بهبود کارایی سیستم کاوش‌گر کلمات تلفنی با استفاده از نرمالیزاسیون امتیاز اطمینان مبتنی بر روش برنامه‌ریزی خطّی

یاسر شکفته[۱و۲]، جهانشاه کبودیان[۱و۳]، محمّدمحسن گودرزی[۱و۲]، ایمان صراف رضایی[۱]
[۱] گروه پردازش صوت، پژوهشکدهٔ پردازش هوشمند علائم (RCISP)، تهران، ایران.
[۲] آزمایشگاه پردازش گفتار، دانشکدهٔ مهندسی‌پزشکی، دانشگاه صنعتی امیرکبیر، تهران، ایران.
[۳] گروه مهندسی کامپیوتر، دانشکدهٔ فنی- مهندسی، دانشگاه رازی، کرمانشاه، ایران.

چکیده

سیستم‌های متداول کاوش‌گر کلمات دارای یک مدل بازشناسی گفتار هستند که وظیفهٔ آن تعیین کلیدواژه‌های کاندید شده و امتیاز اطمینان آن‌هاست. به‌طور معمول قبول و یا رد کلیدواژه‌های کاندید شده بر مبنای مقایسهٔ این امتیاز با یک مقدار آستانهٔ ثابت انجام می‌گیرد. از آن‌جا که عمل‌کرد مدل بازشناس در تشخیص واحدهای زیرکلمه‌ای متفاوت، یکسان نمی‌باشد؛ بنابراین اختصاص امتیاز اطمینان برای هر کلیدواژه بدون در نظر گرفتن ساختار واحدهای زیرکلمه‌ای آن مناسب نیست. از این رو در این مقاله یک روش به‌طور کامل جدید نرمالیزاسیون امتیاز اطمینان بر اساس ساختار واجی کلیدواژه‌ها و روش برنامه‌ریزی خطی ارائه شده است. هدف این روش امتیازدهی به اجزای واجی هر کلیدواژه، براساس بیشینه نمودن تفکیک توزیع امتیاز اطمینان اولیه کلیدواژه‌های درست و غلط تشخیص داده شده است. نتایج به دست آمده نشان می‌دهد که استفاده از روش پیشنهادی منجر به بهبود دو درصدی معیار FOM نسبت به سیستم پایه خواهد شد.

واژگان کلیدی: سیستم کاوش‌گر کلمات، کلید واژه، مدل پنهان مارکوف، امتیاز اطمینان، برنامه‌ریزی خطی، نرمالیزاسیون امتیاز.

۱- مقدمه

جستجو و تشخیص کلیدواژه‌ها[۱] (KWS) یک شاخهٔ کلیدی از حوزهٔ بازشناسی خودکار گفتار[۲] (ASR) است که وظیفهٔ تشخیص مجموعه‌ای از کلمات از پیش تعیین شده را در یک گویهٔ خاص گفتاری برعهده دارد. به‌طور معمول سیستم‌هایی که جستجوی کلیدواژه‌ها را انجام می‌دهند، با سرعتی چندین برابر یک سیستم زمان واقعی عمل می‌کنند. ازاین‌رو حجم بزرگی از فایل های صوتی و گفتاری را می‌توان در مدّت زمان کوتاهی جستجو کرد. حوزهٔ تشخیص کلیدواژه‌ها در دو دههٔ اخیر بسیار توسعه یافته است؛ به‌گونه‌ای که مؤسسهٔ NIST، مسابقاتی را برای بررسی و ارزیابی این سیستم‌ها برگزار نموده است *(NIST, 2006)*.

اولین سیستم کاوش‌گر کلمات در سال ۱۹۷۳ میلادی پیشنهاد شد *(Bridle, 1973)*. این سیستم که مبتنی بر الگو[۳] نامیده می‌شد، به دنبال یافتن دنباله‌ای از قاب‌های ویژگی گفتاری بود که با الگوی کلیدواژهٔ تطابق بیشتری داشته باشند. عمل تطابق نیز به‌وسیلهٔ الگوریتم پیچش زمانی پویا[۴] (DTW) انجام می‌گرفت. در دههٔ ۸۰ میلادی این الگوریتم به‌وسیلهٔ برخی از محقّقان توسعهٔ بیشتری یافت *(Richard, et al., 1977; Myers, et al., 1980)*. سپس در سال ۱۹۸۹ توسعه این روش با مدل مخفی مارکوف[۵] (HMM) انجام یافت *(Wilpon, et al., 1989)*. امّا در سال ۱۹۹۰ بود که *ویلپون* علاوه بر تولید مدل کلیدواژه‌ها از مفهوم مدل زباله[۶] (GM) برای مدل نمودن اجزای گفتاری غیرکلیدواژه استفاده کرد *(Wilpon, et al., 1990)*. مفهوم مدل زباله تعریف شده توسط وی معادل با مدل پرکننده[۷] (FM) در مقالات ارائه شدهٔ امروزی است (به‌طور مثال در *(Manos, et al., 1997)*). در ادامه مدل‌های متنوعی برای

[1] Keywords
[2] Automatic Speech Recognition
[3] Template-based Keyword Spotting
[4] Dynamic Time Warping
[5] Hidden Markov Model
[6] Garbage Model
[7] Filler Model

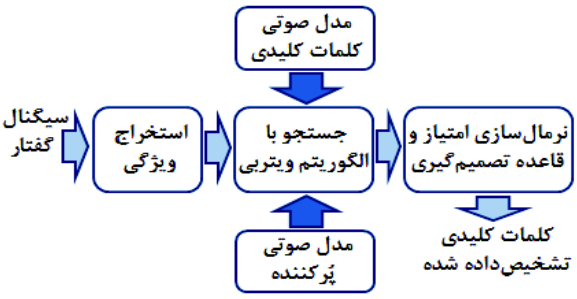
توسعهٔ مدل پرکننده ارائه شد که بهترین آن بوسیلهٔ رُز و با استفاده از مدل‌های سه آوایی[1] گزارش شده است *(Rose, 1990)*. سپس مدل ضدّکلمه‌ای[2] آن معرّفی شد که در آن به‌ازای هر کلیدواژهٔ انتخابی، یک مدل غیرکلمه‌ای متناظر با آن کلیدواژه نیز می‌بایست مورد تعلیم قرار می‌گرفت *(Rahim, et al., 1995)*. از این‌رو اغلب سیستم‌های کاوش‌گر کلمات اولیه، براساس مدل کلیدواژه و مدل پرکننده بودند. در همین اواخر سیستم‌هایی بر اساس مقادیر احتمالاتی پسین مدل‌های زیرکلمه نیز مطرح شده است *(Meliani, et al., 1997; Szoke, et al., 2005; Pinto, et al., 2008;Tejedor, et al., 2008)*. در این روش امتیاز اطمینان[3] هر کلیدواژه براساس نسبت مقادیر احتمالاتی پسین به دست آمده برای مدل‌های حاوی کلیدواژه و مدل پرکننده است؛ درحالی‌که در سیستم‌های مبتنی بر مدل کلیدواژه و پرکننده با HMM، این امتیاز از تفاضل امتیاز لگاریتم درست‌نمایی[4] مدل‌های کلیدواژه و مدل پرکننده تعیین می‌شود.

از آن‌جا که در مدل‌های بازشناس گفتار پیوستهٔ مبتنی بر HMM، تمامی کلمات براساس مدل‌های زیرکلمه‌ای (مانند مدل‌های سه‌آوایی برای هر واج) تعریف می‌شوند، درنظر گرفتن امتیاز لگاریتم درست‌نمایی برای کل کلمه و بدون درنظرگرفتن اجزای واجی آن یک ضعف عمده در محاسبهٔ مقدار امتیاز اطمینان مربوط به آن کلمه محسوب می‌شود. از این‌رو در این مقاله روشی پیشنهاد می‌شود که بتواند با توجه به اجزای زیرکلمه‌ای هر کلیدواژه، یک امتیاز کلی برای ساختار واجی آن کلیدواژه درنظر بگیرد تا به این صورت خطای دقّت مدل بازشناس گفتار در فرایند جستجوی کلیدواژه‌ها در سیستم کاوش‌گر کلمات درنظر گرفته شود.

