



تحقیق توسعه قابلیت به کار گیری مدل مخلوط گوسی برای مسئله تشخیص گفتار در حوزه مسائل یادگیری ماشین

دانشجویان: زهرا رضاپور و سروش عاملی

استاد راهنما: دکتر رشاد حسینی

دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران

نتایج

در این پروژه از مجموعه داده ی `vystadial_en` استفاده شده است. نسخه انگلیسی آن تشکیل شده از تماس های ضبط شده بین انسان ها راجع به محل مناسب برای شام خوردن در شهر و به زبان محاوره است. گویش در این تماس ها آمریکایی است. اطلاعات تکمیلی در جدول زیر نشان داده شده است:

	Audio [hour]	# sentences	# words
Training	41:30	47463	178110
Development	1:45	2000	7376
Test	1:46	2000	7772

معیار ارزیابی مورد استفاده در این پروژه، نرخ خطای تشخیص (WER) در نظر گرفته شده است. این شاخص بر اساس نسبت تعداد رخداد های غلط تشخیص داده شده به تعداد کل رخداد ها تعریف می شود.

پس از پیاده سازی الگوریتم، نتایج زیر حاصل شد. دقت شود که این پیاده سازی نسخه اولیه از پیاده سازی است و این الگوریتم هنوز هم در حال توسعه و تحقیق است.

خطا با حضور الگوریتم پیشنهادی	خطا با حضور الگوریتم STC
7.19.57%	12%

در پیاده سازی، بنا به توصیه های انجام شده، به جای به اشتراک گذاری پارامترها در هر مدل مخلوط، پارامترها بین تمام مدل های مخلوط به اشتراک گذاشته شده است. از طرفی مدل هنوز نقص هایی دارد که در حال رفع هستند.

جمع بندی

در این پروژه مطالعه و بررسی کامل بر روی مدل مخلوط گوسی که یکی از پرکاربردترین قویترین مدل ها برای یادگیری ماشین است، انجام گردید. برای این کار کارهای انجام شده اخیر مطالعه و بررسی شد و با توجه به اینکه استفاده از این مدل در مسائل تشخیص گفتار قابلیت بهبود را دارا بود، مدل جدید پیشنهادی پیاده سازی شد تا ضعف مدل های قبلی را جبران کند و عملیات تشخیص گفتار را کارا تر و سریع تر نماید. مدل STC بر این اساس استوار است که پارامترهای مدل مخلوط گوسی را بین همه ی گوسی ها به اشتراک بگذارد. این ایده اساس این پایان نامه است. با این تفاوت که در مدل پیشنهادی پیاده سازی شده توسط ما، بهبودهایی بر روی روش بهینه سازی آن اعمال شده است که با توجه به مدل های مشابه دارای نقاط قوت بارزی است. علی رغم در حال توسعه بودن، نتایج نشان میدهد که این مدل قابلیت تبدیل شدن به یک مدل جامع و فراگیر را پس از کامل شدن توسعه خواهد داشت. نوآوری این مدل در واقع شیوه برخورد با مسئله و طریقه به روزرسانی متغیرها و کاهش تعداد پارامترها است. محدودیت هایی که در این پروژه موجود بود، نحوه پیاده سازی کلدی بوده که باعث شد که با به اشتراک گذاری پارامترها بین مدل های مخلوط، از مدل اصلی اندکی فاصله بگیریم.

مراجع اصلی

1. C. Biernacki, G. Celeux, G. Govaert, and F. Langrognat, "Model-based cluster and discriminant analysis with the mixmod software," *Computational Statistics & Data Analysis*, vol. 51, no. 2, pp. 587–600, 2006.
2. M. J. Gales, "Semi-tied covariance matrices for hidden markov models," *IEEE Transactions on speech and audio processing*, vol. 7, no. 3, pp. 272–281, 1999.
3. R. Hosseini and S. Sra, "Matrix manifold optimization for Gaussian mixtures," in *Advances in Neural Information Processing Systems*, 2015, pp. 910–918.
4. I. Visser, "Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series," *IEEE transactions on speech and audio processing*, vol. 7, pp. 272-281, 1999.

مقدمه

زبان گفتار یکی از غریزی ترین اشکال ارتباط در بین جوامع بشری است. سیستم های گفتاری زبانی اجازه می دهد تا مردم با ماشین ها به شکل طبیعی صحبت کنند. کیفیت گفتگو با زبان محاوره به شکل قابل توجهی بستگی به دقت تشخیص صوت دارد. یکی از مسائلی که در پردازش صوت توجهات را به خود معطوف نموده است، مدلسازی سیگنال است. مدل های گاوسی، زنجیره مارکوف و مدل مخفی مارکوف از جمله روش هایی هستند که توسط آنها تشکیل مدل سیگنال سراسر است و تنها کافی ست مقادیر پارامترهای مدل تخمین زده شود. در مدل های آماری سعی در ایجاد مدل با استفاده از خواص آماری سیگنال است. فرض اساسی در مدل های آماری این است که می توان خواص سیگنال را به شکل یک فرآیند تصادفی پارامتری مدل نمود. در طی سالیان اخیر روش های متفاوتی برای تشخیص صوت ارائه شده است. هر کدام از این روش ها مزایا و معایب خاص خود را دارا می باشند. یکی از این روش ها، روش STC است که مبنای عملکردی آن، به اشتراک گذاری تعدادی از پارامترها بین تابع های توزیع گوسی در مدل مخلوط گوسی است و در سال ۱۹۹۸ توسط مارک گیلز (Mark Gales) ارائه شد. مبنای عملکرد در این پایان نامه همین مدل است.

