چهاردهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران
دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران) ایران، تهران – ۲۰ و ۲۱ اسفندماه ۱۳۸۷
Archive of SID

ارائه روشی برای تعیین معیار اطمینان در پذیرش کلمات کلیدی در زبان فارسی

مریم اسدی
آزمایشگاه سیستم های هوشمند صوتی گفتاری
دانشکده مهندسی کامپیوتر و فن‌آوری اطلاعات، دانشگاه صنعتی امیرکبیر
Ma_asadi@aut.ac.ir

محمد مهدی همایون‌پور
آزمایشگاه سیستم های هوشمند صوتی گفتاری
دانشکده مهندسی کامپیوتر و فن‌آوری اطلاعات، دانشگاه صنعتی امیرکبیر
homayoun@aut.ac.ir

سید محمد حسین معطر
آزمایشگاه سیستم های هوشمند صوتی گفتاری
دانشکده مهندسی کامپیوتر و فن‌آوری اطلاعات، دانشگاه صنعتی امیرکبیر
moattar@aut.ac.ir

چکیده: تشخیص کلمات‌کلیدی، یک شاخه مهم از بازشناسی اتوماتیک گفتار با شناسایی تعداد محدودی از کلمات‌کلیدی در یک گفتار است. این مقاله یک روش جدید محاسبه درجه اطمینان برای پذیرش/رد کلمه کلیدی شناسایی شده، ارائه می‌کند. در روش پیشنهادی از مدل مخفی مارکوف برای آموزش مدل واحدهای بازشناسی استفاده می‌شود. مدلهای بازشناسی عبارتند از مدلهای واجی زبان فارسی و نیز مدلهای کلمات‌کلیدی. تشخیص کلمات‌کلیدی در دو مرحله انجام می‌شود. در مرحله اول یکسری کلمات‌کلیدی شناسایی می‌شوند و در مرحله دوم یک معیار اطمینان برای رد یا قبول هر یک از کلمات‌کلیدی تشخیص داده شده در مرحله قبل، محاسبه و استفاده می‌شود. در مرحله اول از مدلهای مخفی مارکوف واجی و در مرحله دوم از مدلهای مخفی مارکوف کلمات‌کلیدی همراه با سطوح آستانه تصمیم گیری استفاده می‌شود. آزمایش‌های انجام شده برای تشخیص ۲۵ کلمه کلیدی فارسی بیانگر کارایی خوب روش پیشنهادی و بیشتر بودن کارایی آن به میزان ۶.۵٪ نسبت به روشی است که از معیار اطمینان پیشنهادی استفاده نشود.

واژه های کلیدی: شناسایی گفتار، تشخیص کلمات‌کلیدی، معیاراطمینان، مدل مخفی مارکوف.

۱- مقدمه

در مقابل جستجو در متن، جستجو در گفتار یا به طور کلی جستجو در داده‌های چندرسانه‌ای از جایگاه ویژه‌ای برخوردار است. به منظور جستجو در داده‌های گفتاری نیازمند انجام بازشناسی گفتار هستیم. شاخه‌ای از بازشناسی گفتار بازشناسی کلمات‌کلیدی است[۲و۳].

هدف از بازشناسی کلمات‌کلیدی، آشکارسازی مجموعه‌ای محدود از کلماتی خاص در گفتار پیوسته ورودی است که این کلمات معمولا حامل مفاهیم اصلی گفتار هستند. همزمان با افزایش استفاده از بازشناسی گفتار در کاربردهای مختلف مانند پاسخگوئی تلفنی، محاوره انسان و ماشین و مانند آن، نیازمند سیستم‌هایی هستیم که به اندازه کافی با دامنه وسیعی از رفتارهای کاربران مثل صداهای تاکیدی، مکث‌ها، تکیه‌ها و فشارها روی یک کلمه خاص انعطاف‌پذیری لازم را در بازشناسی گفتار کاربران داشته باشند. مشکل اصلی در شناسایی گفتار این است که دقیقا نمی‌دانیم کاربر چه می‌خواهد بگوید[۴]. در بسیاری مواقع ما اطلاعات خوبی از کلمات کلیدی مورد نیاز برای یک سیستم گفتاری در اختیار داریم. از طرفی مایلیم که کاربر بتواند سوالات خود را بصورت عادی و در قالب گفتار پیوسته بیان نماید و مجبور به بکار بردن کلمات گسسته و یا گفتار پیوسته در قالبی خاص نباشد. با بکارگیری یک سیستم تشخیص کلمات کلیدی در کنار سیستم‌های بازشناسی گفتار می‌توان مشکلات فوق را تا حد زیادی برطرف نمود[2-4]. همچنین امکان بیان گفتار پیوسته و بدون قید و شرط و با سهولت بیشتر را برای کاربران فراهم آورد.

