج، حذف و جایگزینی در نتایج بازشناسی واج است. ضعف دیگر واژه‌یابی گفتار مبتنی بر جستجوی واجی آن است که مدل پرکننده قادر است هر دنباله واجی اعم از دنباله واجی متناظر با کلمات کلیدی را مدل کند. روش‌های مختلفی برای رفع این نقیصه در مقالات ارائه شده‌اند [۱۲] تا [۱۴].

از مزایای روش‌های واژه‌یابی گفتار مبتنی بر مدل مخفی مارکف می‌توان به سرعت مناسب آنها برای کاربردهای برخط، در نظر گرفتن دانش اولیه از دادگان آموزش، امکان استفاده از مدل‌های زبانی و امکان استفاده از اطلاعات وابسته به محتوا اشاره نمود. در مقابل این مزیت‌ها، واژه‌یابی گفتار مبتنی بر مدل‌های مخفی مارکف دارای چندین ضعف مهم می‌باشد. به دلیل استفاده از دانش اولیه از دادگان آموزش و در اختیار نبودن تخمین‌های دقیق در اغلب موارد، الگوریتم آموزش مدل‌های مخفی مارکف در بهینه‌های محلی گرفتار شده و دقت واژه‌یابی گفتار کاهش

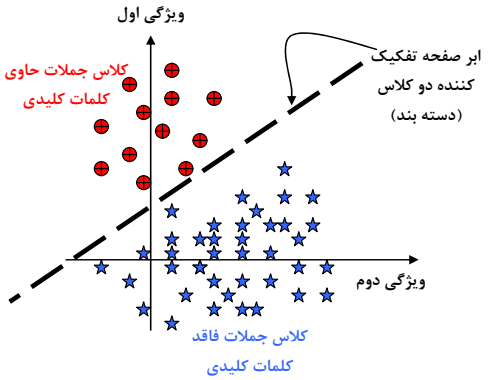
این مقاله در تاریخ ۱۹ دی ماه ۱۳۹۱ دریافت و در تاریخ ۲۹ شهریور ماه ۱۳۹۳ بازنگری شد.

شیما طبیبیان، پژوهشگاه هوافضا، وزارت علوم، تحقیقات و فناوری، تهران، (email: tabibian@ari.ac.ir).

احمد اکبری، آزمایشگاه پردازش صوت و گفتار، دانشکده کامپیوتر، دانشگاه علم و صنعت ایران، تهران، (email: akbari@iust.ac.ir).

بابک ناصرشریف، دانشکده مهندسی برق و کامپیوتر، دانشگاه صنعتی خواجه نصیرالدین طوسی، تهران، (email: bnasersharif@eetd.kntu.ac.ir).

1. Evolutionary Discriminative Spoken Term Detection
2. Keyword Spotting or Spoken Term Detection
3. Discriminative
4. Large Vocabulary Continuous Speech Recognition (LVCSR)-Based Keyword Spotting
5. Phonetic-Based Search Keyword Spotting
6. Offline
7. Online
8. Filler Model (Garbage Model)
9. Context-Independent Models
10. Context-Dependent Models



شکل ۱: واژه‌یاب گفتار به عنوان یک دسته‌بند دودویی.

می‌یابد. از دیگر نقاط ضعف مدل‌های مخفی مارکف آن است که الگوریتم آموزش مورد استفاده برای بهینه‌سازی پارامترهای مدل‌ها، بدون توجه به معیار ارزیابی کارایی سیستم‌های واژه‌یاب گفتار عمل می‌کند. بنابراین واژه‌یاب گفتار آموزش‌یافته مبتنی بر مدل مخفی مارکف، لزوماً معیار ارزیابی کارایی خود را بیشینه نمی‌کند.

روش‌های آموزش متمایزساز مختلفی برای برطرف نمودن معایب مدل‌های مخفی مارکف پیشنهاد شده‌اند که از آن جمله می‌توان به تخمین اطلاعات متقابل بیشینه (MMIE)[1] [۱۵]، خطای دسته‌بند کمینه (MCE)[2] [۱۶]، خطای واج کمینه (MPE)[3]، خطای کلمه کمینه[4] [۱۷] و [۱۷] و مدل‌های مخفی مارکف حاشیه- وسیع[5] [۱۸] و [۱۹] اشاره نمود. اگرچه این رویکردها شیوه آموزش مدل‌ها را بهبود داده‌اند، مشکل همگرایی به بهینه‌های محلی همچنان به قوت خود باقی است. همچنین معیار خطای مورد استفاده در این روش‌ها با معیار ارزیابی سیستم‌های واژه‌یاب گفتار منطبق نمی‌باشد. راهکار دیگر استفاده از مدل‌های متمایزساز به جای مدل‌های مخفی مارکف است. واژه‌یابی گفتار مبتنی بر مدل‌های متمایزساز در مقابل واژه‌یابی گفتار مبتنی بر مدل‌های مخفی مارکف، پیشنهاد شده است. واژه‌یابی گفتار مبتنی بر شبکه‌های عصبی [۲۰] و [۲۱] و رویکردهای حاشیه- وسیع [۲۲] تا [۲۷]، به ویژه ماشین بردار پشتیبان[6] [۲۸] از روش‌های معروف واژه‌یابی گفتار مبتنی بر مدل‌های متمایزساز محسوب می‌شوند.

از مزایای مدل‌های متمایزساز برای واژه‌یابی گفتار می‌توان به منطبق‌بودن الگوریتم آموزش مدل‌ها با معیار ارزیابی سیستم‌های واژه‌یاب گفتار و در نتیجه دقت و کارایی مطلوب واژه‌یاب گفتار، عدم نیاز به رویکرد مجزایی برای برخورد با چالش دادگان خارج از واژگان و عدم نیاز به دانش اولیه از دادگان آموزش اشاره نمود. در مقابل این مزایا، مدل‌نکردن مستقیم کشش زمانی واج‌ها، پیچیدگی‌های محاسباتی و زمانی، عدم امکان استفاده از اطلاعات وابسته به محتوا و عدم امکان استفاده از مدل‌های زبانی از نقاط ضعف مدل‌های متمایزساز محسوب می‌شود. به منظور مدل‌نمودن کشش زمانی واج‌ها از روش‌های دیگری در کنار رویکرد متمایزساز استفاده می‌شود [۲۵] و [۲۹] تا [۳۱]. پیچیدگی‌های محاسباتی و زمانی بالا، امکان استفاده از واژه‌یاب گفتار را به صورت برخط با مشکل مواجه می‌سازد. به دلیل وجود خطاهای اجتناب‌ناپذیر در بازشناسی واج که جزء لاینفک سیستم‌های واژه‌یاب گفتار است، لزوماً بخشی از عبارت گفتار که بیشترین میزان اطمینان از حضور کلمه کلیدی در آن وجود دارد، مکان درست رخداد کلمه کلیدی نمی‌باشد. این موضوع در سیستم‌های واژه‌یاب گفتاری که تشخیص مکان درست کلمه کلیدی از اهمیت بالایی برخوردار است ایجاد خلل می‌کند. استفاده از اطلاعات وابسته به محتوا، منجر به افزایش دقت بازشناسی واج و در نتیجه بهبود کارایی سیستم واژه‌یاب گفتار می‌گردد.

در این مقاله، واژه‌یاب گفتار متمایزساز به عنوان یک دسته‌بند[7] دودویی دودویی در نظر گرفته شده که کلاس جملات حاوی کلمات کلیدی را از کلاس جملات فاقد کلمات کلیدی تفکیک می‌کند. رویکرد واژه‌یابی گفتار متمایزساز پیشنهادی از دو بخش اصلی تشکیل شده است: استخراج ویژگی و دسته‌بندی. برای بخش دسته‌بندی، از یک الگوریتم تکاملی مبتنی بر ایده حاشیه- وسیع- که در کارهای قبلی ما ارائه شده است [۲۶]- برای آموزش پارامترهای دسته‌بند مذکور استفاده شده است. در این مقاله روی بخش استخراج ویژگی تمرکز کرده‌ایم. ابتدا یک رویکرد استخراج ویژگی مستقل از محتوا (مبتنی بر واج) ارائه شده و سپس در مرحله دوم، راهکار مناسبی برای استفاده از اطلاعات وابسته به محتوا (سه واج‌ها) به منظور افزایش دقت بازشناسی واج پیشنهاد داده‌ایم.

ساختار مقاله به صورت ذیل تدوین شده است. در بخش دوم مقاله چارچوب پایه برای واژه‌یابی گفتار متمایزساز مورد بررسی قرار خواهد گرفت. در بخش سوم، روش پیشنهادشده برای استخراج ویژگی مستقل از محتوا ارائه خواهد شد. نحوه به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی در بخش چهارم مورد بحث قرار می‌گیرد. در بخش پنجم، شرایط آزمایش و نتایج ارزیابی‌ها ارائه می‌شوند و نهایتاً در بخش ششم به جمع‌بندی مقاله می‌پردازیم.

۲- چارچوب پایه برای واژه‌یابی گفتار متمایزساز

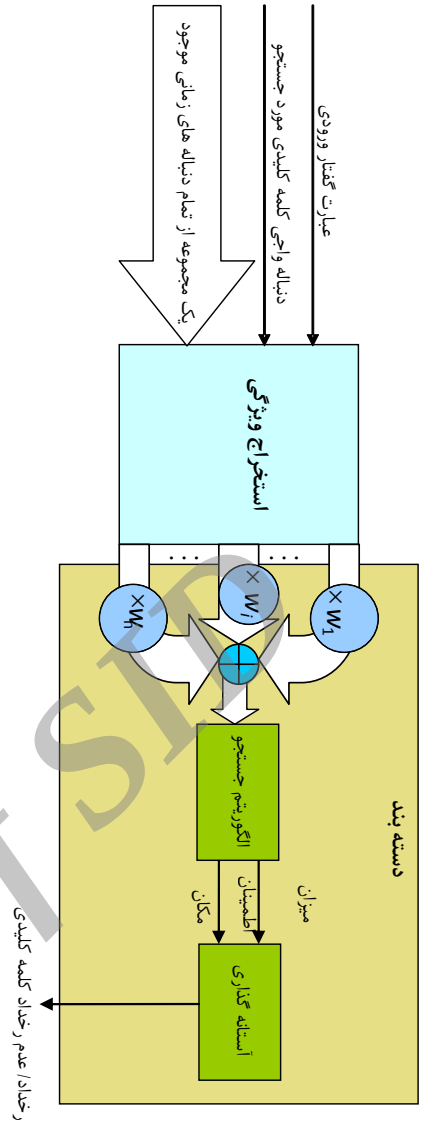
در این مقاله، یک واژه‌یاب گفتار به عنوان یک دسته‌بند دودویی در نظر گرفته شده که کلاس جملات حاوی کلمات کلیدی را از کلاس جملات فاقد کلمات کلیدی تفکیک می‌کند. هر دسته‌بند از دو بخش مهم تشکیل شده است: بخش استخراج ویژگی و بخش دسته‌بندی. با فرض آن که ویژگی‌های استخراج‌شده برای نمونه‌های متعلق به دو کلاس مذکور، در دو بعد قابل نمایش باشند، شکل ۱ نمونه‌ای از کلاس جملات حاوی کلمات کلیدی و کلاس جملات فاقد کلمات کلیدی را نشان می‌دهد.

