سیزدهمین کنفرانس ملی انجمن کامپیوتر ایران
جزیره کیش، خلیج فارس، ایران
۱۹ الی ۲۱ اسفند ۱۳۸۶
مقاله کامل
انجمن کامپیوتر ایران
Computer Society of Iran
دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

ارائه روشی برای بکارگیری ترکیبی مدل زبانی پیش بین مونوگرام و ترایگرام در بازشناسی گفتار پیوسته با واژگان بزرگ

مهدی صوفی‌فر
دانشجوی کارشناسی ارشد
آزمایشگاه پردازش گفتار
دانشگاه صنعتی شریف
soufifar@ce.sharif.edu

دکتر حسین ثامتی
استادیار دانشگاه صنعتی شریف
sameti@sharif.edu

باقر باباعلی
دانشجوی دکترا
آزمایشگاه پردازش گفتار
دانشگاه صنعتی شریف
sameti@sharif.edu

خسرو حسین زاده
کارشناس ارشد
آزمایشگاه پردازش گفتار
دانشگاه صنعتی شریف
hosseinzadeh@ce.sharif.edu

چکیده: در این مقاله، یک روش محاسباتی و تقریبی جدید برای مدل زبانی پیش بین[1] ارائه شده است. مهمترین اثر مدل زبانی پیش بین هرس تیز تر فضای جستجو با تقریب زدن امتیاز مدل زبانی لغت تحت بازشناسی و استفاده هرچه سریع تر آن در پروسه دیکد می باشد. به هرحال مدل زبانی پیش بین دارای سربارهای محاسباتی و حافظه ای مخصوص خود است که با بزرگ شدن واژگان و بالاتر رفتن مرتبه مدل زبانی مورد استفاده، به شکل نا مطلوبی افزایش می یابد و روند بازشناسی را تحت تاثیر قرار می دهد. روش ارائه شده برای به طور مشخص به این موضوع می پردازد و سعی دارد با استفاده از یک روش اشتراک گیری در بین داده های موجود در گره های درخت واژگان، علاوه بر فراهم کردن امکان استفاده از مدلهای زبانی ترایگرام در هرس مدل زبانی پیش بین، تعداد عملیات محاسبه احتمال مدل زبانی را نیز کاهش دهد به نحوی که در نهایت هم سرعت بازشناسی افزایش یابد و هم دقت بازشناسی بهبود یابد. در این مقاله روش پیشنهادی بر روی واژگان فارسی و انگلیسی آزمایش شده است و در نهایت نشان داده شده است که با استفاده از این روش می توان سرعت و دقت بازشناسی را به شکل قابل توجهی بهبود داد.

واژه های کلیدی : مدل زبانی پیش بین، درخت واژگان، دیکدر کلمات، بازشناسی گفتار پیوسته با واژگان بزرگ.

1- مقدمه

اصول بازشناسی آماری گفتار بر اساس فرمول احتمالات شرطی بیز می باشد.با توجه به محدودیتهای زبانی موجود در بازشناسی گفتار، رایج ترین سطح دیکد، سطح کلمات می باشد. با توجه به تعداد کلمات زیاد در بازشناسی گفتار پیوسته با واژگان بزرگ، یک واحد زیر کلمه ای برای به اشتراک گذاشتن قسمتهای مشترک در معادل تلفظی کلمات لازم است. این واحد زیر کلمه ای که می تواند به عنوان واحد آوایی مورد استفاده سیستم نیز مطرح باشد، هم برای حل مشکل کمبود داده های آموزشی در فاز آموزش مدل های آوایی مورد استفاده قرار می گیرد و هم در کاهش فضای جستجویی که در پروسه دیکدینگ با آن مواجه هستیم نقش بسیار مهمی دارد.

در عمل واژگان سیستم با استفاده از واحد آوایی انتخاب شده به صورت یک درخت پیشوندی تلفظی[2] (PPT) سازمان دهی می شود. فرایند دیکد در این فضا هدایت می شود و به این وسیله تضمین می شود که خروجی سیستم دنباله کلماتی باشد که در واژگان وجود دارند. بر این اساس توسعه فرضیه های فعال و تعیین گره های فعال درخت واژگان در فریم های بعدی باید بر اساس گره های قابل دسترس از گره های های فعال در فریم جاری صورت گیرد.

در سیستمهایی که واژگان آنها بر اساس درخت پیشوندی تلفظی ارائه می شود با توجه به اینکه گره های ریشه درخت واژگان ممکن است در بین چندین کلمه مشترک باشند، امکان استفاده از احتمالات مدل زبانی در حین ورود به گره ریشه درخت واژگان وجود ندارد. بر این اساس هدف اصلی مدل زبانی پیش بین، تخمین زدن احتمال مدل زبانی کلمه ای است که در درخت واژگان در حال بازشناسی می باشد . مدل زبانی پیش بین سعی دارد احتمال مزبور را هرچه سریعتر تخمین بزند تا این احتمال را در ارزشیابی فرضیه های مختلفی که برای بازشناسی کلمه کاندید هستند، استفاده کند. هرچه مدل زبانی ای که در مدل زبانی پیش بین، تخمین زده می شود از مرتبه بالاتری باشد، در ارزش یابی فرضیه های مختلف هرس بهتری را می توان اعمال کرد.