یکی از مشکلات مهم در تشخیص کلمات‌کلیدی، شناسایی کلمات غیرکلیدی است[۱]. کلمات غیرکلیدی می‌توانند یکسری کلمات مشخص که در یک واژگان گردآوری شده است، یا هر کلمه ادا شده (عبارات زباله[1] یا فیلر[2]) خارج از واژگان (OOV[3]) باشند. یک مدل فیلر (مدل گفتار غیرکلیدی) به منظور فیلتر کردن کلمات خارج از واژگان بکار می‌رود. بدین ترتیب سیستم بازشناسی قادر به فیلتر نمودن کلمات خارج از واژگان در گویش گفتار کاربر بوده و فقط اطلاعات مفید (کلمات‌کلیدی) مربوط به درخواست مورد نظر کاربر را استخراج می‌نماید[۵].

تا آن‌جا که می‌دانیم، هیچ مدل فیلر بهینه جامع که بتواند برای هر تشخیص‌دهنده اتوماتیک گفتار (ASR[4]) در زبان طبیعی بکار رود، وجود ندارد. محققان زیادی به ساخت مدل‌های قوی و قابل اطمینان توجه داشته‌اند ولی در استفاده از مدل‌های گفتار غیرکلیدی، فضای جستجوی شناسایی‌کننده گفتار و زمان اجرا بسیار بزرگ می‌شود[۴]. در بسیاری از کارهای جدید بر روی روش‌هایی تمرکز شده است که بتوانند در کنار یک مدل ساده از کلمات غیرکلیدی و با تکیه بر درصد درستنمایی[5] و

[1] Garbage
[2] Filler
3 Out-Of-Vocabulary
[4] automatic speech recognizer
[5] Likelihood

www.SID.ir

Archive of SID

معیاراطمینان چند کاندید بهتر، اطلاعات مفید را از گویشی که حاوی کلمات‌کلیدی و گفتار غیرکلیدی است، استخراج نمایند[۵]. معیاراطمینان (CM[۶]) عددی بین ۰ و ۱ است که در خروجی پردازش گفتار بکار می‌رود. یک معیار اطمینان، شاخصی است که بیانگر میزان اطمینان ما از صحت تشخیص گویش مورد نظر (مثلا یک عبارت، یک کلمه، یا یک آوا) است. در واقع می‌بایست درجه اطمینان کمتری به خروجی یک تشخیص‌دهنده با تعداد زیادی کلمه خارج از واژگان یا اصوات مبهم، اصوات نویزی و یا همراه با پس‌زمینه بلند نسبت داد. کلمات خارج از واژگان و اصوات مبهم، هر دو از منابع مهم خطای تشخیص هستند که با بکار بردن معیار اطمینان به عنوان یک شاخص در تست فرضیه‌های آماری می‌توان این خطاها را کاهش داد. درجه اطمینان می‌تواند یکی از درجه‌های اطمینان حاصل از منابع اطلاعاتی اطمینان نظیر میزان درستنمایی، نرخ میزان درستنمایی، استمرار، نرخ استمرار، احتمال‌های مدل زبانی و غیره باشد[۳]. هدف از این مقاله ارائه روشی به عنوان معیار اطمینان برای رد یا قبول کلمات‌کلیدی است. ابتدا به وسیله مدل مخفی مارکوف (HMM[۷]) مدل‌های واجی را آموزش می‌دهیم. همچنین در واژگان مورد استفاده برای بازشناسی، تک‌تک واجها و نیز دنباله واجی معادل هر یک از کلمات‌کلیدی را تعریف می‌نماییم. آنگاه گرامر مورد استفاده را به گونه‌ای ارائه می‌کنیم که نتیجه بازشناسی بتواند حاوی هر یک از کلمات‌کلیدی و نیز واجها (به عنوان مدلهای زباله) باشد. سپس فایل تست را با استفاده از مدلهای واجی و به کمک الگوریتم ویتربی مورد بازشناسی قرار داده و دنباله واجی متناظر با فایل تست ورودی را مشخص می‌کنیم. در اینکار چنانچه دنباله‌ای از واجهای بازشناسی شده با یکی از کلمات‌کلیدی متناظر باشد، کلمه‌کلیدی شناسایی و محدوده متناظر با آن کلمه‌کلیدی در فایل گفتاری ورودی مشخص می‌شود. آنگاه بردارهای ویژگی محدوده گفتاری بدست آمده به عنوان یک کلمه‌کلیدی را با مدل مخفی مارکوف آن کلمه‌کلیدی که از قبل آموزش داده‌ایم، مقایسه و مقدار شباهت آنها را بدست می‌آوریم. در انتها مقدار شباهت بدست آمده را با یک سطح آستانه از پیش تعیین شده مقایسه و نسبت به قبول یا رد آن کلمه کلیدی اقدام می‌کنیم. در اینجا مقایسه با مدل کلمه‌ای کلمات‌کلیدی و استفاده از سطح آستانه تصمیم‌گیری، بعنوان روشی برای اعمال معیار اطمینان در تشخیص کلمات کلیدی بکار گرفته شده است.