چنانچه شکل ۱ نشان می‌دهد، نمونه‌های مشخص‌شده با ستاره، نماینده نمونه‌های کلاس جملات فاقد کلمات کلیدی و دایره‌ها، نماینده نمونه‌های کلاس جملات حاوی کلمات کلیدی می‌باشند. دسته‌بند تفکیک‌کننده دو کلاس ابرصفحه‌ای است که در قالب خط‌چین نمایش داده شده است. شکل ۱ در حالت ایده‌آل رسم شده و در حالت واقعی نمونه‌ها در هم رفتگی داشته و بدون خطا از هم تفکیک نمی‌شوند.

شکل ۲ چارچوب پایه واژه‌یاب گفتار متمایزساز را نشان می‌دهد. این چارچوب پایه در [۲۵] تا [۲۷] ارائه و به کار گرفته شده است.

چنانچه شکل ۲ نشان می‌دهد، سیستم واژه‌یاب گفتار متمایزساز از دو بخش استخراج ویژگی و دسته‌بند تشکیل شده است. به منظور آموزش پارامترهای دسته‌بند تفکیک‌کننده دو کلاس مذکور، از رویکرد مبتنی بر ماشین بردار پشتیبان [۲۸] استفاده شده است. بنابر ایده حاشیه- وسیع

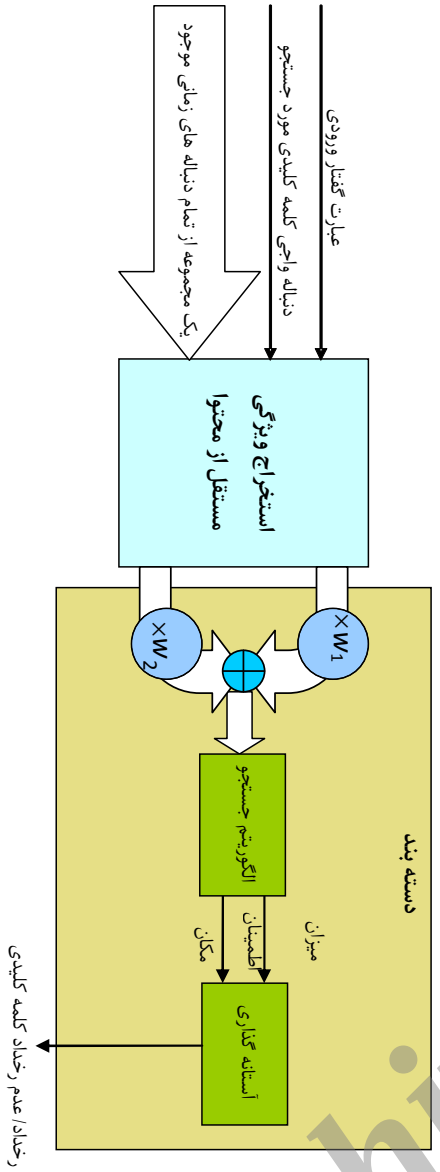
1. Maximum Mutual Information Estimation
2. Minimum Classification Error
3. Minimum Phone Error
4. Minimum Word Error
5. Large Margin
6. Support Vector Machine
7. Classifier



شکل ۲: چارچوب پایه واژه‌یاب گفتار متمایزساز.

موجود در رویکردهای مبتنی بر ماشین بردار پشتیبان، پارامترهای دسته‌بند (ابرصفحه) تفکیک‌کننده دو کلاس مذکور به گونه‌ای آموزش داده می‌شوند که فاصله میان نمونه‌های مرزی کلاس‌ها (بردارهای پشتیبان) و ابرصفحه تفکیک‌کننده دو کلاس بیشینه شود. با توجه به ادبیات ماشین بردار پشتیبان [۲۸] در انجام این بهینه‌سازی، یک یا چند معیار خطا مورد نظر گرفته شده و بهینه‌سازی با هدف کمینه‌نمودن خطای محاسبه‌شده توسط معیارهای مذکور انجام می‌گیرد. از آنجا که دسته‌بند آموزش‌یافته در کاربرد واژه‌یابی گفتار مورد استفاده قرار می‌گیرد، معیار خطای مورد استفاده در این کاربرد خاص، یکی از معیارهای ارزیابی سیستم‌های واژه‌یاب گفتار خواهد بود. n، تعداد ویژگی‌های تفکیک‌کننده استخراج‌شده در بخش استخراج ویژگی را نشان می‌دهد. w_1، w_2، ... و w_n، معرف پارامترهای ابرصفحه تفکیک‌کننده (دسته‌بند مبتنی بر ماشین بردار پشتیبان) کلاس جملات حاوی کلمات کلیدی و کلاس جملات فاقد کلمات کلیدی هستند که تعداد آنها با توجه به نوع دسته‌بند مورد استفاده، مساوی با ویژگی‌های استخراجی است.

در زیربخش الگوریتم جستجو از دسته‌بند شکل ۲، با بهره‌گیری از یک ایده متمایزساز، مکان رخداد آن را به گونه‌ای تعیین می‌کند که میزان اطمینان از رخداد کلمه کلیدی در آن مکان از عبارت گفتار ورودی بیشینه شود. در نهایت با آستانه‌گذاری میزان اطمینان به دست آمده، رخداد یا



شکل ۳: جایگاه روش پیشنهادی برای استخراج ویژگی مستقل از محتوا در چارچوب پایه واژه‌یاب گفتار متمایزساز.

عدم رخداد کلمه کلیدی در عبارت گفتار و به عبارت بهتر کلاس عبارت گفتار ورودی تعیین می‌شود.

این مقاله روی بخش استخراج ویژگی از سیستم پایه برای واژه‌یابی گفتار متمایزساز متمرکز شده است. نوآوری‌های این مقاله به صورت خلاصه در ادامه ارائه شده است:

الف) مدل‌نمودن دو مفهوم میزان اطمینان از رخداد کلمه کلیدی در عبارت گفتار ورودی و مکان رخداد آن در قالب دو تابع اندازه اطمینان به عنوان ویژگی‌های تفکیک‌کننده مورد استفاده در بخش استخراج ویژگی.

ب) ارائه راهکاری برای به کارگیری اطلاعات مربوط به توالی واج‌ها (اطلاعات وابسته به محتوا) با هدف افزایش دقت بازشناس واج مورد استفاده در بخش استخراج ویژگی.

۳- روش استخراج ویژگی پیشنهادی مستقل از محتوا

در این بخش از مقاله روی قسمت استخراج ویژگی از چارچوب پایه واژه‌یابی گفتار متمایزساز (شکل ۲) تمرکز نموده و رویکردی را برای استخراج ویژگی مستقل از محتوا ارائه می‌نماییم. شکل ۳ جایگاه روش

پیشنهادی برای استخراج ویژگی مستقل از محتوا را در چارچوب پایه واژه‌یاب گفتار متمایزساز نشان می‌دهد.

تفاوت شکل ۳ در مقایسه با شکل ۲ (چارچوب پایه واژه‌یاب گفتار متمایزساز)، تعیین تعداد دقیق ویژگی‌های تفکیک‌کننده (تعیین مقدار پارامتر n) می‌باشد. w_1 و w_2 در شکل ۳، معرف پارامترهای ابرصفحه تفکیک‌کننده (دسته‌بند) کلاس جملات حاوی کلمات کلیدی و کلاس جملات فاقد کلمات کلیدی هستند. به منظور آموزش پارامترهای دسته‌بند تفکیک‌کننده دو کلاس مذکور از یک الگوریتم تکاملی مبتنی بر ایده حاشیه- وسیع [۲۶] استفاده شده است.

برای واژه‌یابی گفتار متمایزساز، ویژگی‌های تفکیک‌کننده مختلفی در مقالات پیشنهاد شده‌اند [۲۲]، [۲۳]، [۲۵] تا [۲۷] و [۲۹] تا [۳۱]. این ویژگی‌های تفکیک‌کننده به گونه‌ای ارائه شده‌اند که دو مفهوم اصلی واژه‌یابی گفتار، میزان اطمینان از رخداد کلمه کلیدی در عبارت گفتار ورودی و مکان رخداد آن را مدل نمایند. به عنوان مثال ویژگی‌هایی که نرخ صحبت گوینده، کشش زمانی واج‌ها و گذار میان واج‌ها را مدل می‌کنند [۲۲]، [۲۳]، [۲۵] تا [۲۷] و [۲۹] تا [۳۱] برای تشخیص مکان درست کلمه کلیدی کاربرد داشته و ویژگی‌های بیانگر میزان اطمینان از حضور واج‌های تشکیل‌دهنده کلمه کلیدی در قاب‌های متناظر گفتار [۲۲]، [۲۳]، [۲۵] تا [۲۷] و [۲۹] تا [۳۱] برای تعیین میزان اطمینان از رخداد کلمه کلیدی در عبارت گفتار ورودی به کار می‌روند. در این مقاله به منظور مدل‌نمودن دو مفهوم اصلی واژه‌یابی گفتار، میزان اطمینان از رخداد کلمه کلیدی در عبارت گفتار ورودی و مکان رخداد آن، دو ویژگی تفکیک‌کننده پیشنهاد داده شده‌اند. ویژگی تفکیک‌کننده اول، میزان اطمینان از حضور دنباله واجی کلمه کلیدی مورد جستجو در عبارت گفتار ورودی را محاسبه می‌نماید و به آن تابع اطمینان از حضور[1] (PCF) گفته می‌شود. ویژگی تفکیک‌کننده دوم بیانگر میزان اطمینان از اعتبار کشش زمانی پیشنهادشده برای واج‌های تشکیل‌دهنده کلمه کلیدی بوده و به آن تابع اطمینان از کشش زمانی[2] (DCF) گفته می‌شود. جزئیات مربوط به این ویژگی‌های تفکیک‌کننده در دو زیربخش بعدی ارائه شده‌اند.

۳-۱ تابع اطمینان از حضور (PCF)

تابع اطمینان از حضور، میزان اطمینان دنباله واجی کلمه کلیدی k را در عبارت گفتار ورودی X با توجه به دنباله زمانی S محاسبه می‌کند. دنباله زمانی S، بیانگر دنباله زمان‌های پیش‌بینی شده برای رخداد واج‌های تشکیل‌دهنده کلمه کلیدی k می‌باشد. به منظور محاسبه میزان اطمینان از رخداد واج‌های کلمه کلیدی k در مکان پیشنهادشده توسط دنباله زمانی S، از یک بازشناس واج استفاده خواهد شد. توانمندی‌های مورد نیاز بازشناس واج مذکور، داشتن دقت مطلوب در بازشناسی واج، مقاوم‌بودن در شرایط نویزی و قابلیت تشخیص نواحی گذار میان واج‌ها می‌باشد. در [۲۲] و [۲۳] برای مدل‌نمودن محدوده‌های گذار میان واج‌ها از چهار تابع ویژگی مجزا استفاده شده که این چهار تابع ویژگی بر اساس فاصله اقلیدوسی میان ویژگی‌های MFCC استخراج‌شده از قاب‌های مجاور واج فعلی، محاسبه می‌شوند. روشن است که ویژگی‌های MFCC گزینه چندان مناسبی برای مدل‌نمودن محدوده‌های گذار میان واج‌ها نمی‌باشند و از طرف دیگر، وجود پنج تابع ویژگی برای محاسبه میزان اطمینان از حضور واج‌های کلمه کلیدی k در عبارت گفتار ورودی X و محدوده‌های گذار میان واج‌ها، پیچیدگی‌های محاسباتی و زمانی را در بخش استخراج ویژگی و در واژه‌یاب گفتار متمایزساز نهایی در پی دارد.