ساختار مقاله به‌صورت زیر نگاشته شده است: در بخش دوم، سیستم کاوش‌گر کلمات مبتنی بر HMM معرّفی می‌شود. سپس در بخش سوم، روش پیشنهادی برای یافتن امتیاز هر واحد زیرکلمهٔ واجی و چگونگی نرمالیزاسیون امتیازهای اطمینان کلیدواژه‌های کاندید شده در فرایند جستجوی کلمات معرفی خواهد شد. بخش چهارم معیارهای ارزیابی مختلف را در حوزهٔ جستجوی کلمات معرفی می‌نماید. در بخش پنجم، مجموعه دادگان مورد استفاده معرفی خواهد شد و سپس به بررسی و ارزیابی روش پیشنهادی با ارایهٔ مدل سیستم مناسب پایه جستجوی کلمات خواهیم پرداخت. در بخش آخر نیز جمع‌بندی و نتیجه‌گیری مقاله آورده شده است.

۲- معرفی روش پایه سیستم کاوش‌گر کلمه مبتنی بر HMM

در سیستم کاوش‌گر کلمه مبتنی بر HMM، متناظر با سیگنال گفتار ورودی، یک رشته خروجی متشکل از کلیدواژه‌ها و واحدهای زیرکلمه‌ای دیگر (مانند واج‌ها) در خروجی سیستم حاصل می‌شود *(Rose, et al., 1990)*. مدلی که وظیفهٔ تولید بخش‌های متناظر با غیرکلیدواژه را دارد، مدل پرکننده (FM) نامیده می‌شود. مرحلهٔ دوم سیستم، شامل بررسی صحّت اعتبار کلیدواژه‌های کاندید شده است. از این‌رو امتیازات آکوستیکی متناظر با هر کلیدواژه واقع در رشته خروجی، در یک واحد تصمیم‌گیری مورد پردازش قرار می‌گیرد تا امتیاز اطمینان برای هر کلیدواژهٔ کاندیدشده، تعیین و سپس با مقایسه این امتیاز با یک مقدار آستانه از پیش تعیین شده، اعتبار کلیدواژهٔ کاندید شده قبول یا رد شود. در (شکل ۱) بلوک دیاگرام این سیستم نشان داده شده است:



(شکل ۱): بلوک دیاگرام سیستم کاوشگر لغات مبتنی بر سیستم بازشناس گفتار پیوسته HMM.

اگر در محاسبهٔ امتیاز اطمینان برای کلیدواژه کاندید شده دوباره از مدل مخفی مارکوف استفاده شود، به سیستم حاصل، سیستم کاوش‌گر کلمه مبتنی بر HMM دو مرحله‌ای گفته می‌شود *(Manos, et al., 1997)* که الگوریتم کامل مراحل آن در ادامه آورده شده است:

[1] Triphone
[2] anti-word
[3] Confidence Score
[4] Log-Likelihood

۱- تعیین کلیدواژه‌ها: کلیدواژه‌های مورد نظر به همراه دنبالهٔ واجی و کلیه تنوعات تلفظی آن به مجموعه لغات[1] اضافه می‌شود.

۲- تولید شبکهٔ کلمات: برای هر یک از کلیدواژه‌های مورد نظر، یک شاخه به شبکهٔ کلمات[2] اضافه می‌شود. این شبکهٔ کلمات، مشابه شبکه‌ای است که در حالت عادی برای بازشناسی واجی مورد استفاده قرار می‌گیرد، با این تفاوت که شاخه‌هایی نیز برای کلیدواژه‌ها در آن در نظر گرفته شده است. (شکل ۲) نمونه‌ای از شبکهٔ کلمات را نشان می‌دهد که در این‌جا Net1 نامیده می‌شود.

۳- بازشناسی اولیه با استفاده از شبکهٔ کلمات و مدل‌های آکوستیکی تعلیم‌یافته: در این مرحله نتیجه بازشناسی اولیه شامل ترکیبی از واجها (مدل‌های پُرکننده) به همراه برخی از کلیدواژه‌های کاندید شده، تعیین می‌شود.

۴- تعیین کلیدواژه‌های کاندیدشده: در این مرحله با استفاده از خروجی مرحلهٔ سه، کلماتی که به عنوان کلیدواژه کاندید شده‌اند، شناسایی و دنبالهٔ قاب‌های گفتاری مربوط به آن‌ها از فایل‌های ویژگی متناظرشان جدا شده (با استفاده از محدوده زمانی متناظر با آن‌ها) و به‌صورت جداگانه در فایل‌های جدیدی جهت بازشناسی مجدد ذخیره می‌شود.

۵- بازشناسی مجدد کلیدواژه‌های کاندید شده توسط سیستم بازشناس فقط واجی: در این مرحله ویژگی‌های متناظر با کلیدواژه‌های کاندید شده در مرحلهٔ چهار، دوباره توسط سیستم مبتنی بر HMM ای بازشناسی می‌شود که به‌انحصار متشکل از دنبالهٔ واحدهای زیرکلمه‌ای (بدون کلیدواژه) می‌باشد. از این رو نتیجهٔ بازشناسی این مرحله فقط شامل دنباله‌ای از واجها خواهد بود. به این نحوهٔ مدل‌سازی، مدل پس‌زمینه[3] (BM) نیز گفته می‌شود. ساختار این مدل (که Net2 نامیده شده است) در (شکل ۳) نشان داده شده است.

۶- محاسبه امتیاز اطمینان: در این مرحله برای محاسبهٔ امتیاز اطمینان متناظر با هر کلیدواژه کاندید شده، تفاضل مقدار لگاریتم درست‌نمائی به دست آمده از شبکهٔ کلمات Net1 و مدل پس زمینه Net2 هر کلیدواژه کاندید شده به صورت زیر محاسبه می‌شود:

۱- تعیین امتیاز کلیدواژهٔ کاندید شده (KW Score):

$$Score(i,j) = \frac{LogLikelihood(Net1)}{NF(j)} \qquad (۱)$$

۲- تعیین امتیاز مدل پس‌زمینه (BG Score):

$$Score(BG(j)) = \frac{LogLikelihood(Net2)}{NF(j)} \qquad (۲)$$

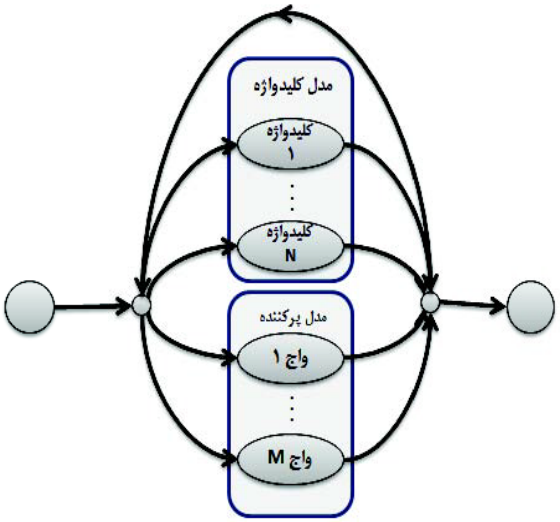
۳- محاسبهٔ امتیاز اطمینان[4]:

$$S_j^i = Score(i,j) - Score(BG(j)) \qquad (۳)$$

که در آن $NF(j)$ تعداد قاب‌های گفتاری[5] مرتبط با j اُمین رخداد کلید واژه i اُم از مجموعه لغات کلیدواژه‌ها می‌باشد. لازم به ذکر است که روش تقسیم امتیاز بر تعداد قاب‌ها (روابط ۱ و ۲) یک روش نرمال‌سازی امتیاز محسوب می‌شود که در سیستم پایهٔ این مقاله از آن استفاده شده است.

از آن‌جا که انتظار داریم امتیاز به دست آمده از مدل پس‌زمینه همواره بیشتر از مدل کلیدواژه باشد، بنابراین امتیاز اطمینان به دست آمده به‌طور عمومی منفی خواهد بود. هم‌چنین هر چه اندازهٔ اختلاف این دو امتیاز کم‌تر باشد، احتمال صحیح بودن فرض قبولی کلیدواژهٔ کاندیدشده بیشتر و با افزایش اندازهٔ اختلاف (منفی‌تر شدن امتیاز اطمینان)، احتمال فرض ردشدن کلیدواژهٔ کاندید شده قوت می‌یابد.