مطالب باقی‌مانده در این مقاله به شرح زیر می‌باشد: در قسمت ۲ تکنیک‌های تشخیص کلمات‌کلیدی توضیح داده خواهند شد. در قسمت ۳ دادگان مورد استفاده معرفی می‌شود. قسمت ۴ به تشریح پیاده‌سازی‌ها و نتایج آزمایش‌ها اختصاص دارد و در آخر بخش ۵ به نتیجه‌گیری از این مقاله می‌پردازد.

۲. تکنیک های تشخیص کلمات کلیدی

در بازشناسی اتوماتیک گفتار از روش‌های مبتنی بر مدل مخفی مارکوف بطور گسترده‌ای استفاده شده است[8-6]. یک سیستم تشخیص کلمات کلیدی مبتنی بر HMM شامل دو فاز عمده است: تشخیص‌دهنده (شناسایی کننده) و تصدیق کننده برای رد یا قبول کلمات کلیدی تشخیص داده شده[۲]. بهبود هر یک از این دو فاز منجر به افزایش کارایی سیستم می‌شود. روشهایی که سعی در بهبود فاز تشخیص‌دهنده دارند، بر روی یک مدل فیلر که بیانگر کلمات غیرکلیدی است، تمرکز دارند[۹]. معمولا در فاز تصدیق، پذیرش یا عدم پذیرش کلمات خارج از واژگان با محاسبه یک معیار اطمینان و مقایسه آن با یک حد آستانه تصمیم‌گیری انجام می‌شود[۵].

مشکل تشخیص کلمات کلیدی با سه روش عمده قابل حل است[۳]. مهم‌ترین و معروف‌ترین روش، استفاده از یک سیستم بازشناسایی گفتار پیوسته با واژگان بزرگ(LVCSR[۸])، برای تولید یک رشته کلمه و سپس جستجو برای کلمه کلیدی در این رشته از کلمات است. داده‌های آموزشی به دو دسته داده‌های کلیدی و غیرکلیدی تقسیم می‌شوند. کلمات کلیدی با مدلهای HMM که با گفتار کلمات کلیدی آموزش داده شده‌اند و مدل‌های گفتار غیرکلیدی با گفتار کلمات غیرکلیدی آموزش داده می‌شوند. عیب اصلی این روش ها نیاز به یک لغت نامه بزرگ از کلمات غیرکلیدی است که در نتیجه آن، مشکلاتی در رابطه با کلمات خارج از واژگان، شروع‌های غلط، تردیدها، تکرار و بی‌نظمی‌های دیگر به وجود می‌آید.

روش دوم بر پایه تحلیل خروجی بازشناسی کننده آواهای مبتنی بر اطلاعات اکوستیکی است.