بر اساس نتایج پیاده‌سازی‌های ارائه‌شده در [۳۲] و [۳۳] استفاده از ویژگی‌های طیفی سیگنال گفتار منجر به افزایش صحت تشخیص محدوده‌های گذار میان واج‌ها می‌شود. در این مقاله به دنبال رابطه‌ای برای محاسبه PCF هستیم که نه تنها میزان اطمینان از حضور واج‌های کلمه کلیدی k در عبارت گفتار ورودی X را محاسبه کند، بلکه با بهره‌گرفتن از ویژگی‌های طیفی، محدوده‌های گذار میان واج‌ها را نیز بهتر مدل نماید. با ارائه این رابطه، پنج تابع ویژگی ارائه‌شده در [۲۲] و [۲۳] به تنها یک تابع ویژگی در این مقاله کاهش می‌یابد که به نوبه خود منجر به کاهش پیچیدگی‌های محاسباتی و زمانی در بخش استخراج ویژگی و در واژه‌یاب گفتار متمایزساز نهایی خواهد شد. رابطه (۱) نحوه محاسبه PCF را نشان می‌دهد

$$PCF(P^k, X, S) = \frac{1}{|P^k|} \sum_{i=1}^{|P^k|} \frac{1}{s_{i+1} - s_i} \sum_{t=s_i}^{s_{i+1}-1} f(p_i, x_t) \qquad (1)$$

که $s_{i+1} - s_i$ کشش زمانی واج p_i بر حسب تعداد قاب گفتار و s_i شماره قاب ابتدای واج p_i می‌باشد. $f(p_i, x_t)$ تابعی است که واج p_i و شماره قاب x_t را به عنوان آرگومان ورودی دریافت نموده و میزان اطمینان از انتساب واج p_i به قاب گفتار x_t را به عنوان خروجی بر می‌گرداند. P^k نماینده دنباله واج‌های کلمه کلیدی k می‌باشد. اگر چه در [۲۲] و [۲۳] میزان اطمینان از حضور دنباله واج‌های کلمه کلیدی k در عبارت گفتار X به کمک رابطه‌ای مشابه با (۱) محاسبه می‌شود، دو تفاوت مهم میان آنها وجود دارد. تفاوت اول مربوط به روش مورد استفاده برای محاسبه $f(p_i, x_t)$ است. در [۲۲] و [۲۳]، $f(p_i, x_t)$ با استفاده از یک بازشناس واج مبتنی بر ماشین بردار پشتیبان محاسبه می‌شود که در [۳۴] به طور مبسوط توضیح داده شده، اما در این مقاله، مقادیر $f(p_i, x_t)$ با استفاده از یک بازشناس واج مبتنی بر ماشین بردار پشتیبان که از الگوریتم آموزش SMO [۳۵] و [۳۶] استفاده می‌کند، انجام می‌شود. بنابراین، آموزش و آزمون ماشین بردار پشتیبان سریع‌تر انجام می‌شود. تفاوت دوم مربوط به ویژگی‌های ورودی به دسته‌بند مبتنی بر ماشین بردار پشتیبان می‌باشد. در [۲۲] و [۲۳] بردار ویژگی‌های ورودی، تنها شامل ویژگی‌های MFCC[3] مربوط به قاب‌های عبارت گفتار ورودی می‌باشد، در در حالی که بردار ویژگی‌های ورودی به دسته‌بند مبتنی بر ماشین بردار پشتیبان مورد استفاده در مقاله، علاوه بر ویژگی‌های MFCC، در برگیرنده ویژگی‌های طیفی قاب‌های گفتار می‌باشد. ویژگی‌های آکوستیکی از ۱۲ ضریب MFCC، انرژی و مشتقات اول و دوم که با استفاده از روش نرمال‌سازی CMVN[4] [۳۷] نرمال شده‌اند، تشکیل شده است. ویژگی‌های ویژگی‌های طیفی در برگیرنده آنتروپی طیفی[5] [۳۸]، همواری طیفی[6] [۳۹]، درجه از هم‌پاشی[7] [۴۰] و فرکانس نیمساز[8] [۴۰] می‌باشند.

۳-۲ تابع اطمینان از کشش زمانی (DCF)

تابع اندازه اطمینان DCF با دریافت دو پارامتر دنباله واجی کلمه

1. Presence Confidence Function
2. Duration Confidence Function
3. Mel Frequency Cepstral Coefficients
4. Cepstral Mean and Variance Normalization
5. Spectral Entropy
6. Spectral Flatness
7. Burst Degree
8. Bisector Frequency

کلیدی k و دنباله رخدادهای زمانی S، میزان اطمینان از کشش زمانی پیش‌بینی شده برای واج‌های تشکیل‌دهنده کلمه کلیدی k را تعیین می‌کند. در [۲۲] و [۲۳] برای محاسبه اعتبار کشش زمانی پیش‌بینی شده برای هر واج از یک توزیع گوسی به صورت زیر استفاده شده است

$$DCF(P^k,S)=\frac{\sum_{i=1}^{|P^k|}Gaussian(p_i,mean\_duration_{p_i},std\_duration_{p_i})}{|P^k|} \quad (۲)$$

که $mean\_duration_{p_i}$ معرف میانگین و $std\_duration_{p_i}$ معرف انحراف معیار کشش زمانی‌های مختلف واج p_i می‌باشد. تابع Gaussian نیز همان تابع توزیع گوسی است که از (۳) محاسبه می‌شود

$$Gaussian(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{\sigma^2}} \quad (۳)$$

ایراد وارد بر این روش آن است که توزیع کشش زمانی واج‌ها لزوماً از تابع گوسی پیروی نمی‌کند. به علاوه، تابع گوسی زمانی مقدار بیشینه خود را به دست می‌آورد که کشش زمانی واج متناظرش دقیقاً برابر با مقدار میانگین کشش زمانی برای آن واج باشد و با دورشدن مقدار کشش زمانی منتسب به واج از مقدار میانگین، مقدار تابع گوسی و به عبارت دیگر اعتبار کشش زمانی پیش‌بینی شده برای واج کاهش می‌یابد. حال آن که یک واج، در گویش‌ها و لهجه‌های مختلف و همچنین بر اساس سرعت صحبت گویش‌ور با کشش‌های زمانی مختلفی ادا می‌شود که همه آنها معتبر هستند. در این مقاله به منظور رفع این ایرادها از رویکرد جدیدی برای محاسبه تابع اندازه اطمینان DCF استفاده شده که به صورت زیر عمل می‌کند

$$DCF(P^k,S)=\frac{\sum_{i=1}^{|P^k|}\text{Histogram}(p_i,(s_{i+1}-s_i))}{|P^k|} \quad (۴)$$

که $\text{Histogram}(p_i,(s_{i+1}-s_i))$، احتمال اختصاص داده شده به مقدار $s_{i+1}-s_i$ در هیستوگرام حاصل‌شده برای واج p_i بوده و s_i معرف قاب شروع واج p_i می‌باشد. مقدار این احتمال، میزان اعتبار کشش زمانی پیش‌بینی شده برای واج p_i را نشان می‌دهد. هیستوگرام مذکور بر اساس کشش زمانی‌های مختلف هر واج در دادگان آموزش حاصل شده است.

از میان دو ویژگی تفکیک‌کننده معرفی‌شده در این بخش، ویژگی PCF که تعیین‌کننده میزان اطمینان از رخداد واج‌های کلمه کلیدی در عبارت گفتار ورودی می‌باشد، از اهمیت بالاتری برخوردار است [۴۱]. با توجه به نقش بازشناس واج مبتنی بر ماشین بردار پشتیبان در محاسبه PCF، دقت بازشناسی واج تأثیر بسزایی در میزان تفکیک‌کنندگی ویژگی PCF دارد. افزایش میزان تفکیک‌کنندگی ویژگی PCF منجر به بهبود کارایی واژه‌یاب گفتار متمایزساز خواهد شد. در بخش بعد راهکاری برای افزایش میزان تفکیک‌کنندگی ویژگی PCF است که از اطلاعات وابسته به محتوا بهره می‌گیرد.

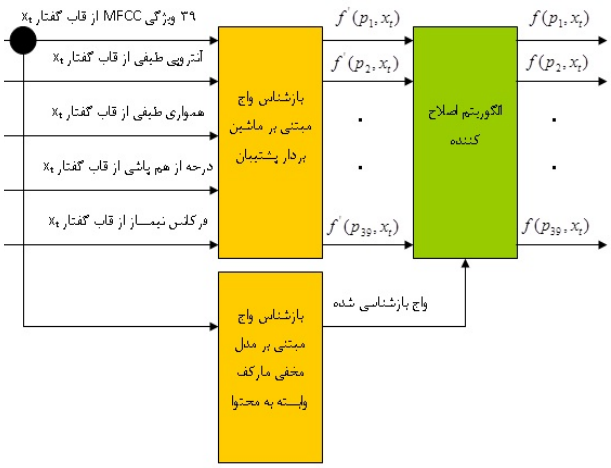
۴- روش پیشنهادی برای به کارگیری اطلاعات وابسته به محتوا در استخراج ویژگی

یکی از معایب بازشناس واج مبتنی بر ماشین بردار پشتیبان عدم به کارگیری اطلاعات مربوط به توالی میان واج‌ها در راستای بهبود دقت بازشناسی واج است. این در حالی است که در روش‌های مبتنی بر مدل مخفی مارکف امکان به کارگیری اطلاعات مربوط به توالی میان واج‌ها در قالب مدل‌های مخفی مارکف وابسته به محتوا (دو واج یا سه واج) فراهم می‌شود. در مقابل، ماشین بردار پشتیبان در مقایسه با مدل‌های مخفی مارکف، در شرایط نویزی از مقاومت بیشتری برخوردار است. از طرفی، بررسی‌های آزمایشگاهی انجام‌شده روی نتایج بازشناسی واج روی یک مجموعه دادگان یکسان، در هر دو رویکرد مبتنی بر ماشین بردار پشتیبان و مدل مخفی مارکف، حاکی از آن است که در برخی موارد نتایج بازشناسی مدل مخفی مارکف و در مواردی نیز نتایج بازشناسی ماشین بردار پشتیبان صحیح بوده است. بنابراین به دنبال راهکاری برای ترکیب نتایج بازشناسی دو رویکرد به منظور بهره‌گرفتن از مزایای هر دو رویکرد و جبران‌نمودن کاستی‌هایشان می‌باشیم.