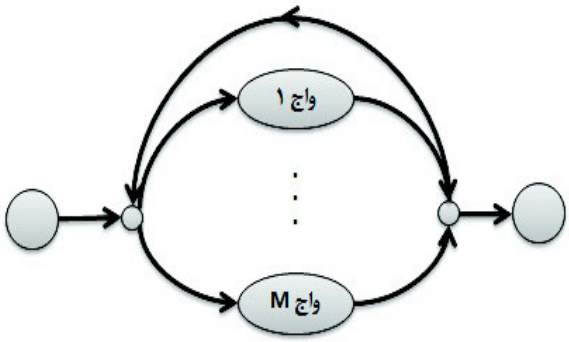
۷- تصمیم‌گیری رد یا قبول کلیدواژه کاندیدشده: با توجه به امتیاز اطمینان به دست آمده از مرحلهٔ ۷ و مقایسهٔ آن با یک مقدار آستانهٔ[6] از پیش تعیین شده، کلیدواژه کاندید شده قبول یا رد می‌شود.



(شکل ۲): مدل شبکهٔ کلمات (Net1) متشکل از N مدل کلیدواژه و M مدل پرکننده برای تولید رشته خروجی متناظر با سیگنال گفتار ورودی شامل کلیدواژه‌ها و مدل‌های پرکننده

[1] Dictionary
[2] Word Network
[3] Background Model

[4] Confidence Score
[5] Number of Frames
[6] Threshold



(شکل ۳): مدل پس‌زمینه (Net2) متشکل از M واج متناظر با مدل کلی هر کلیدواژه کاندیدشده

۳- روش پیشنهادی برای نرمالیزاسیون امتیاز اطمینان کلیدواژه بر مبنای اجزای تشکیل دهندهٔ آن

در بخش دو اشاره شد که پس از تشخیص کلیدواژه‌های کاندیدشده، متناظر با هر یک از کلیدواژه‌ها، امتیاز اطمینانی محاسبه می‌شود که نشان دهندهٔ احتمال درست بودن تشخیص کلیدواژهٔ متناظر با آن است. از آن‌جا که برخی از واحدهای زیرکلمهٔ واجی (به‌خصوص واک‌دارها) دارای درصد صحّت بازشناسی بالا و برعکس برخی از واحدهای زیرکلمهٔ واجی (به‌خصوص انفجاری‌ها) دارای درصد صحت بازشناسی کمتری توسط سیستم‌های متداول بازشناسی گفتار هستند؛ از این‌رو استفاده از روشی برای نرمالیزاسیون امتیاز پیشنهاد می‌شود که بتواند با توجّه به نوع و تعداد واج‌های تشکیل دهندهٔ یک کلیدواژه، به نرمال‌سازی امتیاز اطمینان آن‌ها بپردازد. با این رویکرد می‌توان برای تلفظ‌های مختلف از یک کلیدواژه، یک امتیاز براساس اجزای زیرکلمه‌ای آن که نشان‌دهندهٔ احتمال درست تشخیص داده شدن آن توسط مدل بازشناس است، تعیین نمود و در مرحلهٔ تصمیم‌گیری سیستم کاوش‌گر کلمات از آن استفاده کرد. به طور مثال اگر کلیدواژه کاندید شده از واج‌هایی با درصد صحّت بازشناسی بیشتری تشکیل شده باشد، صحت اعتبار این کلمه می‌تواند با مقادیر آستانهٔ امتیاز اطمینان کم‌تری نسبت به سایر کلمات کلیدی دیگر مورد پذیرش و قبول واقع شود. بنابراین در روش پیشنهادی، امتیاز رخداد jاُم از کلیدواژه i اُم توسط یک مقدار ثابت G و یک مقدار ΔS_i که وابسته به کلمهٔ کلیدی i اُم است، تصحیح خواهد شد.

با فرض در نظر گرفتن واحد آوایی واج برای بخش‌های زیرکلمه‌ای هر کلیدواژه، در این حالت میزان تصحیح در امتیاز کلیدواژه i اُم را می‌توان با استفاده از رابطهٔ زیر محاسبه کرد:

$$\Delta S_i = \frac{w_1 \mid Ph_1^i \mid + \ldots + w_M \mid Ph_M^i \mid}{\sum_{l=1}^{M} \mid Ph_l^i \mid} = \frac{\sum_{l=1}^{M} w_l \mid Ph_l^i \mid}{L_i} \qquad (۴)$$

در رابطهٔ فوق، $\mid Ph_l^i \mid$ تعداد تکرار واج l ام از زبان فارسی در کلیدواژه i اُم و w_l وزن مرتبط با اهمیت آن واج است. مقدار L_i نیز برابر با تعداد واج‌های تشکیل دهندهٔ کلیدواژه i اُم است. هم‌چنین عدد M نشان‌گر تعداد واحدهای واجی مورد استفاده در مدل بازشناس می‌باشد. از آن‌جا که در پیاده‌سازی‌های این مقاله از تعداد ۲۹ واج فارسی (به‌عنوان واحدهای زیرکلمه) برای واج‌نویسی کلیدواژه‌ها استفاده شده است، بنابراین بایستی تعداد ۲۹ وزن متناظر با تعداد واج‌های زبان فارسی (M=29) به‌صورت بهینه محاسبه شود.

پس از تعیین امتیاز نرمالیزهٔ هر کلیدواژه (با استفاده از روش پیشنهادی) و هم‌چنین تعیین امتیاز اطمینان مرتبط با هر کلیدواژه (محاسبه شده از طریق الگوریتم هفت مرحله‌ای بخش دو)، از رابطهٔ زیر برای محاسبهٔ امتیاز نهایی هر کلیدواژهٔ کاندید شده، استفاده خواهیم نمود:

$$\hat{S}_j^i = S_j^i - G + \Delta S_i \qquad (۵)$$

در رابط، فوق مقدار G به‌عنوان مقدار بایاس تخمین روش نرمالیزاسیون است که مفهوم آن معادل با مقدار امتیاز مناسب برای جداسازی توزیع امتیازهای اطمینان مرتبط با رخدادهای درست و نادرست از کلیدواژه‌های سیستم در دادگان مورد استفاده است.

آن‌گونه که در بخش دوم مقاله نیز بحث شد، توزیع امتیاز اطمینان رخدادهای درست از کلیدواژه‌ها حول مقادیر منفی نزدیک به صفر و توزیع امتیاز اطمینان رخدادهای نادرست از کلیدواژه‌ها حول مقادیر بسیار منفی‌تر قرار دارد. اگر فرض کنیم که جداسازی توزیع امتیاز اطمینان متناظر با رخدادهای درست و نادرستِ کلیدواژه‌ها یک مسأله جداسازی دوکلاسه است؛ مطلوب آن است که این دو توزیع به‌طور کامل از هم تفکیک‌پذیر باشند و با انتخاب یک مقدار

سطح آستانه که در بین مقادیر این دو توزیع قرار می‌گیرد، خطای رد رخدادهای درست و خطای قبول رخدادهای نادرست صفر شود. درعمل چنین فرضی غیرممکن است؛ امّا می‌توانیم پارامترهای مجهول وزن w_l و بایاس G را به‌گونه‌ای تعیین کنیم که دو توزیع امتیاز $\hat{S}_j^i$ متناظر با رخدادهای درست و نادرست بیشترین تفکیک‌پذیری را داشته باشند.

در این‌جا می‌توانیم برای محاسبهٔ پارامترهای مجهول وزن w_l و بایاس G از مفاهیم مرتبط با ماشین بردار پشتیبان[1] استفاده نماییم. برای این منظور فرض می‌کنیم که استفاده از امتیازات اطمینان نهایی نرمالیزه شده مرتبط با هر کلیدواژهٔ کاندیدشده باعث تفکیک بیشتر دو کلاس مرتبط با تمامی رخدادهای درست و نادرست از کل کلیدواژه‌ها در دادگان شود. هم‌چنین فرض می‌کنیم که مقدار فاصلهٔ بین مرز دو کلاس مقدار 2H باشد که در این‌جا مقدار H را مقدار حاشیه[2] می‌نامیم. اگر رخداد j اُم از کلیدواژه i اُم در دادگان آموزشی درست باشد، مطلوب آن است که رابطهٔ زیر برقرار باشد:

$$\hat{S}_j^i \geq +H \qquad (۶)$$

از آن‌جا که چنین شرطی لزوماً همیشه برقرار نمی‌باشد، یک متغیر کمکی ξ به نام متغیر منعطف[3] (SV) را به این صورت تعریف می‌کنیم:

$$\begin{aligned} &\hat{S}_j^i \geq +H - \xi_j^i, \\ &s.t. \qquad \xi_j^i \geq 0 \end{aligned} \qquad (۷)$$