برای استفاده از مدلهای زبانی مرتبه بالاتر روشی در[1] ارائه شده است که به صورت مجزا برای هر کلمه بازشناسی شده در سابقه فرضیه های فعال، یک درخت واژگان جدید می سازد و احتمالات مدل زبانی فاکتور گیری شده بایگرام به ازاء آن کلمه را در آن نگهداری می کند. این روش، مصرف حافظه بسیار بالایی دارد و به دلیل ایجاد و حذف درخت های واژگان به صورت پویا و در حین بازشناسی، سربار

Archive of SID
www.SID.ir

محاسباتی بالایی را به سیستم وارد می کند و باعث کندی فرایند دیکد می شود. علاوه بر این مدل زبانی ترایگرام در این روش قابل استفاده نیست.

دراین مقاله ما با بررسی چگونگی اثر بخشی مدل زبانی پیش بین در دقت و سرعت بازشناسی، یک روش محاسبه ای جدید را برای استفاده از احتمالات مدل زبانی ترایگرام ارائه خواهیم کرد. در نهایت الگوریتمی برای استفاده ترکیبی از مدل زبانی پیش بین با استفاده از اطلاعات مونوگرام و ترایگرام ارائه شده است. این روش با کم کردن بار حافظه مورد نیاز این کار، علاوه بر کاهش خطای بازشناسی، سرعت بازشناسی را نیز افزایش می دهد.

در ادامه این مقاله در بخش 2 اصول روش LMLA و مشکلات آن را به اختصار توضیح می دهیم. در بخش 3 اصول روشهای جستجو و روش جستجوی مورد استفاده در این مقاله را توضیح می دهیم. در بخش 4 ساختار درخت واژگان استفاده شده تشریح شده است و در بخش 5 نحوه نگهداری و استفاده از مدل زبانی توضیح داده شده است. در بخش 6 روش پیشنهادی برای بهبود سیستم بازشناسی گفتار پایه ارائه شده است و در بخش 7 نیز نتایج بدست آمده از اعمال پیشنهادی بر روی دادگان انگلیسی و فارسی ارائه شده است. در بخش 8 که بخش نتیجه گیری می باشد نتایج بدست آمده بررسی شده است.

2- مدل زبانی پیش بین

به روشهای مختلف می توان مدل زبانی پیش بین را در فرایند دیکد اعمال کرد. در [1] و [2] اصول روش مدل زبانی پیش بین توضیح داده شده است. در اصل چگونگی اعمال مدل زبانی پیش بین بسیار وابسته به الگوریتم جستجوی مورد استفاده (در [3] مروری بر استراتژی های جستجو و الگوریتم های دیکد آورده شده است) و چگونگی گسترش فضای جستجو است. ولی به هر حال تمامی روشهای ارائه شده برای این کار دارای مشکلاتی بشرح زیر هستند:

- درصورت استفاده از مدل زبانی بایگرام یا ترایگرام در مدل زبانی پیش بین باید به نحوی معین کنیم تخمین ما از بهترین احتمال مدل زبانی مربوط به چه سابقه ای از کلمات بازشناسی شده می باشد. برای این کار روشی در [1] ارائه شده است. یک راه دیگر می تواند این باشد که لیست کلماتی که از هر گره می توانیم به آن برسیم را در هر گره نگه داریم. هر دو این روشها نیازمند یک حافظه بسیار بزرگ می باشد که عملا استفاده از این روش را با مشکل مواجه می کند. برای حل این مشکل یک روش به اشتراک گذاشتن لیست کلمات قابل دسترس در بین ریشه های درخت واژگان تلفظی پیشوندی و بچه های آن ارائه شده است.
- با توجه به ساختار درخت واژگان تلفظی پیشوندی، لغت مورد بازشناسی در گره های میانی معین نیست و در گره های نزدیک به ریشه محاسبه احتمالات مدل زبانی در بین تمامی کلمات قابل دسترس از آن گره و استفاده از بهترین احتمال ممکن در پروسه دیکد کار مناسبی نمی باشد زیرا محاسبات احتمالات مدل زبانی در بین کلماتی انجام می شود که اگرچه اشتراکات تلفظی با یکدیگر دارند ولی از نظر گرامری لزوما ارتباتی در بین آنها نمی باشد. برای کاهش اثر این مشکل بر نتایج بازشناسی از مدل آوایی سه واجی به عنوان واحد آوایی برای ساخت واژگان پیشوندی استفاده شده است تا هم اشتراک مدلهای آوایی در ابتدای کلمات حفظ شود و هم با بالا بردن فاکتور انشعاب در درخت واژگان، لیست کلمات قابل دسترس از هر گروه کوچکتر شود. برای اعمال مدل زبانی پیش بین در گره های درخت واژگان نیز از یک حد آستانه[3] استفاده شده است.

3- الگوریتم جستجو

به صورت عمده دو دسته دیکدر در سیستمهای بازشناسی گفتار پیوسته ای که بر اساس توسعه پویای درخت واژگان کار می کنند، وجود دارد، دیکدینگ همزمان[4] و دیکدینگ غیر همزمان[5].