روش‌های متعددی برای ترکیب مدل‌های مخفی مارکف و ماشین بردار پشتیبان ارائه شده که در دسته‌ای از آنها ماشین بردار پشتیبان برای بهبود کارایی رویکرد مبتنی بر مدل مخفی مارکف مورد استفاده قرار می‌گیرد [۲۴]، [۴۲] و [۴۳]. در دسته دیگری [۴۴] تا [۴۶] ابتدا با استفاده از ماشین بردار پشتیبان برخی از پارامترهای مدل‌های مخفی مارکف تخمین زده شده، سپس با استفاده از مدل‌های مخفی مارکف تنظیم‌شده بر اساس پارامترهای تخمین زده شده مذکور، بازشناسی واج انجام می‌شود. بر خلاف کارهای انجام‌شده در جهت ترکیب رویکردهای مبتنی بر مدل مخفی مارکف و ماشین بردار پشتیبان، هدف از این مقاله ترکیب دو بازشناس واج و به دست آوردن یک بازشناس واج با کارایی بیشتر نیست، بلکه هدف آن است که از نتایج بازشناسی واج مبتنی بر مدل‌های مخفی مارکف وابسته به محتوا برای اصلاح نتایج بازشناسی واج مبتنی بر ماشین بردار پشتیبان استفاده شود که به این ترتیب اطلاعات وابسته به محتوا در ماشین بردار پشتیبان نیز در نظر گرفته شود.

الگوریتم اصلاح باید به صورتی عمل نماید که در مواردی که بازشناس واج مبتنی بر ماشین بردار پشتیبان به درستی عمل می‌کند، نتایج تغییر نکنند و در مواردی که بازشناس واج مبتنی بر ماشین بردار پشتیبان عملکرد نادرستی دارد، نتایج بازشناس واج مبتنی بر مدل مخفی مارکف جایگزین نتایج نادرست شوند. روشن است که تشخیص موارد مذکور بدون دسترسی به دنباله واجی صحیح متناظر با عبارت گفتار ورودی امکان‌پذیر نمی‌باشد. در ادامه راهکاری ارائه شده که بدون نیاز به دانستن دنباله واجی عبارت گفتار ورودی، دستیابی به شرایط الگوریتم اصلاح را تا حد قابل قبولی امکان‌پذیر کرده است. شکل ۴ بلوک دیاگرام مربوط به روند به کارگیری اطلاعات وابسته به محتوا را در جهت اصلاح نتایج بازشناسی واج مبتنی بر ماشین بردار پشتیبان نشان می‌دهد.

چنانچه شکل ۴ نشان می‌دهد، در مرحله اول با استفاده از یک بازشناس واج مبتنی بر ماشین بردار پشتیبان، ویژگی‌های آکوستیکی و طیفی یک قاب از عبارت گفتار ورودی به ۳۹ عدد اعشاری (از $f'(p_1,x_t)$ تا $f'(p_{39},x_t)$) نظیر می‌شود. عدد اعشاری iام $(f'(p_1,x_t))$ معرف میزان اطمینان از حضور واج p_i در قاب ورودی مذکور است. بیشترین مقدار در میان این ۳۹ عدد اعشاری، واج منتسب به قاب گفتار ورودی را معین می‌نماید. در مرحله دوم با استفاده از خروجی بازشناس واج مبتنی بر مدل مخفی مارکف وابسته به محتوا- که در ازای قاب گفتار ورودی x_t، واج منتسب به این قاب گفتار است- و با استفاده از یک الگوریتم اصلاح‌کننده، نتایج بازشناس واج مبتنی بر ماشین بردار پشتیبان اصلاح می‌شوند. الگوریتم اصلاح‌کننده به گونه‌ای عمل می‌کند که چنانچه واج خروجی بازشناس واج مبتنی بر مدل مخفی مارکف با واج متناظر با



شکل ۴: به کارگیری اطلاعات وابسته به محتوا در جهت اصلاح نتایج بازشناسی واج مبتنی بر ماشین بردار پشتیبان.

بیشترین مقدار خروجی بازشناس واج مبتنی بر ماشین بردار پشتیبان، یکسان بود مقادیر $f'(p_1,x_t)$ تا $f'(p_{39},x_t)$ بدون تغییر به مقادیر $f(p_1,x_t)$ تا $f(p_{39},x_t)$ نگاشت می‌شوند. در غیر این صورت با در نظر گرفتن دو مقدار پاداش و جریمه و با فرض آن که خروجی $f'(p_i,x_t)$ بازشناس واج مبتنی بر ماشین بردار پشتیبان بیشترین مقدار را داشته است، مقدار $f'(p_i,x_t)$ با مقدار جریمه (که یک عدد منفی است) جمع شده و به $f(p_i,x_t)$ نگاشت می‌شود. همچنین با فرض آن که خروجی بازشناس واج مبتنی بر مدل مخفی مارکف، p_j باشد مقدار $f'(p_j,x_t)$ با مقدار پاداش (که یک عدد مثبت است) جمع شده و به $f(p_j,x_t)$ نگاشت می‌شود. سایر مقادیر خروجی بازشناس واج مبتنی بر ماشین بردار پشتیبان تغییر نمی‌کنند. شکل ۵ روندنمای مربوط به الگوریتم اصلاح‌کننده را نشان می‌دهد.

در ادامه اثبات صحت عملکرد الگوریتم اصلاح‌کننده، ارائه شده است.

اثبات صحت عملکرد الگوریتم اصلاح

چنانچه مطرح شد، میزان اطمینان از حضور کلمه کلیدی k در عبارت گفتار ورودی x با استفاده از (۱) محاسبه می‌شود که $f(p_i,x_t)$ میزان اطمینان از حضور واج p_i در قاب گفتار x_t می‌باشد. با توجه به الگوریتم اصلاح‌کننده و روندنمای ارائه‌شده در شکل ۵ دو حالت وجود دارد:

حالت اول

نتایج بازشناسی واج در دو رویکرد مبتنی بر ماشین بردار پشتیبان و مبتنی بر مدل مخفی مارکف یکسان است و در این حالت نتایج بازشناسی واج بدون تغییر باقی می‌مانند. چنانچه هر دو رویکرد درست عمل کرده باشند، واج p_i به درستی به قاب گفتار x_t نظیر شده و چنانچه هر دو رویکرد نادرست عمل کرده باشند، امکان اصلاح و بهبود نتایج از طریق ترکیب نتایج دو رویکرد وجود نداشته و بهبودی در کارایی واژه‌یاب گفتار نهایی حاصل نمی‌شود.

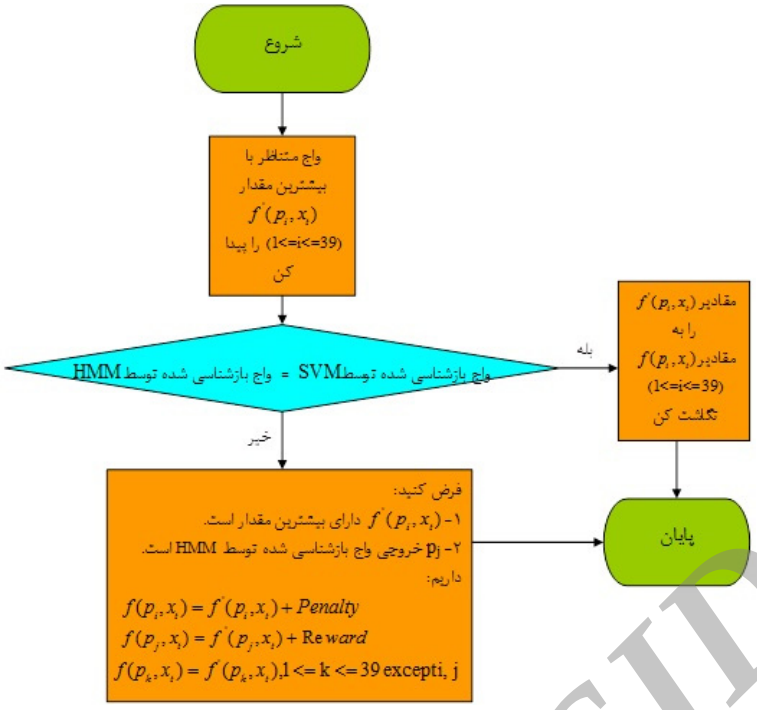
حالت دوم

نتایج بازشناسی واج در دو رویکرد مبتنی بر ماشین بردار پشتیبان و مبتنی بر مدل مخفی مارکف متفاوت است. در این حالت اصلاح به صورتی که قبلاً توضیح داده شد، انجام می‌شود. قابل ذکر است مقادیر پاداش و جریمه به گونه‌ای لحاظ می‌شوند که هر دو مقدار $f(p_i,x_t)$ و $f(p_j,x_t)$ تفاوت قابل توجهی نسبت به سایر ۳۷ مقدار خروجی بازشناس واج مبتنی بر ماشین بردار پشتیبان داشته باشند. به عبارتی متناسب با مقادیر خروجی بازشناس واج مبتنی بر ماشین بردار پشتیبان[۱] (محاسبه‌شده روی فایل‌های گفتاری مجموعه اعتبارسنجی) جریمه کوچک (۰/۱- در کار حاضر) و پاداش بزرگ (۰/۵ در کار حاضر) انتخاب می‌شود. لازم به ذکر است که بررسی‌های انجام‌شده روی مقادیر خروجی بازشناس مبتنی بر ماشین بردار پشتیبان (محاسبه‌شده روی فایل‌های گفتاری مجموعه اعتبارسنجی) بیانگر آن است که اگر نتیجه بازشناسی رویکرد مبتنی بر ماشین بردار پشتیبان درست باشد، مقدار عددی اختصاص داده شده به واج بازشناسی‌شده ($f(p_i,x_t)$) بزرگ‌تر از ۰/۸ و بسیار نزدیک به ۱ می‌باشد و در غیر این صورت، این مقدار اندکی بزرگ‌تر از ۰/۵ خواهد بود. بنابراین چنانچه نتیجه بازشناسی رویکرد مبتنی بر ماشین بردار پشتیبان صحیح و نتیجه بازشناسی رویکرد مبتنی بر مدل مخفی مارکف نادرست باشد، با توجه به مقدار کوچک جریمه و مقدار عددی خروجی ماشین بردار پشتیبان (بزرگ‌تر از ۰/۸ و نزدیک به ۱)، واج p_i با اندازه اطمینان قابل توجهی به قاب گفتار x_t نظیر می‌شود. در این حالت از آنجا که $f(p_i,x_t)$ بزرگ‌تر از ۰/۸ و نزدیک به ۱ است، پس از اعمال جریمه، مقدار آن بین ۰/۷ و ۱ خواهد بود. از سوی دیگر $f(p_j,x_t)$ پس از اعمال پاداش اندکی بزرگ‌تر از ۰/۵ و قطعاً کوچک‌تر از ۰/۷ خواهد شد و بنابراین واج p_i با اندازه اطمینان قابل توجهی به قاب گفتار x_t نظیر می‌شود. چنانچه نتیجه بازشناسی رویکرد مبتنی بر ماشین بردار پشتیبان نادرست و نتیجه بازشناسی رویکرد مبتنی بر مدل مخفی مارکف صحیح باشد، با توجه به مقدار بزرگ پاداش و مقدار کوچک خروجی ماشین بردار پشتیبان (اندکی بزرگ‌تر از ۰/۵)، واج p_j با اندازه اطمینان قابل توجهی به قاب گفتار x_t نظیر می‌شود. در این حالت از آنجا که $f(p_i,x_t)$ اندکی بزرگ‌تر از ۰/۵ است، پس از اعمال جریمه، مقدار آن کاهش یافته و کوچک‌تر از ۰/۵ خواهد شد. از سوی دیگر $f(p_j,x_t)$ پس از اعمال پاداش بزرگ‌تر از ۰/۵ خواهد شد. بنابراین در این حالت واج p_j با اندازه اطمینان قابل توجهی به قاب گفتار x_t نظیر می‌شود. چنانچه هر دو رویکرد نادرست عمل کرده باشند، امکان اصلاح و بهبود نتایج از طریق ترکیب نتایج دو رویکرد وجود نداشته، بهبودی در کارایی واژه‌یاب گفتار نهایی حاصل نمی‌شود.