که اگر یک نمونه از امتیازات رخدادهای درست در سمت مربوط به کلاس خود باشد، مقدار SV صفر است و در صورتی که نمونهٔ مذکور در داخل نوار مرزی و یا در سمت مربوط به کلاس دیگر (رخدادهای غلط) باشد، مقدار آن بزرگ‌تر از صفر و متناسب با دوری از مرز کلاس خود است؛ از طرفی دیگر اگر رخداد j اُم از کلیدواژه i اُم در دادگان آموزشی نادرست باشد، مطلوب آن است که رابطهٔ زیر برقرار باشد:

$$\hat{S}_j^i \leq -H \qquad (۸)$$

به‌طور مشابه از آن‌جا که چنین شرطی لزوماً همیشه برقرار نمی‌باشد، رابطهٔ تصحیح شده را به این صورت تعریف می‌کنیم:

$$\begin{aligned} &\hat{S}_j^i \leq -H + \xi_j^i, \\ &s.t. \qquad \xi_j^i \geq 0 \end{aligned} \qquad (۹)$$

که اگر یک نمونه از امتیازات رخدادهای نادرست در سمت مربوط به کلاس خود باشد، مقدار متغیر منعطف صفر است و در صورتی که نمونهٔ مذکور در داخل نوار مرزی و یا در سمت مربوط به کلاس دیگر باشد، مقدار آن بزرگ‌تر از صفر و متناسب با دوری از مرز کلاس خود است. سپس دو رابطهٔ متناظر با رخدادهای درست و نادرست را در یک رابطه و به‌صورت زیر تجمیع می‌کنیم:

$$\begin{aligned} &T_j^i \times \hat{S}_j^i \geq +H - \xi_j^i, \\ &s.t. \qquad \xi_j^i \geq 0 \end{aligned} \qquad (۱۰)$$

که در آن $\hat{S}_j^i$ امتیاز نهایی و نرمالیزه شدهٔ متناظر با رخداد j اُم از کلیدواژه کاندیدشده i اُم، ξ_j^i متغیر SV متناظر با آن رخداد کلیدواژه و T_j^i برچسب درستی و یا نادرستی آن رخداد است که به‌صورت زیر تعریف می‌شود:

$$T_j^i = \begin{cases} +1 \text{ if } j\text{-th occurence of } KW(i) \\ \quad \text{is True} \\ -1 \text{ if } j\text{-th occurence of } KW(i) \\ \quad \text{is False} \end{cases} \qquad (۱۱)$$

حالت ایده‌آل آن است که تمامی SVهای متناظر با هر ξ_j^i برابر با صفر باشند؛ امّا درعمل به این صورت نیست و باید کوشش کنیم مجموع متغیرهای منعطف را کمینه کنیم.

در مسئلهٔ بهینه‌سازی فوق، هدف یافتن بردار مجهول x شامل پارامترهای مجهول وزن w_l، مقدار بایاس G و متغیرهای منعطف ξ_j^i، در جهت کمینه کردن مجموع SVها است که به‌صورت زیر بیان می‌شود:

$$\begin{aligned} &\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \boldsymbol{f}^T\boldsymbol{x} = \arg \min_{\boldsymbol{x}} \sum_{i=1}^{N}\sum_{j=1}^{N_i^{Occ}} \xi_j^i, \\ &\boldsymbol{x}^T = [w_1,...,w_M,G,...,\xi_j^i,...], \\ &\boldsymbol{f}^T = [0,...,0,0,1,...,1]_{1\times(M+1+N^{Occ})} \end{aligned} \qquad (۱۲)$$

به شرطی که محدودیت‌های زیر برقرار باشد:

$$\begin{aligned} &T_j^i \times \hat{S}_j^i \geq +H - \xi_j^i, \\ &\xi_j^i \geq 0, \\ &0 \leq w_l \leq 1. \end{aligned} \qquad (۱۳)$$

در رابطهٔ (۱۲) N_i^{Occ} مجموع تعداد رخدادهای (کاندیدها) درست و نادرست از کلیدواژه i اُم و N^{Occ} مجموع تعداد کل رخدادهای درست و نادرست از تمامی کلیدواژه‌های سیستم در دادگان آموزشی است. هم‌چنین

[1] Support Vector Machine
[2] Margin
[3] Slack Variable

بردار ستونی f شامل M+1 عنصر با مقدارِ صفر و N^{occ} عنصر با مقدارِ یک می‌باشد. حدود پایین و بالا برای هر یک از عناصر بردار x در رابطهٔ (۱۲) به‌صورت زیر خواهد بود:

$$0 \le w_l \le 1,$$
$$-\infty \le G \le +\infty,$$
$$0 \le \xi_j^i \le +\infty. \qquad (۱۴)$$

در این‌جا برای محاسبه و تخمین بردار مجهول x در معادلات فوق، از روش بهینه‌سازی برنامه‌ریزی خطّی[1] (LP) استفاده شده است. پس از حل مسألهٔ برنامه‌ریزی خطّی فوق و تخمین بردار مجهول x ، مقادیر وزن واجی w_l و مقدار بایاس G به دست خواهد آمد.

۳-۱- روش برنامه‌ریزی خطّی

برنامه‌ریزی خطّی یا همان بهینه‌سازی خطّی، روشی در ریاضیات است که به پیدا کردن مقدار کمینه یا بیشینه از یک تابع خطی از متغیرها روی یک چندضلعی محدب[2] می‌پردازد. این چندضلعی محدب در حقیقت نمایش نموداری تعدادی قید از نوع معادله یا نامعادله روی متغیرهای تابع است. به بیان ساده‌تر با برنامه‌ریزی خطّی می‌توان بهترین نتیجه (به‌عنوان مثال بیش‌ترین سود یا کم‌ترین هزینه) را در شرایط خاص و با محدودیت‌های خاص به‌دست آورد. محل اصلی استفادهٔ برنامه‌ریزی خطّی در اقتصاد است؛ امّا در مهندسی نیز کاربردهای فراوانی دارد.

در این مقاله، هدف یافتن بردار x برای کمینه‌سازی تابع هزینهٔ خطی $\boldsymbol{f}^T\boldsymbol{x}$ با استفاده از روش برنامه‌ریزی خطّی و هم‌چنین درنظر گرفتن برخی از قیود می‌باشد. در این حالت خاص، قیود مورد نظر، حدهای بالا و پایین برای متغیر x و نیز معادلاتی در شکل نامساوی و به‌صورت زیر هستند:

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \boldsymbol{f}^T\boldsymbol{x}$$
$$s.t.\begin{cases} A\boldsymbol{x} \le \boldsymbol{b}, \\ \boldsymbol{\ell} \le \boldsymbol{x} \le \boldsymbol{u} \end{cases} \qquad (۱۵)$$

روش متداول برای حل این نوع مسأله، استفاده از بستهٔ نرم‌افزاری LIPSOL[3] است که برای مسائل با ابعاد و حجم دادهٔ بزرگ توسط ژانگ پیشنهاد شده است *(Zhang, 1995)*. این روش توسعه‌ای از الگوریتم پیشبین- تصحیح کنندهٔ مهروترا[4] محسوب می‌شود *(Mehrotra, 1992)* که در آن از حل‌کننده‌های برنامه‌ریزی خطی تُنُک[5] بر اساس روش‌های نقاط درونی استفاده می‌شود. در روش LIPSOL از ساختار داده‌های ماتریس تُنُک در محیط نرم‌افزاری MATLAB و فایل‌های واسط MEX در جهت ساده‌سازی برنامه و هم‌چنین کارآمد کردن هزینهٔ محاسباتی استفاده می‌شود *(Zhang, 1995)*.

۳-۲- الگوریتم اصلی روش حل مسئله برنامه‌ریزی خطی

در الگوریتم اصلی روش حل مسئلهٔ برنامه‌ریزی خطی، در ابتدا یک‌سری عملیات پیش‌پردازش بر روی متغیرها و قیود مسئله اعمال می‌شود، به‌طور مثال:

- محدودهٔ پایین تمامی متغیرها صفر در نظر گرفته می‌شود.
- تمامی قیود به‌صورت تساوی در نظر گرفته می شود.
- متغیرهای ثابتی که دارای محدودهٔ بالا و پایین یکسان هستند، حذف می‌شوند.
- سطرهای شامل تمام عناصر صفر از ماتریس قیود حذف می شوند.
- ستون‌های شامل تمام عناصر صفر از ماتریس قیود حذف می‌شوند.