مدل مخلوط گاوسی یک مدل پارامتری بسیار مؤثر در حوزه یادگیری ماشین و تشخیص الگوست که برای مسائل مختلف در زمینه های متنوع از مهندسی، علوم پایه، پزشکی و ... بکار می رود. تعداد مؤلفه های به کار رفته در مدل مخلوط یکی از عوامل تعیین کننده در میزان انعطاف پذیری و قدرت بیانی آن است. یک عامل مهم دیگر، نوع ماتریس کوواریانس مؤلفه ها است زیرا هم شکل آن ها را مشخص می کند و هم بخش عمده ای از کل پارامترهای مدل را در بر می گیرد. معمولاً مدل نمی تواند هم تعداد مؤلفه خیلی زیاد و هم ماتریس های کوواریانس با درجه آزادی بالا داشته باشد. این محدودیت از آن جا ناشی می شود که در چنین حالتی مدل مخلوط گاوسی تعداد بسیار زیادی پارامتر خواهد داشت بگونه ای که فرآیند یادگیری و تخمین پارامتر با مشکل مواجه می شود. علاوه بر این، تعداد زیاد پارامتر ممکن است موجب بیش برآزش در مدل شود. هنگامی که بعد داده ها زیاد باشد، مشکلات یادگیری تشدید نیز خواهد شد. این پدیده باعث می شود که نتوان مدل مخلوط گاوسی را برای بسیاری از کاربردها که نیاز به یک مدل مقیاس بزرگ دارند به کار برد. یک راه حل برای اینکه بتوانیم یک مدل مخلوط گاوسی با تعداد زیاد مؤلفه و تعداد پارامتر کم داشته باشیم، به کارگیری روش های تنظیم مدل است. مبنای اصلی این روش ها اعمال یک نظم مشخص و تعریف شده بر روی ساختار مدل است.

مدل پیشنهادی و پیاده سازی آن

به عنوان روش تنظیم، می توان یک مدل مخلوط گاوسی با تعداد زیاد مؤلفه و ماتریس های کوواریانس ساده مانند قطری تعریف نمود. جالب است که چنین تنظیم ساده ای، برای کاربردهای مختلف، بهبود قابل توجهی در عملکرد مدل مخلوط ایجاد نموده است. سوال اساسی که پیش می آید این است که آیا می توان یک روش تنظیم مدل توسعه داد که مجموع تعداد پارامترهای آن کم باشد تا امکان به کارگیری تعداد زیاد مؤلفه محقق شود و در عین حال ساختار ماتریس های کوواریانس خیلی محدود نگردد به گونه ای که قابلیت یادگیری و مؤثر بودن عملکرد آن حفظ شود. به اشتراک گذاری پارامترها پاسخ این سوال است. فرمول کلی مدل مخلوط گاوسی مطابق رابطه (۱) میباشد:

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (1)$$

در این رابطه، K تعداد گاوسی های مورد استفاده برای مدل کردن داده میباشد و $\boldsymbol{\Sigma}$ و $\boldsymbol{\mu}$ به ترتیب میانگین و ماتریس کواریانس مربوط به هر گاوسی میباشد. پارامترهای کل مدل مخلوط گاوسی (۲) در رابطه ی (۲) نمایش داده شده است:

$$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k \quad (2)$$

تعداد این پارامترها در حالیکه n تعداد ابعاد داده و K تعداد مؤلفه های گاوسی است برای دو حالت کلی و حالت مدل پیشنهادی در جدول زیر آورده شده است. همانطور که واضح است تعداد پارامترها بطور قابل توجهی در مدل پیشنهادی کاهش می یابد.

	General	Proposed
Decomposition	$\boldsymbol{\Lambda}_k = \mathbf{U}_{k \times n} \mathbf{D}_{k \times n} \mathbf{U}_{k \times n}^T$	$\boldsymbol{\Lambda}_k \approx \mathbf{U}_{n \times m} \mathbf{D}_{m \times m} \mathbf{U}_{m \times n}^T + \alpha \mathbf{I}_{n \times n}$
# \mathbf{D}_k params	Kn	Km
# \mathbf{U} params	Kn^2	mn
# Total params	$Kn^2 + Kn$	$mn + mK$