با توجه به دو حالت توضیح داده شده، روشن است که الگوریتم اصلاح‌کننده به گونه‌ای عمل می‌کند که در حالتی که یکی از دو رویکرد درست عمل کنند، نتیجه درست به عنوان نتیجه بازشناسی واج نهایی لحاظ شود. لازم به تأکید است که به دلیل در دسترس‌نبودن برچسب واج‌های عبارات گفتار در مرحله آزمون، الگوریتم اصلاح‌کننده نمی‌داند که کدام یک از دو رویکرد درست عمل کرده‌اند. لیکن با تمهیداتی که برای اصلاح نتایج بازشناسی واج مبتنی بر ماشین بردار پشتیبان لحاظ کرده‌ایم، به صورت غیر مستقیم در حالتی که یکی از دو رویکرد درست عمل کنند، نتیجه درست به عنوان نتیجه بازشناسی واج نهایی لحاظ می‌شود.

در این بخش راهکاری برای به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی از واژه‌یاب گفتار متمایزساز ارائه شد. شکل ۶ جایگاه راهکار ارائه‌شده برای به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی را در واژه‌یاب گفتار متمایزساز ارائه‌شده در شکل ۳ نشان می‌دهد.

۱. ۹۵ درصد از مقادیر عددی کوچک‌تر از ۰/۱ (نزدیک صفر)، ۲ درصد از آنها بین ۰/۱ و ۰/۲ و ۱/۶ درصد از آنها بین ۰/۲ و ۰/۵ هستند. از ۱/۴ درصد باقیمانده مقادیر عددی، حدود ۶۰ درصد بزرگ‌تر از ۰/۸ (نزدیک ۱) و ۴۰ درصد بین ۰/۵ و ۰/۸ هستند.



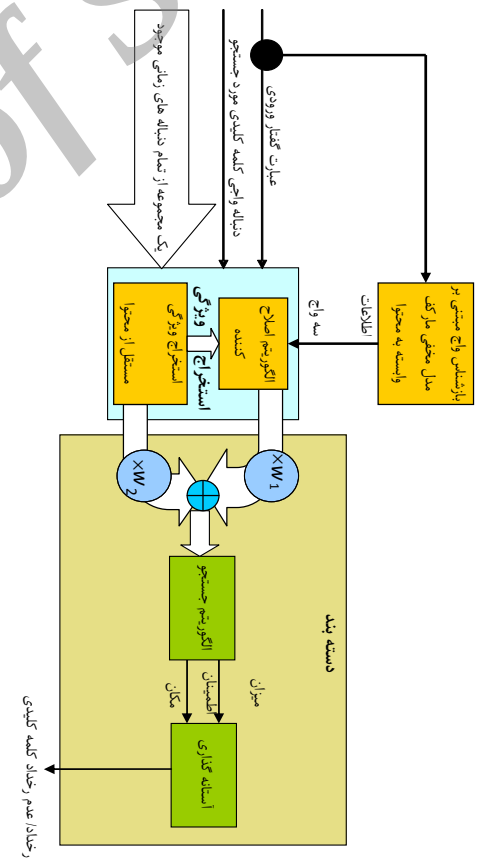
شکل ۵: روندنمای مربوط به الگوریتم اصلاح کننده.

۵- نتایج ارزیابی‌ها

چارچوب نهایی پیشنهادشده برای واژه‌یابی گفتار متمایزساز (در برگیرنده بخش استخراج ویژگی وابسته به محتوا) در شکل ۶ ارائه شد. در این بخش، نتایج ارزیابی چارچوب پیشنهادی برای واژه‌یابی گفتار متمایزساز ارائه شده است. به منظور بررسی دقیق‌تر و تحلیل جامع‌تر رویکرد پیشنهادشده، نتایج ارزیابی‌ها در سه زیربخش مجزا مورد بررسی قرار گرفته‌اند: چارچوب پایه واژه‌یابی گفتار (شکل ۲)، چارچوب پیشنهادشده برای واژه‌یابی گفتار متمایزساز به همراه بخش استخراج ویژگی مستقل از محتوا (شکل ۳) و چارچوب پیشنهادشده برای واژه‌یابی گفتار متمایزساز به همراه بخش استخراج ویژگی وابسته به محتوا (چارچوب نهایی پیشنهادشده برای واژه‌یابی گفتار متمایزساز) (شکل ۶). در راستای تکمیل ارزیابی‌ها، چارچوب نهایی پیشنهادشده برای واژه‌یابی گفتار متمایزساز با یک واژه‌یاب گفتار مبتنی بر مدل مخفی مارکف مقایسه شده است. آزمایشات روی پایگاه داده TIMIT [۴۷] انجام و شرایط آزمایش‌ها در جدول ۱ و روش‌های پیاده‌سازی شده در جدول ۲ ارائه شده‌اند.

برای ارزیابی روش‌های پیاده‌سازی شده، سه معیار ارزیابی مورد استفاده قرار گرفته‌اند: ROC[1]، FOM[2] و RTF[3]. یکی از رایج‌ترین معیارهای ارزیابی سیستم‌های واژه‌یاب گفتار، نمودار ROC می‌باشد. این معیار در قالب یک منحنی دوبعدی نمایش داده می‌شود که محور افقی آن نماینده تعداد اخطارهای اشتباه بر کلمه کلیدی بر ساعت و محور عمودی آن نماینده تعداد تشخیص‌های درست است [۴۸]. FOM، متوسط سطح زیر منحنی ROC برای نرخ اخطار اشتباه مابین ۱ تا ۱۰ به ازای هر کلمه کلیدی و به ازای یک ساعت داده صوتی ضبط‌شده است که به صورت زیر محاسبه می‌شود

1. Receiver Operating Characteristic
2. Figure of Merit
3. Real Time Factor



شکل ۶: جایگاه راهکار ارائه‌شده برای به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی، در واژه‌یاب گفتار متمایزساز ارائه‌شده در شکل ۳.

تفاوت شکل ۶ در مقایسه با شکل ۳، به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی به هدف بهبود دقت بازشناس واج مبتنی بر ماشین بردار پشتیبان و در نتیجه بهبود ویژگی PCF می‌باشد. در بخش بعد به ارزیابی سیستم پیشنهادشده برای واژه‌یابی گفتار متمایزساز (شکل ۶) خواهیم پرداخت.

جدول ۱: شرایط آزمایش.

رویکرد متمایزساز	مجموعه آموزش دسته‌بند مبتنی بر SVM برای استخراج PCF	۱۵۰۰ جمله از دادگان آموزش TIMIT
	مجموعه آموزش دسته‌بند متمایزساز	۸۱۴ جمله از دادگان آموزش TIMIT (۴۵۲ جمله برای آموزش و ۳۶۲ جمله برای اعتبارسنجی)
	دادگان آزمون واژه‌یاب گفتار متمایزساز	حداکثر ۲۰ جمله مثبت و ۲۰ جمله منفی برای هر یک از ۸۰ واژه کلیدی از دادگان آزمون TIMIT
	کمینه طول هر واژه کلیدی	۵ واج
	بردار ویژگی استخراج‌شده از هر قاب گفتار	۳۹ ضریب MFCC هنجارسازی شده با روش CMVN آنتروپی طیفی همواری طیفی درجه از هم پاشی فرکانس نیمساز
رویکرد مبتنی بر مدل مخفی مارکف	مجموعه آموزش بازشناس واج مبتنی بر مدل مخفی مارکف وابسته به محتوا برای بهبود PCF	۱۵۰۰ جمله از دادگان آموزش TIMIT (یکسان با مجموعه آموزش دسته‌بند مبتنی بر SVM)
	مجموعه آموزش واژه‌یاب مبتنی بر مدل مخفی مارکف	۳۶۹۶ جمله از دادگان آموزش TIMIT
	دادگان آزمون واژه‌یاب گفتار مبتنی بر مدل مخفی مارکف	حداکثر ۲۰ جمله مثبت برای هر یک از ۸۰ واژه کلیدی از دادگان آزمون TIMIT (یکسان با مجموعه آزمون واژه‌یاب گفتار متمایزساز)
	بردار ویژگی استخراج‌شده از هر قاب گفتار	۳۹ ضریب MFCC هنجارسازی شده با روش CMVN
	تعداد وضعیت‌ها در هر مدل مخفی مارکف	۳ وضعیت اصلی و ۲ وضعیت برای ورود و خروج
	تعداد مخلوط‌های گوسی در هر وضعیت	۱۶
	مدل پرکننده	بازشناس واج مبتنی بر یک مدل مخفی مارکف چپ به راست سه‌حالته با ۱۶ مخلوط گوسی در هر حالت [۹]
	روش محاسبه اندازه اطمینان برای دنباله واجی شناسایی‌شده توسط مدل پرکننده	تابع هارمونیک نرمال [۲۴]

جدول ۲: روش‌های پیاده‌سازی شده.

نام روش	شرح روش	
CI-DSTD	واژه‌یاب گفتار متمایزساز با ۷ ویژگی تفکیک‌کننده و رویکرد مبتنی بر ایده حاشیه- وسیع برای آموزش دسته‌بند [۳۵] به عنوان چارچوب پایه واژه‌یاب گفتار متمایزساز (شکل ۲)	رویکرد متمایزساز
CI-EDSTD	چارچوب پیشنهادشده برای واژه‌یابی گفتار متمایزساز به همراه بخش استخراج ویژگی مستقل از محتوا و رویکرد مبتنی بر الگوریتم تکاملی برای آموزش دسته‌بند [۲۶] (شکل ۳)	
CD-EDSTD	چارچوب پیشنهادشده برای واژه‌یابی گفتار متمایزساز به همراه بخش استخراج ویژگی وابسته به محتوا و رویکرد مبتنی بر الگوریتم تکاملی برای آموزش دسته‌بند و جستجوی اول- بهترین (شکل ۶)	
CD-HMM	واژه‌یاب گفتار مبتنی بر مدل مخفی مارکف وابسته به محتوا بدون مدل پرکننده	رویکرد مبتنی بر مدل مخفی مارکف
CD-HMM-CD-FM	واژه‌یاب گفتار مبتنی بر مدل مخفی مارکف وابسته به محتوا با مدل پرکننده وابسته به محتوا	

$$FOM = \frac{1}{10}\sum_{FA/KW/H=1}^{10} TDR_{FA/KW/H} \quad (۵)$$

به طوری که TDR نرخ تشخیص درست، FA نرخ اخطار اشتباه سیستم واژه‌یاب گفتار، KW تعداد کلمات کلیدی مورد جستجو و H طول زمانی فایل آزمون بر حسب ساعت می‌باشد.