پس از اعمال مرحلهٔ پیش‌پردازش‌های اولیه، مسئلهٔ برنامه‌ریزی خطّی به فرم زیر خواهد بود:

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \boldsymbol{f}^T\boldsymbol{x}$$
$$s.t.\begin{cases} A\boldsymbol{x} = \boldsymbol{b}, \\ 0 \le \boldsymbol{x} \le \boldsymbol{u} \end{cases} \qquad (۱۶)$$

با اضافه نمودن یک متغیر کمکی s می‌توان معادلهٔ (۱۶) را به‌صورت زیر بیان نمود:

$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{x}} \boldsymbol{f}^T\boldsymbol{x}$$
$$s.t.\begin{cases} A\boldsymbol{x} = \boldsymbol{b} \\ \boldsymbol{x} + \boldsymbol{s} = \boldsymbol{u} \\ \boldsymbol{x} \ge \boldsymbol{0}, \boldsymbol{s} \ge \boldsymbol{0}. \end{cases} \qquad (۱۷)$$

که به آن مسئلهٔ اولیه[6] می‌گویند. دوگان مسئلهٔ اولیه به‌صورت زیر مطرح می شود:

$$\hat{\boldsymbol{y}} = \arg \max_{\boldsymbol{y}} b^T\boldsymbol{y} - u^T\boldsymbol{w}$$
$$s.t.\begin{cases} A^T\boldsymbol{y} - \boldsymbol{w} + \boldsymbol{z} = \boldsymbol{f} \\ \boldsymbol{z} \ge 0, \boldsymbol{w} \ge 0. \end{cases} \qquad (۱۸)$$

[1] Linear Programming
[2] Convex polygon
[3] Linear-programming Interior Point Solvers
[4] Mehrotra's Predictor-Corrector
[5] Sparse LP solver based on interior point methods
[6] primal problem

که در آن y و w شامل متغیرهای دوگان و z نیز دربرگیرندهٔ متغییر کمکی دوگان است. سپس شرایط بهینه برای حل مسئله خطی شامل معادله اولیهٔ (۱۷) و معادلهٔ دوگان (۱۸) به‌صورت زیر تعریف می شود:

$$F(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{s},\boldsymbol{w}) = \begin{pmatrix} A\boldsymbol{x} - b \\ \boldsymbol{x} + \boldsymbol{s} - u \\ A^T\boldsymbol{y} - \boldsymbol{w} + \boldsymbol{z} - \boldsymbol{f} \\ \boldsymbol{x}_i\boldsymbol{z}_i \\ \boldsymbol{s}_i\boldsymbol{w}_i \end{pmatrix} = 0, \qquad (۱۹)$$

$$s.t.\ \boldsymbol{x},\boldsymbol{z},\boldsymbol{s},\boldsymbol{w} \geq 0.$$

که در آن x_iz_i و s_iw_i مؤلّفه‌های ضربی می‌باشند. معادله‌های درجهٔ دوم $x_iz_i=0$ و $s_iw_i=0$ شرایط مکمل کننده[1] برای مسئلهٔ خطی نامیده می‌شوند. معادلات خطی دیگر نیز شرایط امکان‌پذیری[2] نام دارند. مقدار x^Tz+s^Tw نیز فاصلهٔ دوگانی[3] نام دارد که مقدار باقیماندهٔ بخش مکمل F را درحالی‌که $(x,z,s,w)>=0$ هستند، اندازه‌گیری می‌کند. در این‌جا از یک الگوریتم primal-dual استفاده می‌شود که به‌طور هم‌زمان مسئلهٔ اولیه و دوگان آن را حل می‌نماید. این الگوریتم یک گونه از الگوریتم های پیش‌بین - تصحیح کننده[4] است. در هر تکرار الگوریتم ابتدا جهت پیش‌بینی به‌صورت زیر محاسبه می‌شود:

$$\Delta v_p = -(F(v))^{-1}F(v) \qquad (۲۰)$$

سپس جهت تصحیح کننده با معادلهٔ (۲۱) تعیین می شود.

$$\Delta v_c = -(F(v))^{-1}F(v + \Delta v_p) - \mu e \qquad (۲۱)$$

که در آن $\mu > 0$ پارامتر centering است و بایستی مقدار آن به دقت انتخاب شود. بردار e نیز یک بردار شامل عناصر صفر و یک است که متناظر با معادلات درجه دوم در F(v) می‌باشد. سپس با استفاده از هر دو مقدار جهت معرفی شده و پارامتر طول گام $\alpha > 0$، مقدار بردار v به روز می‌شود.

$$\begin{aligned} v^+ &= [x^+;y^+;z^+;s^+;w^+] \\ &= v + \alpha(\Delta v_p + \Delta v_c) \end{aligned} \qquad (۲۲)$$

در این‌جا پارامتر طول گام به‌گونه‌ای انتخاب می‌شود که شرط زیر همواره صادق باشد:

$$[x^+;z^+;s^+;w^+] > 0$$

تکرار الگوریتم تا زمان هم‌گرایی کامل متغیرها ادامه می‌یابد. یک معیار مناسب برای توقف می‌تواند استفاده از شرط زیر باشد:

$$\frac{\|r_b\|}{\max(1,\|b\|)} + \frac{\|r_f\|}{\max(1,\|f\|)} + \frac{\|r_u\|}{\max(1,\|u\|)} + \frac{\left\|f^Tx - b^Ty + u^Tw\right\|}{\max(1,\left\|f^Tx\right\|,\left\|b^Ty - u^Tw\right\|)} \leq Tol \qquad (۲۳)$$

که در آن Tol یک مقدار آستانه می‌باشد و دیگر پارامترهای آن به‌صورت زیر تعریف می‌شوند.

$$\begin{aligned} r_b &= Ax - b \\ r_f &= Ay - w + z - f \\ r_u &= x + s - u \end{aligned} \qquad (۲۴)$$

۴- معیارهای ارزیابی سیستم‌های کاوش‌گر کلمات

سیستم‌های کاوش‌گر کلمات به‌طور کامل از دو دیدگاه دقت در تشخیص و سرعت عملکرد مورد ارزیابی قرار می‌گیرند. سرعت و دقت این سیستم‌ها تحت تأثیر پارامترهای متنوعی است. به‌عنوان مثال خصوصیات زبانی و غیرزبانی کلیدواژه مورد نظر همانند میزان تکرار و خصوصیات نحوی و یا نوع ویژگی‌های آکوستیکی تشکیل دهندهٔ آن می‌توانند بر دقت و سرعت سیستم تأثیر بگذارند. به‌علاوه کیفیت گفتار ضبط شده نیز می‌تواند در این راستا تأثیرگذار باشد. در ادامه معیارهای متداول ارزیابی دقت عملکرد سیستم‌های کاوش‌گر کلمات معرفی می‌شوند *(Rohlicek, et al., 1989)*.

۱- نرخ هشدار نادرست[5] (FA): عبارت است از تعداد دفعاتی که یک کلمه در فایل گویهٔ مورد نظر به اشتباه به‌عنوان کلمهٔ کلیدی تشخیص داده شود. این کمیت بایستی به تعداد کلمات کلیدی و طول زمانی گفتار موردنظر نرمالیزه شود. ازاین‌رو معیار نرخ هشدار نادرست به‌ازای یک کلمهٔ کلیدی در یک ساعت به‌صورت زیر محاسبه می‌شود:

$$FA\ [1\ /\ kw\ /\ h] = \frac{N_{FA}}{T \times N} \qquad (۲۵)$$

که در آن T طول زمانی کل گفتار مورد جستجو برحسب ساعت (با این فرض که مقدار سکوت آن زیاد نباشد) و N_{FA} تعداد کلمات هشدار داده شده نادرست

[1] complementarity conditions
[2] feasibility conditions
[3] duality gap
[4] predictor-corrector algorithm
[5] False Alarm

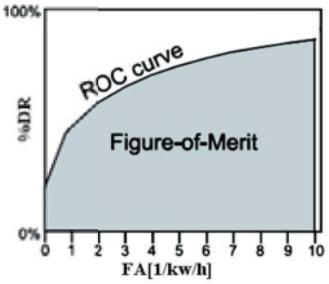
(کلیدواژه‌های به اشتباه کاندیدشده) و N تعداد کل کلیدواژه‌های تعریف شده در مجموعه لغات سیستم است.

۲- نرخ آشکارسازی درست[1] (DR): عبارت است از احتمال این‌که کلیدواژه در گفتار به‌درستی تشخیص داده شود و به‌صورت زیر تعریف می‌شود:

$$\%DR = \frac{N_{CD}}{N_{Tocc}} * 100 \qquad (۲۶)$$

که در رابطهٔ فوق N_{CD} تعداد کلیدواژه‌های به‌درستی تشخیص داده شده و N_{Tocc} تعداد کل وقوع کلیدواژه‌ها در مجموعه گفتار آزمون است (تعداد کل رخدادهای درست کلیدواژه‌ها). در برخی از منابع به این معیار Hit Rate نیز اطلاق می‌شود.