روش سوم، استفاده از یک مدل فیلر برای کلمات خارج از لغت‌نامه است که در آنها از مدل‌های آوایی به عنوان مدل کلیه کلمات غیرکلیدی استفاده می‌گردد. در این روش یک حد آستانه پذیرش با درصد درستنمایی به منظور تشخیص کلمه‌کلیدی بکار می‌رود. جهت بدست آوردن فاکتور نرمال‌سازی، میزان درستنمایی مدل فیلر همزمان با میزان درستنمایی کلمه کلیدی بدست می‌آید. این نرمال‌سازی، الگوریتم را از نحوه ترکیب آوایی کلمات کلیدی مستقل می‌کند. به عبارتی برای اینکه بفهمیم آیا یک کلمه‌کلیدی در صحبت گفته شده وجود دارد یا خیر، امتیاز نرمال شده از کلمه‌کلیدی (KWNS[۹])، با یک حد آستانه تعریف شده، مقایسه می‌شود[4-2]. روشهای ارائه شده برای معیاراطمینان را می توان به سه دسته کلی تقسیم نمود[10-9]:

[6] confidence measure
[7] Hidden Markov Model
[8] large vocabulary continuous speech recognition
[9] Keyword Normalized Score

www.SID.ir

Archive of SID

الف- در بسیاری از کارها برای محاسبه اندازه‌های اطمینان، بر مبنای ترکیب ویژگی‌های پیشگویی‌کننده که در طول فرآیند رمزگشایی اکوستیکی (تشخیص دنباله واحدهای گفتاری) جمع‌آوری شده‌اند عمل می‌شود. این ویژگی‌ها می‌تواند شامل اصوات و نیز اطلاعات زبان مورد تصمیم باشند. در این شیوه همه ویژگی‌های پیشگویی‌کننده با یک روش قطعی با هم ترکیب می‌شود تا یک امتیاز واحد برای مشخص کردن صحت تصمیم تشخیص، تولید گردد[۱۰].

ب- قانون تصمیم‌گیری حداکثر احتمال پسین برای یافتن شبیه‌ترین دنبالهٔ کلمات. مسلم است که احتمال پسین در قانون تصمیم‌گیری حداکثر احتمال پسین (MAP[10]) استاندارد، یک کاندید خوب برای سنجش اطمینان در تشخیص گفتار است، زیرا یک معیار مستقل از میزان خوب بودن تصمیم است[۵]. اما تخمین احتمال پسین با یک شیوه دقیق بسیار مشکل است. در عمل، روش‌های مختلفی برای تخمین آن پیشنهاد شده است که شامل روشهای مبتنی بر فیلر ساده تا روش‌های پیچیده مبتنی بر گراف کلمه می‌گردند[۱].

ج- تحقیقات زیادی درجهت بازبینی محتویات مورد ادعا در یک گویش گفته شده، تحت عنوان تصدیق گویش (UV[11])، انجام شده است. محتویات می‌تواند با یک تشخیص دهنده گفتار یا آشکار کننده کلمه‌کلیدی که توسط اپراتور انسانی آوانویسی شده است، تشخیص داده شود. سنجش اطمینان، تحت چهارچوب تصدیق گویش، می‌تواند بعنوان یک تست فرضیه آماری، فرموله شود[7-8].

هدف اصلی در این مقاله، ارائه روشی برای تخمین دقیق‌تر معیار اطمینان به منظور یافتن راهی برای استخراج اطلاعات مفید از گویشی که حاوی کلمات کلیدی و گفتار غیرکلیدی است، می‌باشد. بدین ترتیب می‌خواهیم یک سیستم تشخیص کلمات‌کلیدی مستقل از لغت‌نامه کلمات غیرکلیدی طراحی نماییم.

۳. معرفی دادگان

دادگان مورد استفاده در این مقاله، TPersianData می‌باشد که در آزمایشگاه سیستم‌های هوشمند صوتی-گفتاری دانشکده مهندسی کامپیوتر دانشگاه صنعتی امیرکبیر ضبط شده‌است و در سطوح آوایی، تقطیع و برچسب‌گذاری شده است.

سیگنال گفتار با نرخ نمونه برداری ۸ کیلو هرتز و با فرمت A-Law نمونه‌برداری شده و سپس به فرم خطی تبدیل و هر نمونه توسط ۱۶ بیت بیان شده است. ویژگی‌های مورد استفاده، ۱۲ ضریب کپسترال مبتنی بر معیار مل MFCC و مشتقات اول و دوم و نیز انرژی فریم‌هایی با طول زمانی 25ms و میزان جابجایی 10ms، شامل برداری با ابعاد ۵۲ بعد می‌باشد. در این سیستم مدلهای مخفی مارکوف بصورت مستقل از گوینده آموزش داده شده‌اند.