RTF معیار رایج محاسبه سرعت یک سیستم بازشناسی گفتار خودکار می‌باشد. اگر P واحد زمانی برای پردازش یک ورودی از طول زمانی I نیاز داشته باشیم، RTF به صورت زیر محاسبه خواهد شد

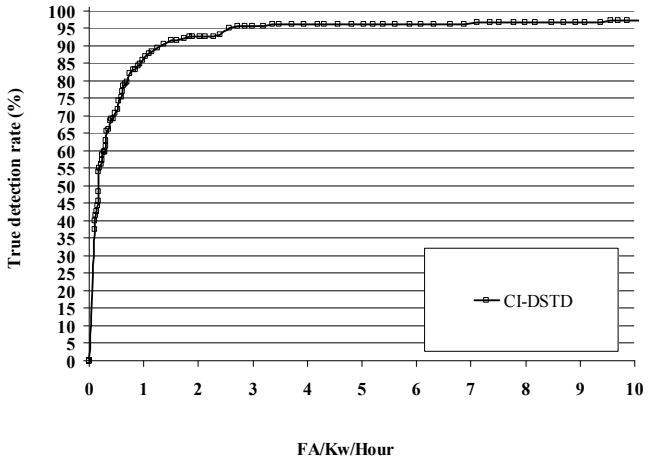
$$RTF = \frac{P}{I} \quad (۶)$$

نتایج ارزیابی‌ها در سه زیربخش مجزا که در ادامه ارائه شده‌اند، مورد تحلیل و بررسی قرار خواهند گرفت.

۵-۱ ارزیابی چارچوب پایه برای واژه‌یابی گفتار متمایزساز

به منظور ارزیابی چارچوب پایه واژه‌یاب گفتار متمایزساز (شکل ۲) نیازمند تعیین مقدار مشخصی برای n و روش آموزش پارامترهای دسته‌بند تفکیک‌کننده کلاس جملات حاوی کلمات کلیدی از کلاس جملات فاقد آنها هستیم. نخستین بار در سال ۲۰۰۹ میلادی، جوزف کشت با تعریف ۷ تابع ویژگی متمایزساز و ارائه یک روش مبتنی بر ایده حاشیه- وسیع به منظور آموزش پارامترهای دسته‌بند تفکیک‌کننده دو کلاس مذکور، چارچوبی را برای واژه‌یابی گفتار متمایزساز ارائه نمود [۲۲]. در این مقاله چارچوب ارائه‌شده در [۲۲] را به عنوان چارچوب پایه واژه‌یاب گفتار متمایزساز لحاظ نموده و آن را CI-DSTD[1] می‌نامیم. شکل ۷ نمودار ROC را برای ارزیابی روش CI-DSTD نشان می‌دهد.

1. Context Independent - EDSTD (CI - EDSTD)



شکل ۷: نمودار ROC برای ارزیابی چارچوب پایه واژه‌یاب گفتار متمایزساز (شکل ۲).

جدول ۳: ارزیابی چارچوب پایه واژه‌یاب گفتار متمایزساز (شکل ۲) بر اساس معیارهای FOM و RTF.

	FOM (%)	RTF
CI-DSTD	۹۵	۸٫۲

جدول ۴: ارزیابی دو روش CI-DSTD و CI-EDSTD بر اساس معیارهای FOM و RTF.

	FOM (%)	RTF
CI-DSTD	۹۵	۸٫۲
CI-EDSTD	۹۲٫۴	۳٫۶۵

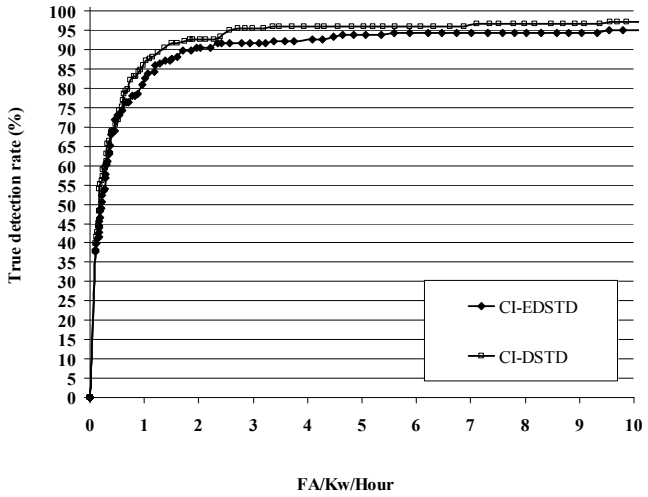
چنانچه شکل ۷ نشان می‌دهد، مقدار *TDR* در *FA* برابر با *FA/KW/H* ۷ با مقداری برابر با ۹۷٪ به اشباع رسیده است. قابل ذکر است که برای محاسبه *TDR* و *FA*، درست‌بودن مکان پیشنهادشده برای رخداد کلمه کلیدی نیز لحاظ شده و این در حالی است که ارزیابی روش CI-DSTD در [۲۲] بدون توجه به مکان پیشنهادشده برای رخداد کلمه کلیدی انجام شده است. تفاوت میان نتایج ارزیابی ارائه‌شده در این مقاله و [۲۲] به همین دلیل می‌باشد. مقادیر FOM و RTF مربوط به روش CI-DSTD در جدول ۳ ارائه شده‌اند.

چنانچه جدول ۳ نشان می‌دهد، در مقابل مقدار FOM مطلوبی که برای روش CI-DSTD حاصل شده، سرعت آن حدود ۸ برابر کمتر از حالت بلادرنگ می‌باشد. بخشی از این سرعت کم به پیچیدگی محاسباتی ناشی از تعداد زیاد ویژگی‌های متمایزساز و پارامترهای دسته‌بند تفکیک‌کننده دو کلاس مربوط می‌شود. در زیربخش بعدی روش CI-EDSTD (چارچوب پیشنهادشده در شکل ۳) در مقابل روش CI-DSTD مورد ارزیابی و مقایسه قرار گرفته است.

۵-۲ ارزیابی چارچوب پیشنهادشده در شکل ۳ (روش CI-EDSTD)

شکل ۸ نمودار ROC مربوط به ارزیابی و مقایسه روش CI-EDSTD را در مقابل روش CI-DSTD که در این مقاله به عنوان چارچوب پایه واژه‌یاب گفتار متمایزساز در نظر گرفته شده است، نشان می‌دهد.

چنانچه شکل ۸ نشان می‌دهد، روش CI-DSTD در مقایسه با روش CI-EDSTD در نرخ اخطار اشتباه یکسان از دقت تشخیص بالاتری برخوردار است. در بهترین حالت، دقت تشخیص روش CI-DSTD حدود ۵٪ بالاتر از دقت تشخیص روش CI-EDSTD می‌باشد اما با توجه به



شکل ۸: نمودار ROC برای ارزیابی دو روش CI-DSTD و CI-EDSTD.

آن که در روش CI-DSTD از ۷ ویژگی تفکیک‌کننده و در روش CI-EDSTD تنها از دو ویژگی تفکیک‌کننده در مراحل آموزش و آزمون واژه‌یاب گفتار استفاده شده، انتظار داریم که پیچیدگی زمانی و محاسباتی روش CI-EDSTD در دو مرحله آموزش و آزمون به مراتب پایین‌تر از روش CI-DSTD باشد که ارزیابی RTF می‌تواند تأییدکننده این حدس باشد. نتایج ارزیابی دو روش مذکور بر اساس معیارهای FOM و RTF در جدول ۴ ارائه شده‌اند.

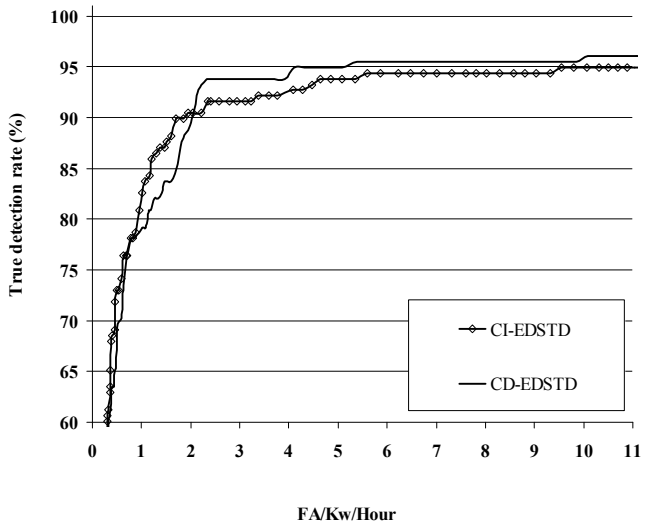
چنانچه جدول ۴ نشان می‌دهد، اگرچه مقدار FOM روش CI-DSTD حدود ۲٫۶٪ بالاتر از مقدار FOM روش CI-EDSTD می‌باشد، اما سرعت روش CI-EDSTD حدود ۲٫۲ برابر بیشتر از روش CI-DSTD است. بنابراین روش CI-EDSTD در مقابل پیچیدگی محاسباتی کمتر در دو مرحله آموزش و آزمون و ۲٫۲ برابر سرعت بیشتر، تنها ۲٫۶٪ افت دقت دارد. در زیربخش بعدی، میزان بهبود کارایی واژه‌یاب گفتار متمایزساز ناشی از به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی مورد ارزیابی قرار خواهد گرفت.

۵-۳ ارزیابی چارچوب پیشنهادشده در شکل ۶ (روش CD-EDSTD)

شکل ۹ نمودار ROC روش CI-EDSTD را در مقایسه با روش CD-EDSTD[1] نشان می‌دهد. روش CD-EDSTD نسبت به رویکرد CI-EDSTD، در نرخ اخطارهای اشتباه بالاتر از ۲ FA/Kw/Hour در بهترین حالت حدود ۳٪ بهبود دقت داشته است. قابل توجه است که استفاده از اطلاعات وابسته به محتوا منجر به افزایش دقت تشخیص واژه‌یاب گفتار متمایزساز شده است. واضح است که افزایش نرخ تشخیص بازشناس واج به دلیل به کارگیری اطلاعات وابسته به محتوا، در تشخیص دقیق‌تر نواحی گذار واج‌ها و در نهایت تشخیص دقیق‌تر مکان رخداد واج‌های تشکیل‌دهنده کلمات کلیدی تأثیر بسزایی دارد. نتایج ارزیابی این دو روش بر اساس معیارهای RTF و FOM در جدول ۵ ارائه شده‌اند.

