دانشگاه آزاد اسلامی واحد تهران جنوب

دانشکده فنی و مهندسی

تکلیف هفته هفتم پروژه نهایی

ارائه ۱۰ مقاله جدید و معتبر با موضوعات مشابه پروژه نهایی به همراه جدول مزایا و معایب آن ها

نام درس:

پردازش سیگنال دیجیتال

نام استاد راهنما:

آقای دکتر مهدی اسلامی

نام دانشجو:

زهرا کربلایی محمدی

شماره دانشجویی:

۴۰۰۱۴۱۴۰۱۱۱۰۳۰

تاریخ:

۱۴۰۱/۹/۴

پاییز ۱۴۰۱

- **ارائه ۱۰ مقاله جدید و معتبر با موضوعات مشابه:**

مقاله۱:

One-shot Voice Conversion with Global Speaker Embeddings

تبدیل صدای تک شات با جاسازی های جهانی بلندگو

ایجاد یک سیستم تبدیل صوتی (VC) برای یک بلندگوی هدف جدید معمولاً به مقدار زیادی داده گفتاری از بلندگوی هدف نیاز دارد. این مقاله روشی را برای ساختن یک سیستم VC برای بلندگوی هدف دلخواه با استفاده از یک گفته داده شده بدون هیچ گونه فرآیند آموزشی انطباق بررسی میکند. با الهام از نشانه های سبک جهانی (GST)، که اخیراً نشان داده شده است که در کنترل سبک گفتار مصنوعی مؤثر است، ما استفاده از جاسازی های بلندگوی جهانی (GSEs) را برای کنترل هدف تبدیل سیستم VC پیشنهاد میکنیم. پسینگرام های آوایی مستقل از بلندگو (PPGs) به عنوان ورودی شرایط محلی به یک سینتسایزر شرطی WaveNet برای تولید شکل موج بلندگوی هدف استفاده می شوند. در همین حال، طیف نگاری ها از گفته داده شده استخراج میشوند و به یک رمزگذار مرجع تغذیه میشوند، سپس جاسازی مرجع تولید شده به عنوان پرس و جوی توجه به GSEs برای تولید جاسازی بلندگو استفاده میشود، که به عنوان ورودی شرایط جهانی به سینت سایزر WaveNet برای کنترل استفاده میشود. هویت بلندگوی شکل موج را ایجاد کرد. در آزمایش ها، در مقایسه با یک سیستم VC مبتنی بر آموزش انطباق، رویکرد VC مبتنی بر GSEs به همان اندازه خوب یا بهتر از نظر طبیعی بودن گفتار و شباهت گوینده، با انعطاف پذیری ظاهراً بالاتر برای مقایسه، عمل میکند [۱].

مقاله ۲:

ONE-SHOT VOICE CONVERSION BY VECTOR QUANTIZATION

تبدیل صدای تک شات توسط کمی سازی برداری

در این مقاله، ما یک رویکرد تبدیل صدای تک شات (VC) مبتنی بر کوانتیزاسیون برداری (VQ) را بدون هیچ نظارتی بر روی برچسب بلندگو پیشنهاد میکنیم. ما جاسازی محتوا را بهعنوان یک سری کدهای گسسته مدلسازی میکنیم و تفاوت بین بردار quantize-before و quantize-after را به عنوان جاسازی بلندگو در نظر میگیریم. ما نشان میدهیم که این رویکرد توانایی قوی برای تفکیک محتوا و اطلاعات سخنران تنها با از دست دادن بازسازی دارد، و بنابراین VC یک شات به دست می آید [۲].

مقاله ۳:

AGAIN-VC: A ONE-SHOT VOICE CONVERSION USING ACTIVATION GUIDANCE AND ADAPTIVE INSTANCE NORMALIZATION

AGAIN-VC: یک تبدیل صدای یک شات با استفاده از راهنمای فعال سازی و عادی سازی نمونه تطبیقی