۳- منحنی ROC[2]: برای آن‌که تصویری مشخص‌تر از عملکرد سیستم کاوش‌گر کلمات به‌دست آید، در منحنی ROC نرخ آشکارسازی درست (DR) برحسب نرخ هشدار نادرست (FA) رسم می شود. هرچه مساحت سطح زیر این نمودار بیشتر باشد، عملکرد سیستم کاوش‌گر کلمه بهتر خواهد بود. در (شکل ۴) نمونه‌ای از منحنی ROC نشان داده شده است.



شکل ۴- نمایش منحنی ROC *(Szoke et al., 2010)*.

۴- معیار FOM[3]: این معیار برابر است با میانگین نرخ تشخیص درست کلمات، وقتی که معیار نرخ هشدار نادرست از مقدار ۱ تا ۱۰ تغییر کند که رابطهٔ آن به‌صورت زیر است:

$$FOM = \frac{1}{10}\sum_{fa=1}^{10} DR(fa) \qquad (۲۷)$$

در رابطهٔ فوق، برای به‌دست آوردن مقدار DR در مکان‌هایی که fa (مقدار نرخ هشدار نادرست که در رابطهٔ (۲۵) آمده است) به‌طور دقیق مساوی با ۱ تا ۱۰ باشد، می‌توان از درون‌یابی استفاده نمود. با توجّه به تعریف معیار FOM، این معیار به نوعی بیان‌گر سطح زیر منحنی ROC می‌باشد.

۵- دادگان مورد استفاده در آزمایش‌ها

دادگان مورد استفاده برای پیاده‌سازی و ارزیابی روش پیشنهادی، از مجموعه دادگان فارس‌دات تلفنی کوچک (TFarsDat) می‌باشد *(Bijankhan, et al., 2003)*. این مجموعه داده که توسط پژوهشکدهٔ پردازش هوشمند علائم (RCISP) تهیه شده است، شامل گفتار محاوره‌ای یک‌طرفه (شماره فایل‌های ۱و۲) و برخی از جملات و کلمات پرکاربرد (شماره فایل‌های ۳ تا ۱۴) از ۶۴ گوینده است که گفتار هریک از گویندگان به‌طور جداگانه از طریق یک کانال تلفنی ضبط شده است. بنابراین برای هر گوینده چهارده فایل گفتاری ضبط شده است. در پیاده‌سازی‌های این مقاله، از تمام فایل‌های گفتاری پنجاه گوینده اول برای تعلیم سیستم بازشناسی گفتار مبتنی بر HMM استفاده شده است. همچنین فایل‌های مربوط به جلسات محاوره‌ای (شماره فایل‌های ۱و۲) از چهارده گویندهٔ باقیمانده، برای بخش آزمایش سیستم درنظر گرفته شده است. مجموع طول زمانی فایل‌های آزمون مورد استفاده حدود هیجده دقیقه می‌باشد $(T = 0.303 \text{ hour})$.

تعداد ۴۶ کلمهٔ پرتکرار (کلمه با طول ۴ واج تا ۱۰ واج) از مجموعه کلمات دادگان مربوط به بخش آزمون نیز به‌عنوان کلیدواژه‌های سیستم کاوش‌گر کلمات انتخاب شدند $(N = 46)$. این کلمات از بین پرتکرارترین کلمات موجود در دادگان آزمون به‌گونه‌ای انتخاب شده‌اند که هر کدام بین ۵ تا ۲۵ مرتبه در مجموعه آزمون تکرار شده‌اند و هر کلمه حداقل چهار واج دارد. در مجموع نیز تعداد وقوع کلیدواژه‌ها در مجموعهٔ آزمون۴۰۲ مرتبه می‌باشد $(N_{Tocc} = 402)$. هم‌چنین برای هر کلمه سعی شده تا تمامی تنوعات تلفظی آن در مرحلهٔ بازشناسی اولیه در نظر گرفته شود.

۶- آزمایش‌ها و بررسی نتایج

سیستم بازشناس مورد استفاده در این مقاله مبتنی بر مدل مخفی مارکوف (HMM) بوده و با استفاده از جعبه ابزار HTK پیاده‌سازی شده است *(HTK, 2006)*. مدل بازشناس گفتار تلفنی تولید شده دارای واحدهای زیرکلمهٔ واجی با مدل‌های سه‌آوایی است و گره‌زدن حالات مدل‌های سه‌آوایی آن براساس درخت تصمیم می‌باشد. هر مدل واجی نیز با توجّه به واک‌دار و یا بی‌واک بودن آن به‌ترتیب شامل ۳ یا ۵

[1] Detection Rate
[2] Receiver Operating Characteristic
[3] Figure of Merit

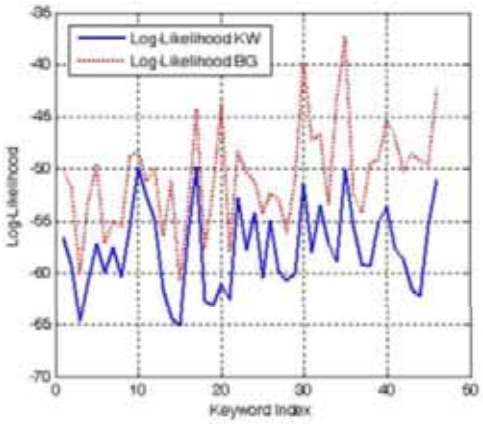
حالت[1] می‌باشد. توزیع احتمالاتی ویژگی‌ها در هر حالت نیز به‌وسیلهٔ مدل تلفیقی گوسی[2] (GMM) در نظر گرفته شده است. در این‌جا تعلیم GMM به‌گونه‌ای پیاده‌سازی شده است که درمجموع برای کل حالات مدل‌های HMM، به‌طور متوسط تعداد هشت گوسین در هر حالت قرار گرفته است. به این معنا که با توجّه به نحوهٔ توزیع چگالی احتمال در هر حالت، امکان دارد در برخی از حالات تعداد گوسین‌ها بیشتر و یا کم‌تر از تعداد هشت گوسین باشد (ر.ج. به دستور PS از تابع اجرایی HHED در صفحهٔ ۲۶۰ از HTKbook نسخهٔ 3.4).

ویژگی‌های اولیهٔ استخراج شده برای هر قاب گفتار تلفنی، شامل دوازده ضریب کپستروم به‌همراه ضریب صفرم آن با استفاده از روش استخراج ویژگی متداول MFCC است *(Davis, et al., 1980)*. هم‌چنین از مشتقات مرتبهٔ اوّل تا سوم ضرایب کپستروم به‌دست آمده نیز به‌عنوان مؤلّفه‌های دینامیک ویژگی‌های گفتاری استفاده شده است *(Furui,1986)*. بنابراین بردار ویژگی تولید شده دارای بُعد ۵۲ خواهد بود. محدودهٔ فرکانسی در نظر گرفته شده برای اعمال بانک‌فیلترها بر روی طیف سیگنال گفتار تلفنی نیز بین فرکانس یکصد هرتز تا چهار کیلوهرتز می‌باشد و تعداد چهارده فیلتربانک مثلثی برای این‌منظور انتخاب شده است. هم‌چنین در جهت مقاوم‌سازی بردار ویژگی به‌دست آمده نسبت به نویزهای جمع شونده (مانند صدای محیط) و اثر کانال انتقال تلفنی از روش پس‌پردازش MVA با درجهٔ فیلتر یک استفاده شده است *(Chen, et al., 2007)*.

پس از تعلیم مدل بازشناس گفتاری، دقت[3] بازشناسی واج معادل با ۵۴/۹۰ درصد بر روی مجموعه دادهٔ آزمون به‌دست آمده است. این در حالی است که دقت بازشناسی واج بر روی دادهٔ تعلیم حدود ۷۶/۲۳٪ می‌باشد.