در دیکدینگ غیر همزمان که به عنوان دیکدینگ پشته نیز شناخته می شود، فرضیه های فعال[6] در بعد زمان به صورت مستقل از یکدیگر توسعه داده می شوند و یک تابع تست[7] برای تخمین هزینه مسیر باقیمانده در تکمیل فرضیه لازم است که باید یک تابع قابل قبول[8] باشد.

در دیکدینگ همزمان تمامی فرضیه های فعال در هر فریم گسترش داده می شوند و در هر فریم فرایند هرس[9] و حذف فرضیه های ضعیف انجام می شود. این هرس بر اساس بهترین امتیاز فرضیه فعال بدست آمده در هر فریم و پارامتر شعاع[10] تعیین شده انجام می گیرد. در حین انتقال بین کلمات[11] تنها بهترین فرضیه گسترش داده می شود. با توجه به همزمان بودن تمامی فرضیه های فعال براحتی می توان امتیاز فرضیه ها را در هر فریم با یکدیگر مقایسه کرد. روش جستجوی انتخاب شده برای این مقاله جستجوی ویتربی شعاعی همزمان است. یکی از مهمترین جنبه های دیکدر های با این مشخصات نوع درخت واژگان مورد استفاده در آنها است. طرح های متفاوتی برای درخت واژگان ارائه و پیاده سازی شده است. مهمترین تفاوت این طرح ها نیز چگونگی استفاده از امتیازات مدل زبانی در فرایند جستجو می باشد. در [4] Ortmanns از استراتژی کپی درختهای واژگان برای اعمال امتیاز مدل زبانی در فرایند جستجو استفاده می کند. ایراد اصلی این روش این است که برای ایجاد و حذف پویای درخت های واژگان باید بار محاسباتی زیادی را تحمل کرد تا بتوان مدیریت حافظه مناسبی داشت و حتی در صورتی که میزان حافظه استفاده شده مهم نباشد، بار پردازشی ایجاد و حذف پویای درخت های واژگان بسیار زیاد می باشد.

روش مورد استفاده ما در این مقاله استفاده از تعداد ثابتی کپی درخت واژگان است. این روش بطور کامل در [6],[5] توضیح داده شده

Archive of SID

www.SID.ir

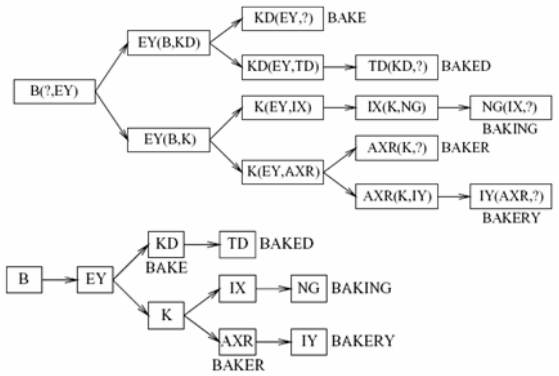
است. در این روش تعداد معینی کپی درخت واژگان ایجاد می شود و در انتقال های بین کلمات سعی می شود که تا حد ممکن کلمه ای که قرار است بازشناسی آن شروع شود، در همان درختی که آخرین کلمه فرضیه در آن بازشناسی شده است، بازشناسی نشود و این کار در یک کپی دیگر از درخت واژگان انجام گیرد. واضح است که به دلیل ثابت بودن تعداد کپی های درخت واژگان، این کار همیشه امکان پذیر نیست ولی در این روش سعی می شود با انتخاب روشهای مختلف این هدف تا حد امکان برآورده شود.

با بازشناسی کلمات متوالی در سابقه یک فرضیه، در درختهای مجزا، احتمال حذف شدن کلمه هایی که برای بازشناسی کاندید می باشند ولی در مقایسه با دیگر کلمات کاندید، امتیاز شروع قوی ای ندارند کاهش می یابد.

4- ساختار واژگان

دو روش عمده برای ارائه واژگان در قالب درخت واژگان تلفظی پیشوندی بر اساس انتخاب واحد آوایی وجود دارد. این واحد می تواند واج یا سه واجی باشد. شکل(1) نمونه ای از درختهای واژگانی که با استفاده از این دو واحد آوایی ساخته شده اند را نشان می دهد.

در درخت واژگانی که بر اساس سه واجی ساخته شده باشد، متن های سمت راست[12] و چپ[13] نیز در حین ساخت درخت در نظر گرفته می شود و یک گره از درخت تنها در بین کلماتی به اشتراک گذاشته می شود که دارای واجهای سمت راست و چپ یکسان در اطراف واج مرکزی[14] باشند. در این شکل از درخت واژگان گره های برگ در بین هیچ دو کلمه ای مشترک نمی باشند و در صورتیکه دو کلمه تلفظ یکسانی داشته باشند، برای آنها گره برگ جداگانه ایجاد می شود. برای گره های ریشه نیز متن سمت چپ به صورت پویا و در حین بازشناسی معین می شود.



شکل(1) بالا)درخت واژگان با استفاده از سه واجی. پایین)درخت واژگان با استفاده از واج[7]

در درخت واژگان تلفظی پیشوندی ای که بر اساس واج ساخته شده باشد دیگر ملاحظات متن سمت چپ و راست در نظر گرفته نمی شود.