به تازگی، تبدیل صدا (VC) به طور گسترده مورد مطالعه قرار گرفته است. بسیاری از سیستمهای VC از تکنیکهای یادگیری مبتنی بر تفکیک برای جدا کردن گوینده و اطلاعات محتوای زبانی از سیگنال گفتار استفاده میکنند. سپس با تغییر اطلاعات بلندگو به بلندگوی هدف، صدا را تبدیل می کنند. برای جلوگیری از نشت اطلاعات گوینده به درون جاسازیهای محتوا، کارهای قبلی یا ابعاد را کاهش میدهند یا تعبیه محتوا را به عنوان یک گلوگاه اطلاعاتی قوی تعیین میکنند. این مکانیسم ها به نوعی به کیفیت سنتز لطمه می زند. در این کار، ما AGAIN-VC را پیشنهاد می کنیم، یک سیستم VC ابتکاری با استفاده از راهنمای فعال سازی و عادی سازی نمونه تطبیقی. AGAIN-VC یک مدل مبتنی بر رمزگذار خودکار است که از یک رمزگذار و یک رمزگشا تشکیل شده است. با فعالسازی مناسب بهعنوان گلوگاه اطلاعاتی در تعبیههای محتوا، مبادله بین کیفیت سنتز و شباهت سخنران گفتار تبدیل شده به شدت بهبود مییابد. این سیستم یک شات VC بدون در نظر گرفتن ارزیابی های ذهنی یا عینی بهترین عملکرد را به دست می آورد.

الگوریتم AdaIN-VC از یک رمزگذار خودکار متغیر متشکل از یک رمزگذار بلندگو، یک رمزگذار محتوا و یک رمزگشا استفاده می کند. با عادی سازی نمونه تطبیقی (AdaIN)، اطلاعات سخنران و اطلاعات محتوا به خوبی قابل تفکیک هستند.

یک محدودیت آشکار دارد که رمزگذار بلندگوی از پیش آموزش دیده صرفاً برای تأیید بلندگو آموزش دیده است. از این رو، استحکام تولید گفتار مشکوک است. AdaIN-VC از دو رمزگذار مستقل برای استخراج تعبیههای بلندگو و جاسازی محتوا به ترتیب استفاده میکند. با این وجود، ما معتقدیم که رمزگذار بلندگو در اینجا تا حدودی اضافی است. یعنی وظایف آنها می تواند توسط یک رمزگذار انجام شود.

الگوریتم AdaIN-VC از یک رمزگذار محتوا و یک رمزگذار بلندگو استفاده می کند، در حالی که AGAIN-VC تنها از یک رمزگذار و یک فعال سازی برای هدایت آموزش استفاده می کند [۳].

مقاله ۴:

High-Quality Many-to-Many Voice Conversion Using Transitive Star Generative Adversarial Networks with Adaptive Instance Normalization

تبدیل صدای چند به چند با کیفیت بالا با استفاده از شبکه های متخاصم مولد ستاره انتقالی با عادی سازی نمونه تطبیقی

این مقاله یک روش جدید تبدیل صدای چند به چند غیر موازی با کیفیت بالا مبتنی بر شبکههای متخاصم مولد ستاره انتقالی با عادیسازی نمونه تطبیقی (Trans-StarGAN-VC با AdaIN) پیشنهاد میکند. اول، ما ساختار مولد را با TransNets بهبود می دهیم تا از ویژگی های سلسله مراتبی مرتبط با طبیعی بودن گفتار استفاده کامل کنیم. در TransNets، بسیاری از اتصالات میانبر ویژگی های سلسله مراتبی را بین بخش رمزگذاری و رمزگشایی به اشتراک می گذارند تا اطلاعات زبانی و معنایی کافی را به دست آورند، که به ارائه گفتار تبدیل شده با صدای طبیعی کمک می کند و همگرایی فرآیند آموزش را تسریع می کند. دوم، با ترکیب AdaIN برای انتقال سبک، ما مولد را قادر میسازیم تا به جای استفاده از برچسبهای ویژگی، اطلاعات کافی از ویژگیهای بلندگو را مستقیماً از گفتار بیاموزد، که همچنین چارچوب امیدوارکنندهای را برای VC یکشات فراهم میکند. آزمایشهای عینی و ذهنی با دادههای آموزشی غیر موازی نشان میدهد که روش ما به طور قابلتوجهی از StarGAN-VC هم در طبیعی بودن گفتار و هم در شباهت گوینده بهتر عمل میکند. میانگین امتیازات میانگین نظر (MOS) و ABX به ترتیب ۲۴,۵ درصد و ۱۰,۷ درصد افزایش یافته است. مقایسه طیف نگاری همچنین نشان می دهد که روش ما می تواند ساختارها و جزئیات هارمونیک کامل تری ارائه دهد و به طور موثر شکاف بین گفتار تبدیل شده و گفتار هدف را پر کند [۴].

