بسمه تعالي

مديريت تحصيلات تكميلي دانشگاه

**پيشنهاديه پايان نامه دوره کارشناسی ارشد**



شماره:

تاريخ:

اين فرم بايد توسط دانشجوی كارشناسي ارشد در يك نسخه با هماهنگی استاد راهنما به **صورت تايپ شده** تکميل گردد

## مشخصات دانشجو:

**نام و نام خانوادگی دانشجو: رسول مومنی زاده شماره دانشجويي: 39713161021**

**رشته تحصیلی/ گرایش: مهندسی کامپیوتر/ نرم افزار دانشکده: فنی مهندسی سال ورود: 1397**

**بورسیه**  **آزاد**  **مامور به تحصیل**

**تلفن: 09392535042**  **E-mail: rasoulmomenizade@gmail.com**

1. **مشخصات اساتيد راهنما و مشاور:**

**الف-مشخصات اساتيد راهنما:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| رديف | **نام و نام خانوادگی** | | | **آخرين مدرک تحصيلي** | | **سال اخذ آخرين مدرک تحصيلي** | | **مرتبه علمي** | **تخصص** | | | **درصد مشاركت** |
| 1- | | دکترای تخصصی | 1393 | | استادیار | | پردازش تصویر و سیگنال | | | 100% |
|  | |  | |  | |  | |  |  | | |  |

**ب-مشخصات اساتيد مشاور:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| رديف | **نام و نام خانوادگی** | **گروه آموزشی** | **دانشکده** | **مرتبه** | **تخصص** | **ملاحظات** |
|  |  | | | | | |

|  |  |
| --- | --- |
| **Email:** | **آدرس و تلفن:**  **محل خدمت:** |

## 3-مشخصات پايان نامه:

## الف-عنوان پايان نامه:

**1-فارسی: تشخیص احساس سیگنال گفتار با استفاده از ویژگی­های فرکانسی-زمانی و الگوریتم­های یادگیری عمیق**

**2-انگلیسی: Emotion Speech signal detection using frequency-temporal features and deep learning algorithms**

## ب-نوع تحقيق:

**1- تجربی**  **2- نظری** **3- تقاضا محور با ارائه مستندات و تاییدیه شورای پژوهش**

**مقدمه**

استفاده از تشخیص احساس از گفتار می­تواند در جهت استخراج معانی مفید واقع شود. در زمینه ارتباط بین انسان و ماشین، تشخیص حس از گفتار امری ضروری است. از جمله این کاربردها نظیر آموزش مجازی، ایجاد موتورهای جستجوی احساسی، برنامه­های کامپیوتری، پیام­های صوتی بر اساس حس گوینده و برنامه­های کاربردی مشابه است. سیگنال گفتار سریع­ترین و طبیعی­ترین روش برای ارتباط متقابل میان انسان هاست. همین موضوع انگیزه­ی ایجاد دانشی به نام تشخیص حس گفتار شده است که هدف آن استخراج احساس گوینده از میان گفتار اوست. به علاوه تشخیص حس گوینده درک بهتر معنا می­کند و کارایی سیستم­های پردازش گفتار را افزایش می­دهد. اگر چه این دانش به صراحت در دسترس نمی­باشند، اما مغز انسان قادر به تشخیص این اطلاعات از جمله حالت عاطفی گوینده، با استفاده از شکل موج گفتار است. این توانایی انسان در تشخیص وضعیت احساسی گوینده از روی گفتارش نسبت به گویندگان مختلف نیز صادق بوده، و انسان قادر به تشخیص موفقیت­آمیز آن در بسیاری از موارد است. با توجه به عدم وجود تکنیکی ایده­آل، از تکنیک­های طبقه­بندی الگو، برای تصمیم­گیری در مورد حالت عاطفی براساس ویژگی­های انتخاب شده استفاده می­شود. این ویژگی­ها به طور کلی به دو گروه سطح بالا و یا سطح پایین طبقه بندی می­شوند. ویژگی­های سطح پایین خواص صوتی، آوابی و با طیفی از سیگنال گفتار را بدون در نظر گرفتن محتوای زبانی، توصیف می­کنند. از سوی دیگر ویژگی­های سطح بالا محتوای زبانی را بدون در نظر گرفتن هر گونه تغییری در پارامترهای صوتی سیگنال گفتار، در نظر می­گیرند، اگر چه شواهد نشان می­دهد که هر نوع از این ویژگی­ها حاوی اطلاعات احساسی خاصی، به منظور محدود کردن پیچیدگی سیستم­های تشخیص احساسات هستند، تعداد زیادی از سیستم های تشخیص خودکار احساسات[[1]](#footnote-1) از اطلاعات معنایی با زبانی استفاده نکرده و صرفا روی ویژگی­های صوتی، آوایی و یا طیفی تاکید دارند.