مهمترین مزیت استفاده از واحد آوایی سه واجیدر ساخت درخت واژگان تلفظی پیشوندی این است که این درخت مذکور در مقایسه با حالتی که از واج در ساخت درخت واژگان استفاده می شود، فاکتور انشعاب بالاتری در نزدیک ریشه دارد و این موضوع باعث می شود که عملا در سطوح بعد از 3 با ساختار تقریبا صاف مواجه باشیم. این موضوع علاوه بر این که از فواید اشتراک بین معادلهای تلفظی کلمات بهره می برد، با بالا بردن فاکتور انشعاب در سطوح بعدی باعث کوچکتر شدن لیست کلمات قابل دسترسی از هر گره می شود و استفاده از مدل زبانی پیش بین را موثر تر می سازد. استفاده از سه واجی در ساخت درخت واژگان اگرچه حجم فضای جستجو را افزایش می دهد ولی این افزایش حجم با هرس تیزتر[15] فضای جستجو با استفاده از مدل زبانی پیش بین جبران میشود.

4-1- نگهداری لیست کلمات در درخت واژگان

یکی از مهمترین مسائل بازدارنده در اعمال مدل زبانی پیش بین، نگهداری لیست کلمات قابل دسترس از هر گره می باشد. در درخت واژگان تلفظی پیشوندی در گره ریشه، لیست کلیه کلماتی که از آن ریشه قابل دسترسی هستند نگهداری می شود و تمامی گره های بچه ای که از این گره ریشه منشعب می شوند، بخشی از این لیست کلمات را در لیست کلمات قابل دسترس خود دارند. لذا نگهداری لیست کلمات برای تمامی گره های درخت واژگان باعث زیاد شدن سربار حافظه ای می شود. برای حل این مشکل می توان از یک آرایه بیتی در هر یک از گره های بچه در درخت استفاده کرد که این آرایه بیتی قابل دسترس بودن یا نبودن کلمات در لیست کلمات ریشه، از هر یک از گره های میانی را، معین می سازد. نکته مهم در اینجا این است که حافظه استفاده شده در این روش باید کمتر از حالت نگهداری کل لیست کلمات در درخت واژگان باشد. با توجه به اینکه واحد حافظه در سیستم عامل 32 بیت (4 بایت) می باشد و با توجه با اینکه اندازه آرایه بیتی باید برای نشان دادن اندیس کلمات در لیست کلمات گره ریشه در اکثر گره های منشعب شده از گره های ریشه در درخت واژگان کافی باشد، یک آمارگیری از گره های ریشه ای که لیست کلمات آنها کوچکتر از 32 می باشد انجام شد که نتیجه آن در

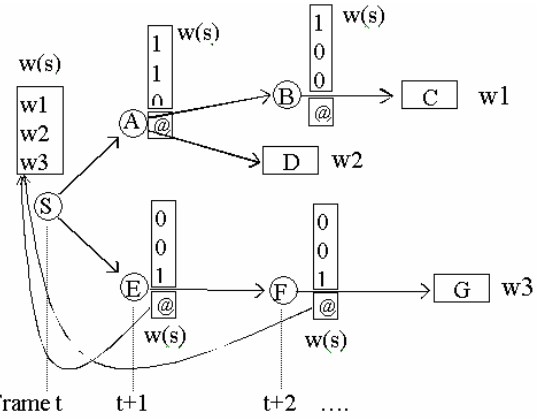
جدول(1) آورده شده است. با استفاده از این جدول می توان دید که تقریبا 98٪ کل گره های ریشه در واژگان متفاوت، لیست کلماتشان کوچکتر از 32 است.

Archive of SID

www.SID.ir

جدول(1) درصد گره هایی از درخت واژگان که لیست کلمات آنها کوچکتر از 32 می باشد

واژگان	زبان	تعداد گره های درخت	گره های با لیست کوچکتر از 32
5k	انگلیسی	26349	٪99.6
20k	انگلیسی	79914	٪98.0
64k	انگلیسی	212610	٪95.78
فارس دات	فارسی	110507	٪97.6



شکل(2) نگهداری لیست کلمات برای گره های میانی با استفاده از آرایه بیتی

بر اساس این جدول آرایه بیتی ای با بلاکهای 32 بیتی برای معین کردن لیست کلمات قابل دسترس در نظر گرفته شده است. با استفاده از این آرایه بیتی و یک اشاره گر به لیست کلمات گره ریشه می توان کلمات قابل دسترس در هر گره را معین کرد. نحوه این کار درشکل(2) نشان داده شده است. در این مرحله یک بهبود دیگر نیز امکان پذیر می باشد. در بسیاری از گره های درخت واژگان با فاکتور انشعاب 1 رو به رو هستیم و در این حالت امتیاز مدل زبانی پیش بین و لیست کلمات قابل دسترس برای گره میانی با فاکتور انشعاب 1 و بچه آن یکسان می باشد. برای مثال درشکل(2) آرایه بیتی برای گره های E و F یکسان می باشد. بهبود مورد نظر به این صورت است که به جای ایجاد آرایه بیتی و مرجع به لیست کلمات ریشه در گره F از اطلاعات گره E در گره F استفاده می کنیم.