داده‌های مورد استفاده از دادگان فوق شامل گفتار تلفنی ۱۱۵ گوینده مختلف است که از هر گوینده ۲ فایل (مجموعا ۲۳۰ فایل صوتی) که در جلسات مختلف ضبط شده‌اند، وجود دارد. از این مجموعه تعداد ۱۸۰ فایل برای آموزش و ۵۰ فایل برای تست استفاده شده است. برای تنوع بیشتر داده‌های آموزشی، از ۴۰ فایل گفتار دادگان تلفنی فارس‌دات تلفنی علاوه بر داده‌های دادگان TPersianDat، برای آموزش مدل‌های واجها استفاده شده است.

واژگان مورد استفاده در این تحقیق شامل ۲۵ کلمه کلیدی است. واحد گفتاری پایه مورد استفاده در بازشناسی نیز واج انتخاب شده است. هر واج با یک مدل HMM سه حالته چپ به راست نمایش داده شده است. هر حالت مدل از مجموعه‌ای از توابع گوسی با تابع چگالی احتمال پیوسته تشکیل می‌گردد. ماتریس کوواریانس توابع گوسی قطری در نظر گرفته شده است. تعداد توابع گوسی مناسب برای حالتهای مدل بصورت تجربی بدست آمده‌اند. مجموعه‌ای از مدل‌های HMM واجی آموزش داده شده که از آنها برای تعیین دنباله واجها جهت تشخیص کلمات‌کلیدی و همچنین برای شناسایی دنباله‌هایی بی‌معنی به عنوان کلمات غیرکلیدی استفاده شده است.

۴. پیاده سازی و نتایج آزمایش‌ها

در این مقاله ترکیبی از مدل واجی و مدل کلمه‌ای به منظور ارائه روشی جدید در تعیین معیار اطمینان برای پذیرش کلمات‌کلیدی در زبان فارسی استفاده شده است. برای مدل‌های کلمات غیرکلیدی از واحدهای گفتار تک‌آوایی[12] و برای کلمات کلیدی از هر دو مدل آوایی و کلمه‌ای استفاده می‌شود. سیستم پیشنهادی از دو قسمت زیر تشکیل شده است:

- ماژول شناسایی کلمات کلیدی
- ماژول سنجش اطمینان

روش کار به این صورت است که ابتدا مدل‌های واجی را آموزش می‌دهیم. همزمان برای هر یک از کلمات کلیدی یک مدل بر مبنای مدل کلمه‌ای می‌سازیم. این مدل کلمه، یک مدل جانبی است که برای تعیین درجه اطمینان بکار می‌رود. سپس گرامر مورد نیاز را بر اساس مدل‌های آوایی تعریف می‌کنیم. در این گرامر کلمات غیرکلیدی بعنوان یک توالی از واجهای جدا از هم و هر کلمه کلیدی نیز با یک توالی از واجها که کلمه‌کلیدی را تشکیل می‌دهند، ارائه می‌گردد. بنابراین کل ساختار HMM، یک شبکه موازی از مدل کلمات کلیدی (ترکیبی از

[10] Maximum A Posterior Probability
[11] Utterance Verification
[12] monoPhone

www.SID.ir

Archive of SID

مدل‌های واجی) و مدل واجهای جدا از هم (مدل‌های غیرکلیدی) است. در مرحله بعد، با استفاده از مدل‌های واجی و گرامر تعریف شده، داده‌های گفتاری تست مورد نظر را پردازش می‌کنیم. شکل ۲، شمایی کلی از سیستم بازشناسی را نشان می‌دهد. همانطور که در شکل می‌بینیم مراحل کار بطور خلاصه به شرح زیر است:

- ابتدا به وسیله مدل مخفی مارکوف (HMM[13]) مدل‌های واجی را آموزش می‌دهیم.
- در لغت‌نامه مورد استفاده در بازشناسی، تک تک واجها و نیز دنباله واجی معادل هر یک از کلمات‌کلیدی را تعریف می‌کنیم و گرامر مورد استفاده را به گونه‌ای تعریف می‌کنیم که نتیجه بازشناسی بتواند حاوی هر یک از کلمات کلیدی و نیز واجها (به عنوان مدلهای زباله) باشد.
- فایل تست را با استفاده از مدلهای واجی و به کمک الگوریتم ویتربی مورد بازشناسی قرار داده و دنباله واجی متناظر با فایل تست ورودی را مشخص می‌کنیم. در اینکار چنانچه دنباله‌ای از واجهای بازشناسی شده با یکی از کلمات کلیدی متناظر باشد، آن کلمه کلیدی شناسایی و محدوده متناظر با آن کلمه کلیدی در فایل گفتاری ورودی مشخص می‌شود.
- بردارهای ویژگی محدوده گفتاری که به عنوان یک کلمه کلیدی بدست آمده را با مدل مخفی مارکوف آن کلمه کلیدی که از قبل آموزش داده‌ایم، دوباره مورد شناسایی قرار می‌دهیم. درستنمایی بدست آمده را با یک سطح آستانه از پیش تعیین شده، مقایسه و نسبت به قبول یا رد آن کلمه کلیدی اقدام می‌کنیم.

مقایسه با مدل کلمه‌ای کلمات کلیدی و استفاده از سطح آستانه تصمیم‌گیری، بعنوان روشی برای اعمال معیار اطمینان در تشخیص کلمات کلیدی بکار گرفته شده است.

برای ارزیابی کارایی روش از دو نرخ ارزیابی استفاده شده است:

نرخ پذیرش غلط (FAR[14]) که به صورت زیر محاسبه می‌شود:

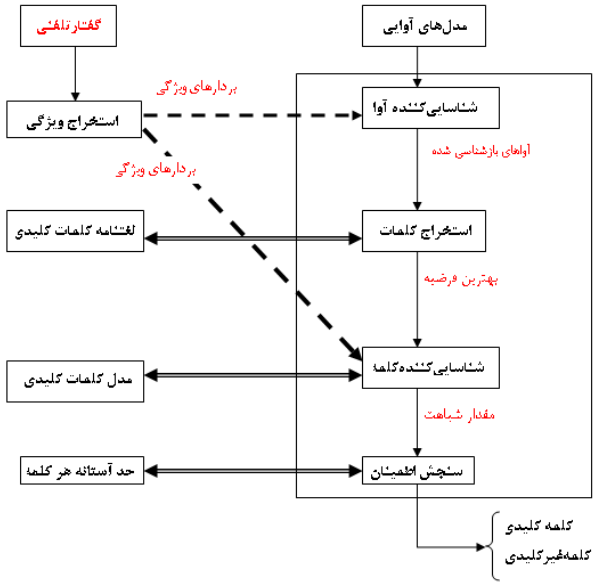
$$FAR = \frac{Total_False_Accep\,tance}{Total_False_Attemts} \qquad (۱)$$

نرخ رد اشتباه (FRR[15]) که به صورت زیر تعریف می‌شود:

$$FRR = \frac{Tota_False_Re\,jection}{Total_True_Attempts} \qquad (۲)$$

با استفاده از این دو مقدار، نرخ تشخیص(DR[16]) را محاسبه می‌کنیم:

$$DR = 100 - \frac{FAR + FRR}{2} \qquad (۳)$$



شکل۴: ساختار کلی سیستم بازشناسی پیشنهادی

در جدول ۱، نتایج حاصل از انجام آزمایش‌ها با استفاده مدل کلمه‌کلیدی و نیز بدون استفاده از مدل کلمه‌کلیدی مضاعف آورده شده است. آزمایش‌ها بروی دادگان TpersianData و برای ۲۵ کلمه کلیدی می‌باشد (توضیح دادگان در بخش ۳ آمده است).

جدول۲- نتایج حاصل از بازشناسی ۲۵ کلمه کلیدی

روش مورد استفاده	نرخ رد اشتباه	نرخ پذیرش اشتباه	دقت
بدون استفاده از مدل کلمه مضاعف	٪۳۸.۷	٪۲۵.۳	٪۶۸
با استفاده از مدل کلمه مضاعف	٪۳۴	٪۱۷.۴	٪۷۴.۳

با توجه به جدول مشاهده می‌شود که استفاده از معیار اطمینان ارائه شده در این مقاله به میزان ٪۶.۵ در تشخیص کلمات کلیدی بهبود ایجاد کرده است. همانطور که قبلا گفته شد، در روش پیشنهادی از هر دو مدل کلمه‌ای و نیز واجی استفاده شده است. از آنجا که گفتار یکبار با مدل واجی تحلیل و بررسی می‌گردد و پس از آن دوباره توسط مدلهای کلمه‌ای مورد ارزیابی مجدد قرار می‌گیرد، لذا انتظار می‌رود که این افزایش کارایی منحصر به دادگان خاص نبوده و با آزمایش و تست روش بر روی دادگان گفتاری دیگر نیز افزایش کارایی را داشته باشیم.