مقاله ۵:

WINVC: One-Shot Voice Conversion with Weight Adaptive Instance Normalization

WINVC: تبدیل صدای یک شات با عادی سازی نمونه های تطبیقی وزن

این مقاله یک راه حل تبدیل صدای یک شات (VC) را پیشنهاد می کند. در بسیاری از راه حل های تبدیل صدای تک شات (به عنوان مثال، روش های VC مبتنی بر رمزگذاری خودکار)، عملکردها به دلیل عادی سازی نمونه و عادی سازی نمونه تطبیقی به طور چشمگیری بهبود یافته است. با این حال، روان تبدیل صدای تک شات هنوز وجود ندارد، و شباهت به اندازه کافی خوب نیست. این مقاله استراتژی عادی سازی نمونه تطبیقی وزن را برای بهبود طبیعی بودن و شباهت تبدیل صدای تک شات معرفی می کند. نتایج تجربی ثابت میکند که تحت مجموعه دادههای VCTK، امتیاز MOS مدل پیشنهادی ما، تبدیل صدای عادیسازی نمونه تطبیقی وزن (WINVC)، با پنج مقیاس به ۳,۹۷ میرسد و SMOS با چهار مقیاس به ۳,۳۱ میرسد. علاوه بر این، WINVC میتواند به ترتیب امتیاز MOS 3.44 و SMOS 3.11 را برای تبدیل صدای یکشات تحت مجموعه دادهای کوچک از ۸۰ بلندگو با ۵ قطعه گفته برای هر نفر به دست آورد[۵].

مقاله ۶:

Disentangling audio content and emotion with adaptive instance normalization for expressive facial animation synthesis

تفکیک محتوای صوتی و احساسات با عادی سازی نمونه تطبیقی برای سنتز انیمیشن چهره رسا

در سالهای اخیر، سنتز انیمیشنهای سه بعدی صورت از طریق صدا مورد توجه قرار گرفته است. با این حال، بیشتر آثار ادبی موجود برای نقشهبرداری محتوای صوتی و تصویری طراحی شدهاند که دانش محدودی در مورد رابطه بین احساسات در انیمیشنهای صوتی و بیانی چهره ارائه میدهند. این کار انیمیشنهای صورت منطبق با صدا را با برچسب احساسی مشخص شده تولید میکند. در چنین کاری، ما استدلال می کنیم که جداسازی محتوا از صدا ضروری است - مدل پیشنهادی باید یاد بگیرد که محتوای چهره را از محتوای صوتی تولید کند در حالی که عبارات از احساسات مشخص شده است. ما آن را با یک ماژول عادی سازی نمونه تطبیقی به دست می آوریم که محتوای موجود در صدا را جدا می کند و احساسات جاسازی شده را از برچسب مشخص شده ترکیب می کند. سپس از تعبیه محتوای مشترک-احساس برای ایجاد رئوس سه بعدی صورت و نقشه های بافت استفاده می شود. ما روش خود را با خطوط پایه پیشرفته، از جمله رویکردهای جداسازی مبتنی بر تقسیمبندی صورت و مبتنی بر تبدیل صدا مقایسه میکنیم. ما همچنین یک مطالعه کاربر برای ارزیابی عملکرد شرطیسازی احساسات انجام میدهیم. نتایج نشان میدهد که روش پیشنهادی ما در کیفیت انیمیشن و دقت طبقهبندی بیان از خطوط پایه بهتر عمل میکند [۶].

مقاله ۷:

Towards low-resource stargan voice conversion using weight adaptive instance normalization

به سمت تبدیل صدای استارگان با منبع کم با استفاده از عادی سازی نمونه های تطبیقی وزن