با استفاده از آرایه بیتی، بار حافظه ای نگهداری لیست کلمات برای اعمال مدل زبانی پیش بین بسیار کاهش می یابد و این امکان فراهم می شود که بتوان از مدل زبانی بایگرام و ترایگرام در مدل زبانی پیش بین استفاده کرد. درجدول(2) و

جدول (3) حافظه مورد نیاز برای روش آرایه بیتی و روش نگهداری لیست کلمات به صورت عادی مقایسه شده است.

جدول(2) بار حافظه ای نگهداری لیست کلمات در حالت عادی

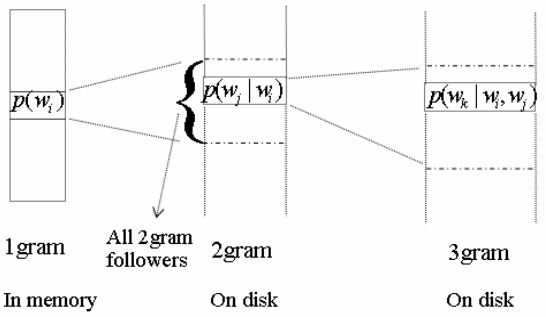
واژگان	زبان	تعداد گره های درخت	سربارحافظه برای هر درخت
5k	انگلیسی	26349	15MB
20k	انگلیسی	79914	57MB
64k	انگلیسی	212610	163MB
فارس دات	فارسی	110507	47MB

جدول (3) بار حافظه ای نگهداری لیست کلمات با روش آرایه بیتی

واژگان	زبان	تعداد گره های درخت	سربارحافظه برای هر درخت
5k	انگلیسی	26349	2.6 MB
20k	انگلیسی	79914	3.0 MB
64k	انگلیسی	212610	3.8 MB
فارس دات	فارسی	110507	2.8MB

5- ساختار مدل زبانی

در بازشناسی گفتار پیوسته با واژگان بزرگ حجم اطلاعات مدل زبانی که جزء ورودی های سیستم بازشناسی گفتار است، بسیار بزرگ می باشد. برای ساده سازی و کاراتر کردن کار با اطلاعات گسترده یک ساختار سلسله مراتبی برای اطلاعات مدل زبانی ارائه شده است. این ساختار درشکل(3) آورده شده است.



شکل(3) ساختار سلسله مراتبی برای نگهداری اطلاعات مدل زبانی

برای کار با این ساختار دو راه داریم

راه اول نگهداری اطلاعات مدل زبانی در حافظه اصلی است. در این حالت کلیه اطلاعات مدل زبانی در زمان بارگذاری برنامه در حافظه اصلی بارگذاری می شود. دراین حالت حافظه بسیار بزرگی صرف نگهداری اطلاعات مدل زبانی می شود ولی مزیت آن این است که زمان محاسبه احتمالات مدل زبانی بسیار کاهش می یابد. با توجه به اینکه حافظه مورد نیاز در این روش زیاد است، از این روش استفاده نمی شود.

روش دوم، نگهداری بخشی از اطلاعات مدل زبانی در حافظه و بخش در حافظه جانبی است. در این روش اطلاعات مونوگرام را در حافظه اصلی نگه می دارند و اطلاعات بایگرام و ترایگرام در حافظه جانبی نگهداری می شود و در صورت نیاز در حافظه بارگذاری می شود.

Archive of SID

www.SID.ir

برای نشان دادن روند کار در این حالت فرض کنید که در حین باز شناسی نیاز به به احتمال مدل زبانی برای دنباله کلمات $w_i w_j$ داریم. در این شرایط با استفاده از w_i کلیه بایگرام های ممکن برای w_i را معین می کنیم و با استفاده از w_j بایگرام مورد نظر را از بین آنها بدست می آوریم. در شرایطی که نیاز به احتمال مدل زبانی برای دنباله کلمات $w_i w_j w_k$ داشته باشیم ابتدا با استفاده از w_i کلیه بایگرام های ممکن برای w_i رامعین می کنیم. سپس با استفاده از w_j کلیه ترایگرام های ممکن برای دنباله کلمات $w_i w_j$ را معین می کنیم و در نهایت با استفاده از w_k ترایگرام مورد نظر را پیدا میکنیم. برای کاهش بار ورودی خروجی در تایین احتمال مدل زبانی مورد نظر، از یک مکانیسم حافظه نهان[16] به صورت زیر استفاده می شود:

- هر گاه نیاز به یک بایگرام برای کلمه آخر w_i داشتیم که در حافظه موجود نبود کلیه بایگرام های مربوط به w_i در حافظه نهان بارگذاری می شود.
- هر گاه برای جفت کلمه $w_i w_j$ نیاز به یک ترایگرام داشتیم که در حافظه موجود نبود ابتدا تمامی بایگرام های مربوط به w_i در حافظه نهان بار گذاری می شود و سپس با استفاده از w_j تمامی ترایگرام های مربوط به $w_i w_j$ در حافظه نهان بار گذاری می شود.
- برای جلوگیری از افزایش بیش از حد حجم حافظه نهان از یک مکانیسم جمع آوری آشغال[17] استفاده می کنیم. این ماژول در انتهای باز شناسی هر گویش[18] اجرا می شود و خانه های بدون استفاده حافظه نهان را خالی می کند و دوباره به سیستم بر می گرداند.