چنانچه جدول ۵ نشان می‌دهد، FOM روش CD-EDSTD در مقایسه با روش CI-EDSTD حدود ۰٫۶٪ بهبود داشته است. این بهبود دقت هزینه‌ای به اندازه ۰٫۳۶ افت سرعت واژه‌یاب گفتار متمایزساز پایه را دارد. بهبود دقت واژه‌یاب گفتار متمایزساز در مقابل این مقدار افت سرعت مقرون به صرفه می‌باشد.

1. Context Dependent - EDSTD (CD - EDSTD)



شکل ۹: نمودار ROC برای ارزیابی روش Mph - EDSTD در مقایسه با روش Tph - EDSTD.

جدول ۵: ارزیابی دو روش CI - EDSTD و CD - EDSTD بر اساس معیارهای FOM و RTF.

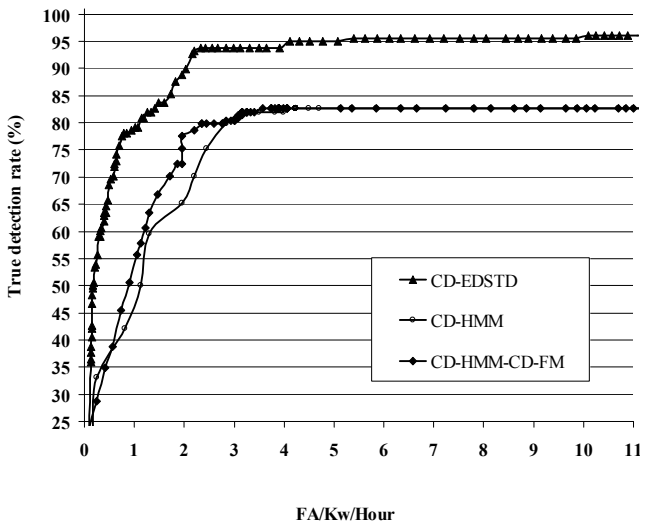
	FOM (%)	RTF
CI-EDSTD	۹۲/۴	۳/۶۵
CD-EDSTD	۹۳	۴/۰۶

به منظور تکمیل ارزیابی‌ها، روش CD-EDSTD با یک رویکرد واژه‌یابی گفتار مبتنی بر مدل مخفی مارکف مقایسه شده است. شکل ۱۰ نمودار ROC را برای ارزیابی روش CD-EDSTD در مقایسه با روش‌های CD-HMM و CD-HMM-CD-FM نشان می‌دهد. چنانچه در شکل ۱۰ نشان می‌دهد، روش CD-EDSTD در مقایسه با روش‌های CD-HMM و CD-HMM-CD-FM اخطار اشتباه یکسان از دقت تشخیص بالاتری برخوردار است. دقت تشخیص روش CD-EDSTD در مقایسه با روش‌های CD-HMM و CD-HMM-CD-FM، به ترتیب به طور متوسط حدود ۲۳٪ و ۲۰٪ بالاتر می‌باشد. همچنین روش CD-HMM-CD-FM که از یک مدل پرکننده به منظور نرمال‌نمودن امتیازات مربوط به کلمات کلیدی بهره می‌برد، در بهترین حالت حدود ۱۰٪ در بهترین حالت، بهتر از روش CD-HMM عمل کرده است. نتایج ارزیابی روش CD-EDSTD در مقایسه با رویکردهای واژه‌یابی گفتار مبتنی بر مدل مخفی مارکف بر اساس معیارهای RTF و FOM در جدول ۶ ارائه شده است.

چنانچه جدول ۶ نشان می‌دهد، روش CD-EDSTD در FOM مقایسه با FOM نسخه‌های مختلف واژه‌یاب گفتار مبتنی بر مدل مخفی مارکف به طور قابل ملاحظه‌ای بالاتر است. در مقابل سرعت روش CD-EDSTD در مقایسه با سرعت رویکردهای مبتنی بر مدل مخفی مارکف افت چشمگیری دارد. با توجه به آن که واژه‌یاب گفتار در چه کاربردی مورد استفاده قرار گرفته و اهمیت سرعت و دقت در آن کاربرد خاص، هر کدام از رویکردهای متمایزساز یا مبتنی بر مدل مخفی مارکف مورد استفاده قرار می‌گیرند.

۶- جمع‌بندی

ما در این مقاله روی سیستم‌های واژه‌یاب گفتار متمایزساز تمرکز کرده‌ایم. هر سیستم واژه‌یاب گفتار متمایزساز از دو بخش اصلی تشکیل



شکل ۱۰: نمودار ROC برای ارزیابی روش CD - EDSTD در مقایسه با روش‌های CD - HMM و CD - HMM - CD - FM.

جدول ۶: ارزیابی روش CD - EDSTD در مقایسه با رویکردهای واژه‌یابی گفتار مبتنی بر مدل مخفی مارکف بر اساس معیارهای FOM و RTF.

	FOM (%)	RTF
CD-EDSTD	۹۳	۴/۰۶
CD-HMM	۷۷/۶	۱/۸
CD-HMM-CD-FM	۷۹/۳	۱/۹

می‌شود: استخراج ویژگی و دسته‌بندی. در بخش دسته‌بندی از الگوریتمی تکاملی که در کارهای قبلی ما ارائه شده است، برای آموزش پارامترهای دسته‌بند استفاده گردید و در این مقاله روی بخش استخراج ویژگی تمرکز کرده‌ایم. در ابتدا یک رویکرد استخراج ویژگی مستقل از واج برای سیستم واژه‌یاب گفتار متمایزساز مبتنی بر الگوریتم تکاملی مذکور (EDSTD) ارائه نمودیم. سپس راهکاری برای به کارگیری اطلاعات وابسته به محتوا در بخش استخراج ویژگی سیستم EDSTD به منظور افزایش کارایی سیستم نهایی پیشنهاد گردید و رویکرد پیشنهادی روی دادگان TIMIT مورد ارزیابی قرار گرفته است. نتایج ارزیابی حاکی از آن است که نرخ بازشناسی درست سیستم EDSTD وابسته به محتوا (CD-EDSTD) در اخطار اشتباه بر کلمه کلیدی بر ساعت بزرگ‌تر از دو، حدود ۳٪ از نرخ بازشناسی درست سیستم EDSTD مستقل از محتوا (CI-EDSTD) بالاتر است. به علاوه مقدار FOM (متوسط نرخ تشخیص درست برای نرخ اخطارهای اشتباه بر کلمه کلیدی بر ساعت بین ۱ تا ۱۰) مربوط به سیستم CD-EDSTD ۰/۶٪ بزرگ‌تر از FOM سیستم CI-EDSTD می‌باشد. هزینه این بهبود دقت، حدود ۰/۳۶ واحد افت سرعت پاسخ‌گویی است که قابل چشم‌پوشی می‌باشد.

به منظور تکمیل ارزیابی‌ها، رویکرد CD-EDSTD با یک رویکرد واژه‌یابی گفتار مبتنی بر مدل مخفی مارکف وابسته به محتوا (CD-HMM-STD) مقایسه شده است. نتایج حاکی از آن است که نرخ تشخیص درست روش CD-EDSTD به طور متوسط به ترتیب حدود ۲۰٪ و ۲۳٪ بالاتر از رویکرد CD-HMM-STD با و بدون استفاده از مدل پرکننده می‌باشد. در مقابل سرعت روش CD-EDSTD در مقایسه با سرعت رویکردهای مبتنی بر مدل مخفی مارکف افت چشمگیری دارد. با توجه به آن که واژه‌یاب گفتار در چه کاربردی مورد استفاده قرار گرفته است و اهمیت سرعت و دقت در آن کاربرد خاص، هر یک از رویکردهای متمایزساز یا مبتنی بر مدل مخفی مارکف مورد استفاده قرار می‌گیرند.

۷- سپاس‌گزاری

از حمایت‌های مرکز تحقیقات مخابرات ایران در طول انجام این کار کمال سپاس‌گزاری را دارم.

مراجع

[1] M. Weintraub, "LVCSR log-likelihood ratio scoring for keyword spotting," in *Proc. of ICASSP'95*, vol. 1, pp. 129-132, May 1995.

[2] I. Szoke, et al., "Comparison of keyword spotting approaches for informal continuous speech," in *Proc. of Interspeech'05*, vol. 1, pp. 633-636, Sep. 2005.

[3] B. Ramabhadran, A. Sethy, J. Mamou, B. Kingsbury, and U. Chaudhari, "Fast decoding for open vocabulary spoken term detection," in *Proc. of NAACL HLT'09*, vol. 1, pp. 277-280, Jan. 2009.

[4] D. Wang, J. Tejedor, S. King, and J. Frankel, "Term-dependent confidence normalization for out-of-vocabulary spoken term detection," *J. of Computer Science and Technology*, vol. 27, no. 2, pp. 358-375, Mar. 2012.

[5] D. Wang, J. Tejedor, J. Frankel, S. King, and J. Colas, "Posterior based confidence measures for spoken term detection," in *Proc. of ICASSP'09*, vol. 8, pp. 4889-4892, Apr. 2009.

[6] R. C. Rose, "Keyword detection in conversational speech utterances using hidden markov model based continuous speech recognition," *Computer Speech & Language J.*, vol. 9, no. 4, pp. 309-333, Oct. 1995.

[7] H. Ketabdar, J. Vepa, S. Bengio, and H. Bourlard, "Posterior based keyword spotting with a priori thresholds," in *Proc. 9th Int. Conf. on Spoken Language Processing, ISCA'06*, vol. 1, 8 pp., May 2006.

[8] J. Tejedor, D. Wang, J. Frankel, S. King, and J. Colas, "A comparison of grapheme and phoneme-based units for Spanish spoken term detection," *Speech Communication*, vol. 50, no. 11-12, pp. 980-991, Nov./Dec. 2008.

[9] A. Shokri, S. Tabibian, A. Akbari, B. Nasersharif, and J. Kabudian, "Robust keyword spotting system for Persian conversational telephone speech using feature and score normalization and ARMA filter," in *Proc. of IEEE GCC Conf. and Exhibition*, vol. 1, pp. 497-500, Feb. 2011.

[10] I. Bazzi, Modeling Out-of-Vocabulary Words for Robust Speech Recognition, Ph.D. Thesis, MIT, 2002.

[11] I. Szoke, *Hybrid Word - Subword Spoken Term Detection*, Ph.D Thesis, Faculty of Information Technology BUT, 2010.

[12] J. Junkawitsch, G. Ruske, and H. Hoge, "Efficient methods for detecting keywords in continuous speech," in *Proc. Eurospeech Conf. on Speech Communication and Technology*, vol. 1, pp. 259-262, Sep. 1997.

[13] U. Yapanel, *Garbage Modeling Techniques for a Turkish Keyword Spotting System*, Master of Science Thesis, Electrical Engineering Department, Bogazici University, 2000.