[13] Hidden Markov Model
[14] False Acceptance Rate
[15] False Rejection Rate
[16] Detection Rate

www.SID.ir

Archive of SID

۵. نتیجه گیری

در این مقاله، یک تکنیک تشخیص کلمات کلیدی جدید دو مرحله‌ای ارائه گردید. در روش جدید، درجه‌های معیار اطمینان بر پایه ترکیب مدل واجی و مدل کلمه‌ای محاسبه می‌شوند. همچنین نشان دادیم که در روش پیشنهادی، تعامل بهتری بین هشدارهای درست و خطا وجود دارد. کارایی تکنیک پیشنهادی در کاربردهای تشخیص دستور تلفنی فارسی، تست شده است و به کارایی خوبی برای این منظور رسیدیم. با استفاده از داده‌های تست گفتار تلفنی و لغت نامه‌ای حاوی ۲۵ کلمه کلیدی، نتایج آزمایش‌ها نشان می‌دهند که روش پیشنهادی نرخ خطای کمتری در مقایسه با روش یک مرحله‌ای دارد و در محاسبه درجه اطمینان از اطلاعات بیشتری استفاده نموده و نتایج بهتری ارائه می‌کند. در مطالعات آتی، یک روش مدلسازی فیلر بر اساس مدلسازی گروه‌های واجی برای دستیابی به یک مدل فیلر عمومی بهینه برای شناسایی کلمات کلیدی را بررسی خواهیم نمود. همچنین استفاده از تکنیک دو مرحله‌ای در فیلتر نمودن گفتار غیرکلیدی با روشی دیگر مورد بررسی قرار می‌گیرد.

قدردانی

این مقاله مورد حمایت مرکز تحقیقات مخابرات ایران قرار گرفته است.

مراجع

[۱] سامان ویسی‌پور، **ایجاد توانایی تشخیص کلمات غیرکلیدی به منظور بهبود سیستم بازشناسی گفتار فارسی**، پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی شریف، تهران، ۱۳۸۵.

[2] H. Shilei, X. Xiang, K. Jingming, "Comparison of Methods for Searching in Speech" INTERSPEECH 2006, 2006.

[3] P. Heracleous, T. Shimizu "A novel approach for modeling non-keyword intervals in a keyword spotter exploiting acoustic similarities of languages" Speech Communications 45, pp. 373–386, 2005.

[4] M. K. Abida "Fuzzy GMM-based Confidence Measure Towards Keyword Spotting Application" A thesis presented to the University of Waterloo in fulfilment of the thesis requirement for the degree of Master of Applied Science in Electrical and Computer Engineering Waterloo, Ontario, Canada, 2007

[5] H. Jiang, "Confidence measures for speech recognition: A survey", Speech Communication, Vol. 45, pp. 455-470, 2004.

[6] R. Rose and, D. Paul, "A Hidden Markov Model Based Keyword Recognition System", IEEE ICASSP, pp.129-132, 1990.

[7] H. Ketabdar, J. Vepa, S. Bengio and H. Bourlard "Posterior Based Keyword Spotting with A Priori Thresholds" INTERSPEECH 2006 , 2006.

[8] Frankel, David S., Model Driven Architecture: Applying MDA to Enterprise Computing, OMG Press, Wiley Publishing, "Improving the Performance of Out-of-vocabulary Word Rejection by Using SVM" INTERSPEECH 2006, 2006.

[9] G. Williams, and S. Renals, "Confidence measures for hybrid HMM/ANN speech recognition", Proceedings of Eurospeech' 97, pp. 1955-1958, 1997.

[10] H., Ketabdar, J. Vepa, S. Bengio and, H. Bourlard, "Developing and enhancing posterior based speech recognition systems", Interspeech'05, Lisbon, Portugal, 2005.

www.SID.ir