6- روش پیشنهادی برای بهبود عملکرد مدل زبانی پیش بین

هدف اصلی از استفاده مدل زبانی پیش بین، تقریب زدن احتمال مدل زبانی کلمه ای می باشد که به طور ناقص بازشناسی شده است و اعمال هر چه سریعتر احتمال مدل زبانی تقریب زده شده در پروسه دیکدینگ می باشد تا بصورت یک امتیاز تقویتی برای کلماتی که امتیاز آوایی آنها ضعیف است عمل کند. این کار را می توان با فاکتور گیری امتیاز مدل زبانی روی گره های درخت واژگان انجام داد[7]. برای یک سابقه فرضیه h ، امتیاز مدل زبانی فاکتور گیری شده $\pi_h(s)$ برای حالت s و دنباله کلمه v تشخیص داده شده در سابقه فرضیه (برای مدل زبانی بایگرام این دنباله شامل یک کلمه و برای ترایگرام این دنباله شامل دو کلمه می باشد)، احتمال مدل زبانی فاکتور گیری شده به صورت زیر محاسبه می شود:

$$\pi_h(s) = \max_{w \in W(s)} p(w \mid v) \qquad (1)$$

در این مقاله هدف ما استفاده از مدل زبانی ترایگرام در مدل زبانی پیش بین در محاسبه احتمال مدل زبانی فاکتور گیری شده است. در صورتی که احتمال مدل زبانی ترایگرام مورد نظر در سیستم وجود نداشته باشد، از مدلهای زبانی مرتبه پایین تر استفاده خواهد شد.

6-1- مدل زبانی پیش بین مونوگرام

در اعمال مدل زبانی پیش بین با استفاده از اطلاعات مدل زبانی بایگرام و یا ترایگرام، با توجه به اینکه سابقه کلمات تشخیص داده شده در حین ساخت درخت واژگان معلوم نمی باشد و در حین بازشناسی و به صورت پویا معین می شود، نگهداری لیست کلمات قابل دسترسی از هر گره برای محاسبه پویای امتیازات مدل زبانی در حین بازشناسی لازم می باشد ولی به دلیل اینکه مدل زبانی یک گرمی مستقل از سابقه فرضیه می باشد، می تواند بدون نگهداری لیست کلمات قابل دسترسی از هر گره اعمال شود. برای این منظور لازم است در زمان ساخت درخت واژگان بهترین احتمال مدل زبانی مونوگرام فاکتور گیری شده در هر گره را با استفاده از رابطه زیر نگه داریم:

$$\pi_h(s) = \max_{w \in W(s)} p(w) \qquad (2)$$

نتایج حاصل از اعمال مدل زبانی پیش بین با استفاده از مدل زبانی مونوگرام در بخش نتایج آورده شده است.

6-2- مدل زبانی پیش بین ترایگرام

همنطور که گفتیم برای استفاده از اطلاعات بایگرام و ترایگرام در حین اعمال مدل زبانی پیش بین، نگهداری لیست کلمات قابل دسترسی اجتناب ناپذیر می باشد. روش دیگری برای این کار در[1] ارائه شده است که برای مدل زبانی پیش بین با استفاده از اطلاعات بایگرام کاربرد دارد. این روش دارای حجم محاسباتی بسیار بالائی برای ایجاد و حذف پویای درختهای واژگان است و علاوه بر این با توجه به حافظه مورد نیاز در این روش، استفاده از اطلاعات ترایگرام در این روش ناممکن است.

برای حل این مشکل با استفاده از ساختار ارائه شده برای درخت واژگان در بخش4-1، روش زیر را برای استفاده ترکیبی مدل زبانی پیش بین با استفاده از احتمالات مدل زبانی مونوگرام و ترایگرام، ارائه می کنیم:

6-2-1- گره های ریشه

با توجه به اینکه لیست کلمات قابل دسترس در گره های ریشه بزرگ می باشد و با توجه به زمان بر بودن انجام محاسبات مدل زبانی، استفاده از مدل زبانی پیش بین در گره ریشه کار درستی نمی باشد و سرعت بازشناسی را به شدت کاهش می دهد. علاوه بر این در گره های ریشه تعداد لغات بسیار زیاد می باشد و فاکتورگیری احتمال مدل زبانی در بین تمامی این کلمات ممکن است به جای هدایت جستجو در مسیر درست، باعث انحراف جستجو از مسیر اصلی شود. برای حل

Archive of SID

www.SID.ir

این مشکل در گره های ریشه تنها از اطلاعات مونوگرام درمدل زبانی پیش بین استفاده می کنیم.