تبدیل صدای چند به چند با داده های آموزشی غیر موازی پیشرفت قابل توجهی در سال های اخیر داشته است. به دلیل فقدان داده های موازی حقیقت، چالش برانگیز است. مدل های مبتنی بر StarGAN به دلیل کارایی و اثربخشی مورد توجه قرار گرفته اند. با این حال، بیشتر کارهای مبتنی بر StarGAN فقط روی تعداد کمی از بلندگوها و حجم زیادی از داده های آموزشی متمرکز شده اند. در این کار، هدف ما بهبود کارایی دادههای مدل و دستیابی به یک تبدیل صوتی غیر موازی مبتنی بر StarGAN برای تعداد نسبتاً زیادی از بلندگوها با نمونه های آموزشی محدود است. به منظور بهبود کارایی داده، مدل پیشنهادی از یک رمزگذار بلندگو برای استخراج تعبیه های بلندگو و لایه های عادی سازی نمونه تطبیقی وزن (W-AdaIN) استفاده میکند. آزمایش ها با ۱۰۹ سخنران در دو موقعیت کم منابع انجام میشود، که در آن تعداد نمونه های آموزشی ۲۰ و ۵ برای هر سخنران است. یک ارزیابی عینی نشان می دهد که مدل

پیشنهادی به طور قابل توجهی از روش های پایه بهتر عمل می کند. علاوه بر این، یک ارزیابی ذهنی نشان میدهد که هم برای طبیعی بودن و هم شباهت، مدل پیشنهادی از روش پایه بهتر عمل میکند [۷].

مقاله ۸:

CLSVC: LEARNING SPEECH REPRESENTATIONS WITH TWO DIFFERENT CLASSIFICATION TASKS.

CLSVC: بازنمایی های گفتار یادگیری با دو وظیفه طبقه بندی متفاوت.

تبدیل صدا (VC) با هدف تبدیل صدای یک گوینده برای تولید یک گفتار جدید همانطور که توسط گوینده دیگر گفته می شود. کارهای قبلی بر یادگیری بازنمایی نهفته با استفاده از دو رمزگذار مختلف برای یادگیری اطلاعات محتوا و اطلاعات تایم از گفتار ورودی به ترتیب تمرکز دارند. با این حال، چه آنها از یک شبکه تنگنا یا فناوری کمیت برداری استفاده کنند، جدا کردن کامل سخنران و اطلاعات محتوا از سیگنال گفتار بسیار دشوار است. در این مقاله، ما یک چارچوب تبدیل صوتی جدید، "ClsVC" را برای رسیدگی به این مشکل پیشنهاد می کنیم. تنها از یک رمزگذار استفاده میکند تا با تقسیم فضای پنهان، اطلاعات مربوط به زمان و محتوا را دریافت کند. علاوه بر این، برخی از محدودیتها پیشنهاد شدهاند تا اطمینان حاصل شود که بخشهای مختلف فضای پنهان تنها حاوی اطلاعات جداکننده محتوا و تایم هستند. ما ضرورت تنظیم این محدودیتها را نشان دادهایم، و همچنین بهطور تجربی ثابت میکنیم که حتی اگر نسبت تقسیم فضای پنهان را تغییر دهیم، اطلاعات محتوا و زمان همیشه به خوبی از هم جدا میشوند. آزمایشات روی مجموعه داده VCTK نشان می دهد که ClsVC یک چارچوب پیشرفته از نظر طبیعی بودن و شباهت گفتار تبدیل شده است.

در این مقاله، ما ClsVC را پیشنهاد کردیم، یک سیستم VC جدید که نمایش گفتار نهفته را یاد میگیرد. در طول آموزش، یک طبقهبندی کننده سخنران معمولی پیشنهاد میشود تا تعبیه تخمینی بلندگو را تشویق کند تا بیشتر و بیشتر با هویت گوینده مرتبط شود و یک طبقهبندی مخالف، محتوای تخمینی را بر روی محتوای تخمینی متمرکز میکند که مستقلتر از سخنران جاسازی شود. ما همچنین توابع هدف دیگری را معرفی می کنیم تا رمزگذار فضای پنهان ایده آل را یاد بگیرد. تمام نتایج تجربی ذهنی و عینی نشان میدهند که روش پیشنهادی ما پیشرفته است [۸].