۶-۱- بررسی امتیاز اطمینان

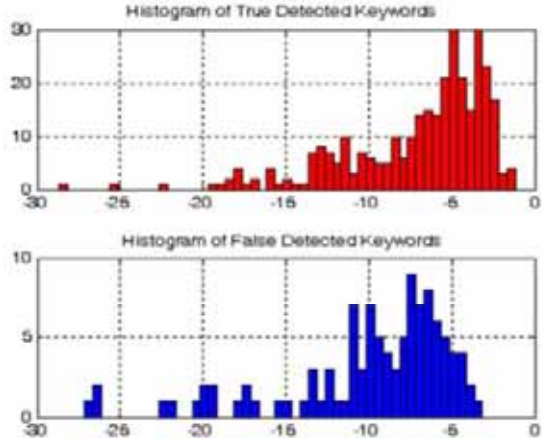
یکی از مهم‌ترین بخش‌ها در سیستم تشخیص کلیدواژه‌ها، محاسبه و ارزیابی امتیاز اطمینان کلیدواژه‌های کاندیدشده و تصمیم‌گیری برای رد یا قبول آن‌ها می‌باشد. در الگوریتم معرفی شده در بخش دوم، برای محاسبهٔ این امتیاز از تفریق مقدار لگاریتم درست‌نمایی کلیدواژهٔ بازشناسی شده از شبکهٔ کلمات و مقدار لگاریتم درست‌نمایی مدل پس‌زمینهٔ متناظر با آن استفاده شده است (رابطهٔ (۳)). (شکل ۶) منحنی‌های دو مقدار لگاریتم درست‌نمایی کلیدواژهٔ بازشناسی شده از شبکهٔ کلمات (LL-KW) و مدل پس‌زمینه (LL-BG) متناظر با آن را برای تعداد ۴۶ کلید واژهٔ متمایز در مجموعه آزمون نشان می‌دهد.



(شکل ۶): نمایش نمونه‌هایی از مقدار لگاریتم درست‌نمایی کلیدواژه کاندید شده (LL-KW) بازشناسی شده و مدل پرکنندهٔ (LL-BG) متناظر با آن برای ۴۶ کلیدواژه انتخاب شده.

همان‌طور که در بخش دوم مقاله بحث شد، مقدار امتیاز درست‌نمایی به‌دست آمده برای کلید واژه، همواره دارای مقدار کم‌تری نسبت به مقدار امتیاز درست‌نمایی به‌دست آمده برای مدل پس‌زمینه می‌باشد که این موضوع در (شکل ۶) به وضوح قابل مشاهده است.

در ادامه به بررسی توزیع امتیاز اطمینان مربوط به کلیدواژه‌های درست و یا غلط تشخیص داده شده خواهیم پرداخت. از این‌رو در (شکل ۷) هیستوگرام امتیاز اطمینان اولیهٔ به‌دست‌آمده از رابطهٔ (۳) برای تمامی کلیدواژه‌ها، به تفکیک تشخیص غلط و یا درست بودن آن‌ها نشان داده شده است.

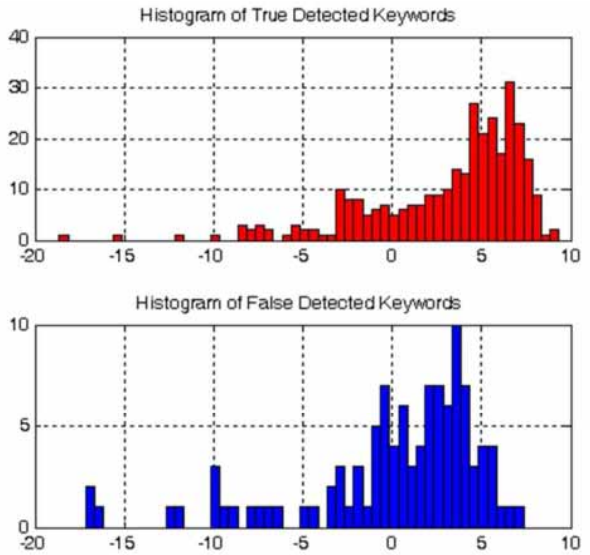


(شکل ۷): هیستوگرام امتیاز اطمینان اولیه کلیدواژه‌های درست (شکل بالا) یا غلط (شکل پایین)

[1] State
[2] Gaussian Mixture Model
[3] Accuracy

همان‌طور که از (شکل ۷) مشهود است و در بخش دو نیز توضیح داده شد، توزیع امتیازهای اطمینان کلیدواژه‌های غلط در مقادیر منفی‌تری نسبت به کلیدواژه‌های درست تشخیص داده شده قرار دارد.

پس از اعمال روش پیشهادی نرمالیزاسیون امتیاز، مقدار امتیاز اطمینان نهایی به‌دست آمده برای هر کلیدواژه تغییر خواهد یافت. این تغییر منجر به جابه‌جایی توزیع‌های نشان داده شده در (شکل ۷) خواهد شد. در (شکل ۸) این تغییرات به‌صورت هیستوگرام امتیاز اطمینان نرمالیزه شده (با استفاده از روش پیشنهادی در بخش سوم و رابطهٔ (۵)) برای هر کلیدواژه درست یا غلط تشخیص داده نشان داده شده است. در این حالت مقدار پارامتر H در رابطهٔ (۱۰) برابر با ۰/۷ در نظر گرفته شده است.

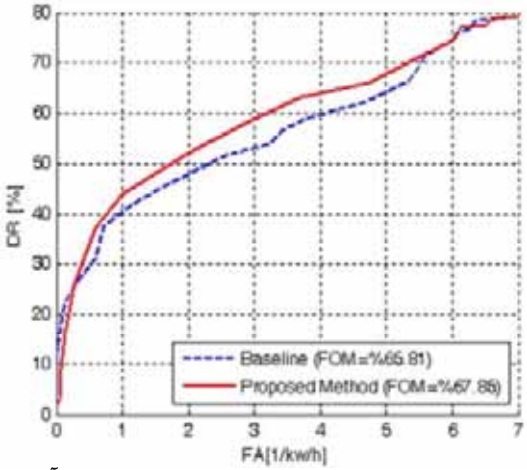


(شکل ۸): هیستوگرام امتیاز اطمینان نرمالیزه شدهٔ کلیدواژه‌های درست (شکل بالا) یا غلط (شکل پایین)

انتظار ما این است که استفاده از روش پیشنهادی نرمالیزاسیون امتیاز اطمینان، منجر به تفکیک‌پذیری بیشتر دو توزیع نشان داده شده در شکل (۷) شده باشد.

۶-۲- بررسی عملکرد روش نرمالیزاسیون امتیاز اطمینان پیشنهادی

در (شکل ۹) منحنی ROC و هم‌چنین نتایج معیار FOM به دست آمده برای سیستم کاوش‌گر کلمات پایه (بدون استفاده از روش نرمالیزاسیون امتیاز اطمینان) و هم‌چنین سیستم کاوش‌گر کلمه با استفاده از روش نرمالیزاسیون امتیاز اطمینان پیشنهاد شده، نشان داده شده است.



(شکل ۹): نمایش منحنی ROC و مقادیر FOM به دست آمده برای سیستم پایه (Baseline) و روش پیشنهاد شده (Proposed Method).

آن‌گونه که از (شکل ۹) پیداست، منحنی‌های مربوط به ROC فقط تا مقدار FA حدود ۷ رسم شده است. این بدان علت است که سیستم کاوش‌گر لغات مورد بررسی، دارای تعداد کلیدواژه‌های کاندید شده به‌نسبه کمی می‌باشد و از این‌رو حداکثر مقدار به‌دست آمده برای FA محدود شده است. در این حالت برای تخصیص مقدار DR در FA های ایجاد نشده در رابطهٔ (۲۷)، از بیشینهٔ مقدار DR محاسبه شده استفاده خواهد شد. با توجه به نتایج به دست آمده از (شکل ۹)، استفاده از روش پیشنهادی منجر به افزایش مقدار DR در اکثر FAها (به جز در مقادیر بسیار کم هشدار نادرست) نسبت به سیستم کاوش‌گر کلمات پایه شده است. از این‌رو بهبودی حدود دو درصد در مقدار FOM سیستم کاوش‌گر کلمات (از مقدار ۶۵/۸۱٪ به مقدار ۶۷/۸۵٪) با استفاده از روش پیشنهادی به‌دست آمده است.

۷- جمع‌بندی و نتیجه‌گیری

در این مقاله یک روش به‌طور کامل جدید برای نرمالیزاسیون امتیاز اطمینان محاسبه شده برای کلیدواژه‌های کاندید شده در سیستم کاوش‌گر کلمات پیشنهاد شده است. روش پیشنهادی برمبنای روش بهینه‌سازی برنامه‌ریزی خطی می‌باشد که هدف آن بیشینه‌نمودن تفکیک‌پذیری توزیع امتیاز اطمینان بین کلید واژه‌های درست و نادرست تشخیص داده شده است. نتایج به دست آمده نشان دهنده مفید بودن روش پیشنهادی در افزایش مقدار معیار FOM حدود ۲٪ در مجموعه آزمون سیستم کاوش‌گر کلمه است.