6-2-2- گره های میانی

با توجه به اینکه گره میانی در نزدیک ریشه قرار گرفته باشد یا نزدیک گره های برگ، لیست کلمات قابل دسترس آنها می تواند بزرگ یا کوچک باشد. علاوه بر این در صورت اعمال مدل زبانی پیش بین با استفاده از اطلاعات ترایگرام در یک گره خاص، باید عمل فاکتور گیری در آن گره را بر روی تمامی کلمات قابل دسترس از آن گره انجام داد تا احتمالات بدست آمده قابل اعتماد باشد. لذا برای کاهش عملیات مدل زبانی، اعمال مدل زبانی پیش بین با استفاده از احتمالات ترایگرام در گره های میانی را تا رسیدن به گره ای که لیست کلمات آن از یک حد آستانه ای کوچکتر باشد، به تعویق می اندازیم. با توجه به اینکه بین اعمال هرچه سریعتر مدل زبانی پیش بین با استفاده از احتمالات ترایگرام در گره های درخت و سرعت بازشناسی یک تقابل[19] وجود دارد، آستانه های مختلف برای اندازه لیست کلمات در نظر گرفته شد. نتایج بازشناسی با در نظر گرفتن آستانه های مختلف و برای واژگان مختلف در بخش نتایج آورده شده است.

میزان پوشش گره های درختان واژگان مختلف با انتخاب آستانه های مختلف برای لیست کلمات در جدول(4) آورده شده است

جدول(4)میزان پوشش گره های درخت واژگان مختلف با آستانه های مختلف برای لیست کلمات

واژگان	تعداد گره	1=	5>	20>
5k	26349	٪52.2	٪89.2	٪98.7
20k	79914	٪46.01	٪82.5	٪95.8
64k	212610	٪43.1	٪78.8	٪93.4
فارس دات	110507	٪51.3	٪85.4	٪96.1

استفاده از این مقادیر برای آستانه اندازه لیست کلمات، می تواند روند کلی تاثیر انتخاب آستانه های مختلف بر روند بازشناسی را نشان دهد.

6-2-3- کاهش عملیات محاسبه مدل زبانی در گره های میانی

بسیاری از عملیات محاسبه مدل زبانی ای که در گره های میانی انجام می شود بی مورد است. برای روشن شدن موضوع با توجه به شکل(2) در گره ریشه دو حالت ممکن است اتفاق بیفتد:

اگر کلمه ای که بهترین احتمال مدل زبانی پیش بین را می دهد W1 باشد، در گره A با توجه به اینکه کلمه W1 هنوز قابل دسترس می باشد دیگر نیازی به محاسبه مجدد احتمال مدل زبانی پیش بین برای لیست کلمات قابل دسترس وجود ندارد زیرا مجددا W1 بهترین احتمال را بدست می آورد. به این دلیل می توان امتیاز بدست آمده در گره ریشه S را در گره A نیز استفاده کرد. این شرایط برای گره B نیز برقرار می باشد. ولی برای گره E با توجه به اینکه کلمه W1 دیگر قابل دسترسی نیست احتمالات مدل زبانی پیش بین، باید مجددا برای لیست کلمات قابل دسترس محاسبه شود.

با استفاده از عملیات بیتی می توان به راحتی شرایط بیان شده را در بالا کنترل کرد و به این ترتیب می توان تعداد عملیات مدل زبانی را کاهش داد.

6-2-4- گره های برگ

در این گره ها نیز دو حالت وجود دارد. اگر گره برگ تنها فرزند والد خود باشد احتمال مدل زبانی پیش بین محاسبه شده در گره والد دقیق بوده و میتوان از این احتمال به عنوان احتمال مدل زبانی صحیح در حین تکمیل بازشناسی یک کلمه استفاده کرد.

اگر گره برگ تنها فرزند گره والد خود نباشد باید احتمال مدل زبانی صحیح در گره برگ مجددا محاسبه شود.

7- نتایج آزمایشات

آزمایش سیستم پایه و روشهای ارائه شده برای بهبود سیستم بازشناسی گفتار پیوسته بر روی دادگان انگلیسی و فارسی انجام شده است. برای آزمایشات انگلیسی مجموعه تست NOV92 که از مجموعه های تست wall street journal می باشد انتخاب شده است. برای آموزش مدلهای آوایی نیز از مجموعه آموزشی WSJ SI-84 که یکی از مجموعه های آموزشی Wall Street Journal است، استفاده شده است. این مجموعه آموزشی شامل 7296 جمله است که توسط 84 گوینده ادا شده اند. مجموعه زمان این گویش ها در حدود 12 ساعت می باشد. مدلهای آوایی آموزش داده شده شامل 4147 زنون می باشد و هر HMM شامل 3 حالت با حداکثر 8 تابع گوسی در هر حالت می باشد.

برای فارسی نیز از مجموعه دادگان آموزشی و آزمایشی فارس دات کوچک استفاده شده است. لازم به ذکر است که درخت واژگان و مدل زبانی استفاده شده برای تست های فارسی بر اساس دادگان 20k می باشد ولی مجموعه جملات تست از واژگان فارس دات کوچک انتخاب شده است. به این ترتیب در آزمایشات فارسی می توان فضای جستجو را معادل فضای جستجوی 20k انگلیسی در نظر گرفت. مدلهای آوایی آموزش داده با این مجموعه دارای 650 زنون می باشد که درقالب 4362 مدل HMM قرار گرفته اند.

در این آزمایشات از یک عدد کامپیوتر با توان پردازشی 3.0GH به همراه 512MB حافظه استفاده شده است.

در جدول(5) دقت و زمان بازشناسی برای سیستم پایه آورده شده است. در جدول(6) اثر انتخاب آستانه های مختلف بر زمان و دقت سیستم بازشناسی نشان داده شده است.