مقاله ۹:

Machine Speech Chain with One-shot Speaker Adaptation

زنجیره گفتار ماشینی با بلندگوی تک شات

در کار قبلی، ما یک مدل زنجیره گفتار حلقه بسته مبتنی بر یادگیری عمیق توسعه دادیم، که در آن معماری اجزای تشخیص خودکار گفتار (ASR) و ترکیب متن به گفتار (TTS) را قادر میسازد تا عملکرد خود را به طور متقابل بهبود بخشند. این امر با آموزش دو بخش با استفاده از داده های برچسب دار و بدون برچسب به یکدیگر انجام شد. این رویکرد می تواند به طور قابل توجهی عملکرد مدل را در مجموعه داده گفتاری تک گوینده بهبود بخشد، اما تنها افزایش جزئی در وظایف چند گوینده به دست می آید. علاوه بر این، این مدل هنوز قادر به مدیریت بلندگوهای دیده نشده نیست. در این مقاله، ما یک مکانیسم زنجیره گفتار جدید را با ادغام یک مدل تشخیص بلندگو در داخل حلقه ارائه میکنیم. ما همچنین پیشنهاد میکنیم قابلیت TTS را برای مدیریت بلندگوهای دیده نشده با اجرای تطبیق بلندگوی تک شات افزایش دهیم. این کار TTS را قادر میسازد تا ویژگیهای صوتی را از یک بلندگو به بلندگوی دیگر تنها با یک نمونه بلندگوی تک شات، حتی از متنی بدون هیچ گونه اطلاعات بلندگو، تقلید کند. در مکانیسم حلقه زنجیره گفتار، ASR همچنین از توانایی یادگیری بیشتر ویژگیهای بلندگوی دلخواه از شکل موج گفتار تولید شده بهره میبرد که منجر به بهبود قابل توجهی در نرخ تشخیص میشود [۹].

مقاله ۱۰:

Voice Conversion Using Sequence-to-Sequence Learning of Context Posterior Probabilities

تبدیل صدا با استفاده از یادگیری توالی به ترتیب احتمالات پسین زمینه

تبدیل صدا (VC) با استفاده از یادگیری توالی به دنباله احتمالات پسین زمینه پیشنهاد شده است. VC معمولی با استفاده از احتمالات پسین زمینه مشترک، پارامترهای گفتار هدف را از احتمالات پسین زمینه برآورد شده از پارامترهای گفتار منبع پیش بینی می کند. اگرچه VC معمولی می تواند از داده های غیر موازی ساخته شود، تبدیل فردیت گوینده مانند ویژگی آوایی و سرعت گفتار موجود در احتمالات پسین دشوار است زیرا احتمالات خلفی منبع مستقیماً برای پیش بینی پارامترهای گفتار هدف استفاده می شود. در این کار، ما فرض میکنیم که دادههای آموزشی تا حدی شامل دادههای گفتاری موازی است و یادگیری ترتیب به دنباله را بین احتمالات پسین منبع و هدف پیشنهاد میکند. مدلهای تبدیل تبدیل غیرخطی و با طول متغیر را از توالی احتمال منبع به هدف انجام میدهند. علاوه بر این، ما یک الگوریتم آموزشی مشترک برای ماژول ها پیشنهاد می کنیم. برخلاف VC معمولی، که به طور جداگانه تشخیص گفتار را که احتمالات پسین را تخمین می زند و سنتز گفتاری که پارامترهای گفتار هدف را پیش بینی می کند، آموزش می دهد، روش پیشنهادی ما به طور مشترک این ماژول ها را همراه با ماژول های تبدیل احتمال پیشنهادی آموزش می دهد. نتایج تجربی نشان میدهد که رویکرد ما از VC معمولی بهتر عمل میکند [۱۰].

جدول مزایا و معایب ۱۰ مقاله مرتبط با موضوع پروژه:

مقاله	مزایا	معایب
مقاله اصلی پروژه	۱. سخنران مبدأ و هدف حتی نیازی به دیده شدن درطول آموزش ندارند. ۲. تبدیل صدای تک شد بدون هیچ نظارتی انجام می‌شود. ۳. مدل می‌تواند جاسازی سخنران را به‌عنوان یک اثر جانبی بیاموزد.	کیفیت پایین صدا
مقاله ۱	هیچ‌گونه فرایند آموزشی در این فرایند استفاده نمی‌شود. با الهام از GST انجام می‌شود و نتایج خوبی از نظر شباهت گفتار و گویش نشان داده است.	به مقدار زیادی داده گفتاری نیاز دارد.
مقاله ۲	هیچ نظارتی در این فرایند وجود ندارد	چون تفکیک محتوا و اطلاعات سخن را تنها با از دست دادن بازسازی انجام می‌شود، ممکن است برخی اطلاعات دیگر نیز از دست بروند.
مقاله ۳	۱. الگوریتم AdaIN-VC از یک رمزگذار محتوا و یک رمزگذار بلندگو استفاده می‌کند. درحالی‌که الگوریتم AGAIN-VC تنها از یک رمزگذار و یک فعال‌سازی برای هدایت و آموزش استفاده می‌کند. ۲. جهت جلوگیری از نشت اطلاعات گوینده به درون جاسازی‌های محتوا ابعاد را کاهش می‌دهد.	AGAIN-VC می‌تواند به نوعی به کیفیت سنتز آسیب بزند.