۸- مراجع

Bijankhan, M., Sheykhzadegan, J., Roohani, M.R., Zarrintare, R., Ghasemi, S.Z., Ghasedi, M.E., 2003. TFarsDat - The Telephone Farsi Speech Database. In Proc. EuroSpeech, Geneva, Switzerland, pp. 1525-1528.

Bridle, J.S., 1973. An Efficient Elastic-Template Method for Detecting Given Words in Running Speech. In Proc. British Acoustic Society Meeting, pp. 1-4.

Chen, C.P., Bilmes, J.A., 2007. MVA Processing Of Speech Features. IEEE Trans. Audio, Speech and Language Processing, vol. 15 (1), pp. 257-270.

Christiansen, R.W., Rushforth, C.K., 1977. Detecting and Locating Keywords in Continuous Speech Using Linear Predictive Coding. IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 25 (5), pp. 361–367.

Davis, S., Mermelstein, P., 1980. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Trans. Acoustics, Speech and Signal Processing. vol. 28 (4), pp. 357-366.

Furui, S., 1986. Speaker Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum. IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 34, pp. 52-59.

HTK, 2006. Hidden Markov ToolKit. Cambridge University Engineering Department, URL: http://htk.eng.cam.ac.uk

Manos, A., Victor, Z., 1997. A Segment-Based Word spotter Using Phonetic Filler Models. In Proc. ICASSP, Munich, Bavaria, Germany, vol. 2, pp. 899–902.

Mehrotra, S., 1992. On the Implementation of a Primal-Dual Interior Point Method. SIAM Journal on Optimization, vol. 2, pp. 575-601.

Meliani, R., OShaughnessy, D., 1997. Accurate Keyword Spotting Using Strictly Lexical Fillers. In Proc. ICASSP, Munich, Bavaria, Germany, pp. 907–910.

Myers, C.S., Rabiner, L.R., Rosenberg, A.E., 1980. An Investigation of the Use of Dynamic Time Warping for Word Spotting and Connected Speech Recognition. In Proc. ICASSP, Denver, Colorado, USA, pp. 173–177.

NIST, 2006. National Institute of Standards and Technology. The Spoken Term Detection (STD) 2006 Evaluation Plan. URL: http://www.nist.gov/speech/tests/std

Pinto, J., Szoke, I., Prasanna, S.R.M., Hermansky, H., 2008. Fast Approximate Spoken Term Detection from Sequence of Phonemes. In Proc. The 31st Annual International ACM SIGIR Conference, Singapore, pp. 28–33.

Rahim, M.G., Lee, C.H., Juang, B.H., 1995. Robust Utterance Verification for Connected Digits Recognition. In Proc. ICASSP, Detroit, Michigan, USA, vol. 1, pp. 285–288.

Rohlicek, W., Russell, S., Roukos, S., Gish, H., 1989. Continuous Hidden Markov Modeling for Speaker-Independent Word Spotting. In Proc. ICASSP, pp. 627-630.

Rose, R.C., Paul, D., 1990. A Hidden Markov Model Based Keyword Recognition System. In Proc. ICASSP, pp. 129-132.

Rose, R.C., 1995. Keyword Detection in Conversational Speech Utterances Using Hidden Markov Model Based Continuous Speech Recognition. Computer Speech and Language, vol. 9, pp. 309-333.

Szoke, I., 2010. Hybrid Word-Subword Spoken Term Detection. PhD Thesis, Brno University of Technology.

Szoke, I., Schwarz, P., Matejka, P., Burget, L., Karafiat, M., Cernocky, J., 2005. Phoneme Based Acoustics Keyword Spotting in Informal Continuous Speech. In Proc. Text, Speech and Dialogue (TSD), pp. 302–309.

Tejedor, J., Wang, D., Frankel, J., King, S., Colas, J., 2008. A Comparison of Grapheme and Phoneme-Based Units for Spanish Spoken Term Detection. Speech Communication, vol. 50, pp. 980–991.

Wilpon, J.G., Lee, C.H., Rabiner, L.R., 1989. Application of Hidden Markov Models for Recognition of a Limited Set of Words in Unconstrained Speech. Glasgow, UK, In Proc. ICASSP, pp. 254–257.

Wilpon, J.G., Rabiner, L.R., Lee, C.H., Goldman, E.R., 1990. Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models. IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 38(11), pp. 1870–1878.

Young S.J., Evermann G., Gales M.J.F., Hain T., Kershaw D., Moore G., Odell J.J., Ollason D., Povey D., Valtchev V., Woodland P.C., 2006. The HTK Book (for HTK Version 3.4). Cambridge University Engineering Department.

Zhang Y., 1995. Solving Large-Scale Linear Programs by Interior-Point Methods under the MATLAB Environment. Technical Report TR96-01, Department of Mathematics and Statistics, University of Maryland, Baltimore.

یاسر شکفته مدارک کارشناسی و کارشناسی ارشد خود را در رشتهٔ مهندسی پزشکی (گرایش بیوالکتریک) به‌ترتیب در سال‌های ۱۳۸۴ و ۱۳۸۷ از دانشگاه صنعتی امیرکبیر اخذ نموده است. وی از سال ۱۳۸۷ دانشجوی مقطع دکتری در رشتهٔ مهندسی پزشکی دانشگاه صنعتی امیرکبیر می‌باشد. زمینه‌های تحقیقاتی مورد علاقه ایشان شامل پردازش سیگنال‌های حیاتی، شناسایی الگو، مدل‌سازی سیستم‌های بیولوژیکی و سیستم‌های بازشناسی گفتار است.
نشانی رایانامک ایشان عبارت است از:
y_shekofteh@aut.ac.ir

جهانشاه کبودیان تحصیلات خود را در هر سه مقطع کارشناسی، کارشناسی ارشد و دکتری در دانشکدهٔ مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر (پلی‌تکنیک تهران) گذراند و مدرک دکترای خود را در سال ۱۳۸۹ از دانشگاه مذکور دریافت نمود. وی هم‌اکنون، استادیار گروه مهندسی کامپیوتر دانشکدهٔ فنی- مهندسی دانشگاه رازی کرمانشاه و هم‌چنین مدیر گروه پردازش صوت پژوهشکدهٔ پردازش هوشمند علائم (RCISP) می‌باشد. زمینه‌های تخصصی مورد علاقه ایشان شامل پردازش سیگنال، پردازش صوت، پردازش گفتار، شناسایی الگو، یادگیری ماشین و مدل پنهان مارکف می‌باشد.
نشانی رایانامک ایشان عبارت است از:
kabudian@{razi,rcisp,aut}.ac.ir

محمدمحسن گودرزی تحصیلات خود را در مقطع کارشناسی در دو رشتهٔ مهندسی پزشکی- بیوالکتریک و مهندسی برق- کنترل به‌ترتیب در سال‌های ۱۳۸۶ و ۱۳۸۸ در دانشگاه صنعتی امیرکبیر به پایان رساند. در سال ۱۳۸۹ در مقطع کارشناسی ارشد رشتهٔ مهندسی پزشکی- بیوالکتریک از همان دانشگاه فارغ التحصیل شد. وی هم اکنون در مقطع دکترای مهندسی پزشکی- بیوالکتریک در دانشگاه صنعتی امیرکبیر در حال تحصیل می‌باشد. زمینه‌های پژوهشی مورد علاقه ایشان شناسایی الگو، پردازش سیگنال‌های تصادفی و پردازش و بازشناسی گفتار می‌باشد.
نشانی رایانامک ایشان عبارت است از:
mm.goodarzi@aut.ac.ir

ایمان صراف رضایی مدارک کارشناسی و کارشناسی ارشد خود را در رشتهٔ مهندسی پزشکی (گرایش بیوالکتریک) به‌ترتیب در سال‌های ۱۳۸۴ و ۱۳۸۷ از دانشگاه صنعتی امیرکبیر اخذ نموده است. وی در حال حاضر عضو گروه پردازش صوت پژوهشکدهٔ پردازش هوشمند علائم می باشد. مدل‌سازی و پردازش سیگنال گفتار زمینهٔ عمومی تحقیقات ایشان است.
نشانی رایانامک ایشان عبارت است از:
imansarraf@aut.ac.ir