Archive of SID

www.SID.ir

جدول(5) سرعت و دقت بازشناسی برای واژگان مختلف درسیستم پایه

واژگان	زبان	زمان بازشناسی(s)	خطای بازشناسی
5k	انگلیسی	1162	6.7
20k	انگلیسی	1625	15.6
64k	انگلیسی	2892	18.1
فارس دات	فارسی	110	19.02

جدول(6)سرعت و دقت بازشناسی برای واژگان مختلف در روش ارائه شده

واژگان	زبان	آستانه	زمان بازشناسی(s)	خطای بازشناسی(WER)
5k	انگلیسی	20	831	5.5
5k	انگلیسی	5	811	5.6
5k	انگلیسی	1	781	5.6
20k	انگلیسی	20	1383	15.1
20k	انگلیسی	5	1208	14.9
20k	انگلیسی	1	1162	14.8
64k	انگلیسی	20	2504	17.6
64k	انگلیسی	5	2113	17.5
64k	انگلیسی	1	2011	17.5
فارس دات	فارسی	20	119	16.85
فارس دات	فارسی	5	110	16.9
فارس دات	فارسی	1	100	17.04

8- نتیجه

با توجه با نتایج ارائه شده در بخش نتایج، با افزایش مقدار آستانه تعداد اعمال محاسبه احتمال مدل زبانی افزایش می یابد و به این دلیل زمان بازشناسی افزایش می یابد. علاوه بر این در گره های با لیست کلمات بزرگ به دلیل عدم ارتباط گرامری بین کلمات، جستجو تا حدی از مسیر اصلی منحرف می شود و به این دلیل هرچه اعمال مدل زبانی پیش بین در گره های نزدیک به برگ ها با لیست کلمات کوچکتر انجام شود دقت بازشناسی افزایش می یابد. با اعمال مدل زبانی پیش بین در گره های نزدیک به برگ که لیست کلمات آنها کوچک می باشد هرس به صورت نرم تری انجام می گیرد ولی به دلیل کاهش عملیات مدل زبانی که بخش عمده ای از زمان بازشناسی را به خود اختصاص داده است، زمان بازشناسی در کل کاهش می یابد. در نهایت بهترین زمان و دقت بازشناسی برای واژگان انگلیسی با آستانه 1 بدست می آید. این بدان معنی است در زبان انگلیسی هر زمان به گره ای رسیدیم که بین هیچ دو کلمه ای مشترک نبود (گره برگ یا گره ای که والد گره برگ می باشد ولی فاکتور انشعاب در آن 1 است) بهتر است که امتیاز قطعی مدل زبانی را اعمال کنیم.

در زبان فارسی نتایج در مورد دقت کاملا معکوس انگلیسی است و بهترین دقت با استفاده از آستانه 20 بدست می آید ولی بهترین زمان بازشناسی همانند انگلیسی در آستانه 1 بدست می آید. لذا در صورتی که دقت مهم باشد می توان از آستانه 20 استفاده کرد و در صورتی که سرعت مهم باشد باید از آستانه 1 استفاده کنیم.

مراجع

[1] S. Ortmanns, H. Ney, and A. Eiden, *"Language-model look-ahead for large vocabulary speech recognition,"* in Proc. ICSLP, Philadelphia, USA, vol.4 pp. 2095 - 2098, 1996.

[2] A. Cardenal-L´opez, F. J. Diguez-Tirado, and C. Garc´ıa-Mateo, *"Fast LM look-ahead for large vocabulary continuous speech recognition using perfect hashing,"* in Proc. ICASSP, Orlando, Florida, USA, pp. 705–708, May 2002.

[3] X. L. Aubert, *"An overview of decoding techniques for large vocabulary continuous speech recognition,"* Computer Speech and Language, vol. 16, pp. 89–114, 2002.

[4] S. Ortmanns, A. Eiden, H. Ney, *"Improved Lexical Tree Search for Large Vocabulary Speech Recognition"* Proc. of ICASSP'98, pp. 817-820, Seattle, May 1998.

[5] M. Ravishankar, *"The 1999 CMU 10X Real Time Broadcast News Transcription System"*, Proc. DARPA Workshop on Automatic Transcription of Broadcast News, Washington, 2000

[6] M. Ravishankar, *"Efficient Algorithms for Speech Recognition*, " Ph.D. Thesis, Carnegie Mellon University, May 1996, Tech Report CMU-CS-96-143.

[7] Ortmanns, S. Ney, H. Seide, E. Lindam, I.*"A comparison of time conditioned and word conditioned searchtechniques for large vocabulary speech recognition"* in Proc ICSLP 96, Philadelphia, PA, USA pp. 2091-2094 vol.4, 1996

زیرنویس‌ها

1 Language Model Look-Ahead
2 Pronunciation Prefix Tree
3 Threshold
4 Time synchronous decoding
5 Time asynchronous decoding
6 Active Hypothesis
7 Probe
8 Admissible
9 Pruning
10 Beam parameter
11 Word transition
12 Right context
13 Left context
14 Base phone
15 Sharper pruning
16 Cache
17 Garbage collector
18 Utterance
19 Trade off

Archive of SID
www.SID.ir