مقاله ۴	۱. ترکیب Trans-StarGAN-VC با AdaIN-VC می‌تواند کیفیت صدا را افزایش دهد. ۲. جزئیات هارمونیک کامل‌تری را ارائه می‌دهد. ۳. در TransNet بسیاری از اتصالات میان‌بر می‌تواند عملیات آموزش را تسریع کند.	_____
مقاله ۵	در تبدیل صدای تک شات شباهت به اندازه کافی خوب نیست و نمونه تطبیق یا وزن می‌تواند به بهبود آن کمک کند.	_____
مقاله ۶	۱. ماژول عادی‌سازی نمونه تطبیقی (AdaIN) در جدا کردن محتوا از صدا در یادگیری مدل پیشنهادی برای سنتز انیمیشن‌های سه‌بعدی از طریق صدا استفاده می‌شود. ۲. دقت و طبقه‌بندی بهتر عمل می‌کند.	_____
مقاله ۷	بهبود کارایی داده‌های مدل و دستیابی به یک تبدیل صوتی غیرموازی مبتنی بر STARGAN برای تعداد زیادی از بلندگوها با نمونه‌های آموزشی.	فقدان داده‌های موازی و تعداد کم بلندگو
مقاله ۸	در چارچوب صوتی ClsVC مشکل جدا کردن کامل سخنران و اطلاعات محتوا از سیگنال گفتار را برطرف می‌کند.	وجود فضای پنهان
مقاله ۹	بهبود عملکرد مدل درمجموعه داده گفتاری تک گوینده	قادر به مدیریت بلندگوهای دیده نشده نیست.

مقاله ۱۰	تبدیل صدا با استفاده از یادگیری توالی به‌طور مشترک ماژول ها را همراه با ماژول های تبدیل احتمال پیشنهادی آموزش می‌دهد.	تبدیل فردیت گوینده مثل ویژگی‌های آوایی و سرعت گفتار موجود در احتمالات پسین دشوار است.

- **منابع:**

[1] Hui Lu, Zhiyong Wu, "One-shot Voice Conversion with Global Speaker Embeddings", INTERSPEECH, pp. 15-19, 2019.

[2] Da-Yi Wu, Hung-yi Lee, " One-Shot Voice Conversion by Vector Quantization", IEEE, 2020.

[3] Yen-Hao Chen, Da-Yi Wu, "Again-VC: A One-Shot Voice Conversion Using Activation Guidance and Adaptive Instance Normalization", IEEE, 2021.

[4] Yanping Li, Zhengtao He, "High-Quality Many-to-Many Voice Conversion Using Transitive Star Generative Adversarial Networks with Adaptive Instance Normalization", Journal of Circuits, Systems and Computers, vol. 30, 2021.

[5] Shengjie Huang, Mingjie Chen, "WINVC: One-Shot Voice Conversion with Weight Adaptive Instance Normalization", springer, pp 559–573, 2021.

[6] Che-Jui Chang, "Disentangling audio content and emotion with adaptive instance normalization for expressive facial animation synthesis", wiley, vol. 33, 2022.

[7] Mingjie Chen, Yanpei Shi, "Towards Low-Resource Stargan Voice Conversion Using Weight Adaptive Instance Normalization", IEEE, 2021.

[8] Tang huaizhen, xulong Zhang, "ClsVC: Learning Speech Representations with two different classification tasks", 2022.

[9] Andros Tjandra, Sakriani Sakti, "Machine Speech Chain with One-shot Speaker Adaptation", arXiv, 2018.

[10] Hiroyuki Miyoshi, Yuki Saito, "Voice Conversion Using Sequence-to-Sequence Learning of Context Posterior Probabilities", arXiv, 2017.