[14] Z. De Greve, Application in Automatic Speech Recognition: Keyword Spotting Based on Online Garbage Modeling, Internship Report (IDIAP Research Institute), 2006.

[15] L. R. Bahl, P. F. Brown, P. De Souza, and R. L. Mercer, "Maximum mutual information estimation of hidden markov model parameters for speech recognition," in *Proc. of ICASSP'86*, vol. 1, pp. 49-52, Apr. 1986.

[16] B. Juang and S. Katagiri, "Discriminative learning for minimum error classification," *IEEE Trans. on Signal Processing*, vol. 40, no. 12, pp. 3043-3054, Dec. 1992.

[17] D. Povey and P. Woodland, "Minimum phone error and I-smoothing for improved discriminative training," in *Proc. of ICASSP'02*, vol. 1, pp. 105-108, May 2002.

[18] H. Jiang, X. Li, and C. Liu, "Large margin hidden markov models for speech recognition," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1584-1595, Sep. 2006.

[19] J. C. Chen and J. T. Chien, "Bayesian large margin hidden markov models for speech recognition," in *Proc. of ICASSP'09*, vol. 7, pp. 3765-3768, Apr. 2009.

[20] K. P. Li, J. A. Naylor, and M. L. Rossen, "A whole word recurrent neural network for keyword spotting," in *Proc. of ICASSP'92*, vol. 2, pp. 81-84, Mar. 1992.

[21] S. Fernandez, A. Graves, and J. Schmidhuber, "An application of recurrent neural networks to discriminative keyword spotting," in *Proc. Conf. on Artificial Neural Networks, ICANN'07*, vol. 4669, pp. 220-229, Sep. 2007.

[22] J. Keshet, D. Grangier, and S. Bengio, "Discriminative keyword spotting," *Speech Communication*, vol. 51, no. 4, pp. 317-329, Apr. 2009.

[23] S. Tabibian, A. Shokri, A. Akbari, and B. Nasersharif, "Performance evaluation for an HMM-based keyword spotter and a large-margin based one in noisy environments," in *Proc. World Conf. on Information Technology, Procedia Computer Science*, vol. 3, pp. 1018-1022, Jun. 2010.

[24] Y. Benayed, D. Fohr, J. P. Haton, and G. Chollet, "Improving the performance of a keyword spotting system by using support vector machines," *in Proc. IEEE Workshop on Automatic Speech Recognition & Understanding, ASRU*'03, vol. 1, pp. 145-149, 30 Nov.-3 Dec. 2003.

[25] J. Keshet and S. Bengio, Automatic Speech and Speaker Recognition, Large Margin and Kernel Methods, John Willy and Sons, 1st Edition, 2009.

[26] S. Tabibian, A. Akbari, and B. Nasersharif, "An evolutionary based discriminative system for keyword spotting," in *Proc. Symp. on Artificial Intelligence and Signal Processing, AISP'11,* vol. 1, pp. 83-88, Jan. 2011.

[27] S. Tabibian, A. Akbari, and B. Nasersharif, "A fast search technique for discriminative keyword spotting," in *Proc. Symp. on Artificial Intelligence and Signal Processing, AISP'12*, vol. 1, pp. 140-144, May 2012.

[28] V. N. Vapnik, *Statistical Learning Theory*, Wiley, 1998.

[29] J. Keshet, S. Shalev-Shwartz, Y. Singer, and D. Chazan, "Phoneme alignment based on discriminative learning," in *Proc. of InterSpeech*, vol. 5, pp. 2961-2964, Sep. 2005.

[30] J. Keshet, S. Shalev-Shwartz, S. Bengio, Y. Singer, and D. Chazan, "Discriminative kernel-based phoneme sequence recognition," in *Proc. of the Int. Conf. on Spoken Language Processing*, vol. 2, pp. 593-596, Sep. 2006.

[31] M. Wollmer, F. Eyben, J. Keshet, A. Graves, B. Schuller, and G. Rigoll, "Robust discriminative keyword spotting for emotionally colored spontaneous speech using bidirectional LSTM networks," in *Proc. of ICASSP'09*, vol. 7, pp. 3949-3952, Apr. 2009.

[32] D. T. Toledano, L. A. H. Gomez, and L. V. Grande, "Automatic phonetic segmentation," *IEEE Trans. on Speech and Audio Processing*, vol. 11, no. 6, pp. 617-625, Nov. 2003.

[33] J. W. Kuo, H. Y. Lo, and H. M. Wang, "Improved HMM/SVM methods for automatic phoneme segmentation," in *Proc. of the 10th European Conf. on Speech Communication and Technology (Interspeech 2007-Eurospeech)*, vol. 1, pp. 2057-2060, Aug. 2007.

[34] O. Dekel, J. Keshet, and Y. Singer, "Online algorithm for hierarchical phoneme classification, *Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Lecture Notes in Computer Science, Springer-Verlag, pp. 146-159, 2004.

[35] C. C. Chang and C. J. Lin, LIBSVM: *A Library for Support Vector Machines*, http://www.csie.ntu.edu.tw/~cjlin, 2009.

[36] J. C. Platt, Advances in Kernel Methods: Support Vector Learning, MIT Press, 1999.

[37] C. P. Chen, J. Blimes, and K. Kirchhoff, "Low-resource noise-robust feature post-processing on AURORA 2.0," in *Proc. of ICSLP'02*, pp. 2445-2448, Sep. 2002.

[38] A. M. Toh, R. Togneri, and S. Nordholm, "Spectral entropy as speech features for speech recognition," in *Proc. of Postgraduate Electrical Engineering and Computing Symposium, PEECS'05*, vol. 1, pp. 22-25, May 2005.

[39] G. Peeters, A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project, Cuidado Project Rep. Ircam, pp. 1-25, 2004.

[40] C. Y. Lin, J. S. Rager Jang, and K. T. Chen, "Automatic segmentation and labeling for Mandarin Chinese speech corpora for concatenation-based TTS," *Computational Linguistics and Chinese Language Processing*, vol. 10, no. 2, pp. 145-166, Jun. 2005.

[41] J. Keshet, S. Shalev-Shwartz, Y. Singer, and D. Chazan, "A large-margin algorithm for speech-to-phoneme and music-to-score alignment," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2373-2382, Nov. 2007.

[42] S. A. Hejazi, R. Kazemi, and S. Ghaemmaghami, "Isolated persian digit recognition using a hybrid HMM-SVM," in *Proc. Int. Symp. on Intelligent Signal Processing and Communication Systems*, 4 pp., Feb. 2008.

[43] S. Chandrakala and C. Chandra Sekhar, "Combination of generative models and SVM based classifier for speech emotion recognition," in *Proc. of Int. Joint Conf. on Neural Networks*, vol. 1, pp. 497-502, Jun. 2009.

[44] J. Stadermann and G. Rigoll, "A hybrid SVM/HMM acoustic modeling approach to automatic speech recognition," in *Proc. of Interspeech'04*, vol. 1, pp. 661-664, Oct. 2004.

[45] Q. Zhi-Yi, L. Yu, Z. Li-Hong, and S. Ming-Xin, "A speech recognition system based on a hybrid HMM/SVM architecture," in *Proc. of the 1st Int. Conf. on Innovative Computing, Information, and Control*, vol. 2, pp. 100-104, 30 Aug.-1 Sep. 2006.

[46] A. R. Ahmad, C. Viard-Gaudin, and M. Khalid, "Lexicon-based word recognition using support vector machine and hidden markov model," in *Proc. of the Int. Conf. on Document Analysis and Recognition*, vol. 1, pp. 161-165, Jul. 2009.

[47] L. Lori, R. Kassel, and S. Stephanie, "Speech database development: design and analysis of the acoustic-phonetic corpus," in *Proc. of DARPA Speech Recognition Workshop*, vol. 2, pp. 161-170, pp. 100-109, Feb. 1986.

[48] D. G. Zacharie and J. P. Pinto, *Keyword Spotting on Word Lattices*, Research Report, IDIAP Research Institute, 2007.

شیما طبیبیان تحصیلات خود را در مقطع کارشناسی مهندسی کامپیوتر- گرایش نرم‌افزار در سال ۱۳۸۳ از دانشگاه صنعتی اصفهان به پایان رسانده است. در سال ۱۳۸۴ در مقطع کارشناسی ارشد رشته مهندسی کامپیوتر- گرایش هوش مصنوعی و رباتیک در دانشگاه علم و صنعت ایران پذیرفته و پس از اتمام تحصیلات خود در این مقطع در سال ۱۳۸۶، در مقطع دکتری همان رشته در دانشکده مهندسی کامپیوتر همان دانشگاه پذیرفته شد. وی در سال ۱۳۹۲ با کسب درجه عالی از رساله دکترای خود دفاع نمود و هم‌اکنون عضو هیأت علمی پژوهشکده سامانه‌های فضانوردی در پژوهشگاه هوافضا می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش و بازشناسی گفتار، واژه‌یابی گفتار، بهسازی گفتار، رابطهای کاربری مبتنی بر گفتار، تشخیص فرامین صوتی، شناسایی آماری الگو، الگوریتم‌های تکاملی و پردازش تصویر.

احمد اکبری در سال ۱۳۶۶ مدرک کارشناسی مهندسی برق و الکترونیک خود و در سال ۱۳۶۹ مدرک کارشناسی ارشد مهندسی برق و مخابرات خود را از دانشگاه صنعتی اصفهان دریافت نمود. وی در سال ۱۹۹۵ میلادی موفق به اخذ مدرک دکتری در پردازش سیگنال از دانشگاه رن۱ فرانسه گردید. دکتر اکبری از سال ۱۳۷۴ تاکنون عضو هیأت علمی دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران می‌باشد. زمینه‌های علمی مورد علاقه نام‌برده متنوع بوده و شامل موضوعاتی مانند بازشناسی گفتار، بهسازی گفتار، سیستم‌های گفتگو، واسط ارتباط انسان و کامپیوتر مبتنی بر گفتار، کاربرد صدا روی وب، شبکه‌های کامپیوتری و امنیت شبکه می‌باشد.

بابک ناصرشریف درجه کارشناسی را در رشته مهندسی کامپیوتر گرایش سخت افزار از دانشگاه صنعتی امیرکبیر در سال ۱۳۷۶ دریافت نمود و موفق به اخذ درجه کارشناسی ارشد و دکتری در رشته مهندسی کامپیوتر گرایش هوش مصنوعی از دانشگاه علم و صنعت ایران در سال‌های ۱۳۷۹ و ۱۳۸۶ گردید. نام‌برده از سال ۱۳۸۶ تا ۱۳۹۰ عضو هیأت علمی گروه مهندسی کامپیوتر در دانشکده فنی گیلان و از سال ۱۳۹۰ تا کنون عضو هیأت علمی دانشکده مهندسی کامپیوتر در دانشگاه صنعتی خواجه نصیر طوسی است. زمینه تحقیقاتی ایشان بهبود گفتار، بازشناسی گفتار و مقاوم سازی آن، واژه‌یابی گفتار و مدل‌سازی و بازشناسی الگو است.