روش­هاي آماري معمولا شامل استفاده از الگوريتم­هاي يادگيري ماشين تحت نظارت است كه در آن مجموعه­اي از داده­هاي حاوي اطلاعات به الگوريتم­هاي سيستم براي يادگيري و پيش­بيني انواع احساسات آموزش داده مي­شود. اين رويكرد به طور معمول شامل دو مجموعه داده است كه شامل مجموعه آموزشي و مجموعه آزمايشي مي­شود. اما يكي از چالش­ها در دستيابي به نتايج خوب در فرايند طبقه­بندي، نياز به يك مجموعه آموزشي كافي است. بعضي از الگوريتم­هاي يادگيري ماشين عبارتند از: ماشین بردار پشتیبانی، دسته­بندی کننده بیز ساده و همچنین یادگیری عمیق که بخشی از یادگیری ماشین است، نیز به طور گسترده­ای در شناخت احساسات مورد استفاده قرار می­گیرند. الگوریتم شناخته شده یادگیری عمیق شامل معماری­های مختلفی از شبکه عصبی مصنوعی مانند شبکه عصبی کانولوشنالی، شبکه عصبی بازگشتی حافظه طولانی کوتاه مدت می­باشند که محبوبیت رویکردهای یادگیری عمیق در زمینه شناسایی احساسات عمدتا به موفقیت آن در کاربردهای مرتبط مانند بینایی ماشین، تشخیص گفتار و پردازش طبیعی زبان مربوط می­شود. لذا در این تحقیق با استفاده از ویژگی های شبکه عصبی کانولوشن به استخراج احساس از گفتار در زبان فارسی پرداخته می­شود.

**پیشینه تحقیق**

پرابها و همکاران (2019) طی پژوهشی به استخراج احساس از گفتار با استفاده از یادگیری عمیق پرداخته­اند. روش مورد استفاده در این تحقیق شبکه عصبی بازگشتی می­باشد. این تکنیک­ها به صورت ترکیبی یا مستقل بر اساس حوزه دامنه کاربرد استفاده می­شوند. تمرکز این تحقیق بر روی طعم­های مختلف روش­های یادگیری عمیق است که در کاربردهای مختلف تحلیل احساسات در سطح جمله و جنبه / هدف مورد استفاده قرار می­گیرد. علاوه بر این، مزایا و اشکال روشها همراه با پارامترهای عملکرد آنها مورد بحث قرار گرفته است. یائو (2018) با ترکیب دو طبقه بندی کننده ماشین بردار پشتیبان و مدل های مخفی مارکوف، نرخ تشخیص بالاتری نسبت به ماشین بردار پشتیبان به تنهایی به دست آورد و نرخ تشخیص ۶۵ درصد را گزارش نمود. کومار و همکاران (2017) با استفاده از یادگیری عمیق به تشخیص احساس از ویژگی­های گفتار پرداختند. در این تحقیق آنها روشی جدید مبتنی بر شبکه پیچشی را جهت تشخیص احساس از گفتار ارائه دادند. این روش باعث افزایش هفت درصدی در دقت تشخیص احساس از گفتار شده است. گوآراو (2017) به بررسی پارامتر نرخ صحبت، در تشخیص احساسات از گفتار می­پردازد. نتایج نشان می­دهد این پارامتر طبقه بندی را آسان می­کند. این گروه از پژوهشگران با استفاده از طبقه بندی کننده MLP نرخ تشخیص ۶۱ / ۴۲ درصد را گزارش کردند.

حریمی (1393) به بررسی تشخیص احساس به کمک پردازش چندبعدی گفتار پرداخته است. از الگوریتم‌های انتخاب متوالی رو به جلو و الگوریتم ژنتیک نیز بعنوان یک مرحله مکمل، برای انتخاب بردار ویژگی بهینه بکار گرفته شدند. به منظور طبقه‌بندی احساس‌های گسسته و پیوسته در پایگاه داده‌ی برلین و VAM از طبقه‌بند و رگرسیون مبتنی بر ماشین بردارهای پشتیبان استفاده شده است. بر اساس نتایج آزمایش‌ها، ویژگی‌های دینامیکی غیرخطی در کاهش تداخل احساس‌های خشم و خوشحالی و به تبع آن در افزایش نرخ بازشناسی بسیار موثر می‌باشند. در پژوهشی احمدی (1396) به تشخیص احساس از روی سیگنال گفتار با استفاده از ویژگی های آمارگان مرتبه بالا پرداخته است. در این سیستم از ویژگی­هایی هم‌چون آمارگان مرتبه بالا برای آموزش و یادگیری استفاده شده است. در روش پیشنهادی برای اولین بار استخراج ویژگی‌های کارآمدی هم‌چون آمارگان مرتبه سوم صورت گرفت و در نتیجه دقت تشخیص احساسات از روی گفتار بهبود داده شد. نتایج نشان دهنده‌ی کارایی روش پیشنهادی در تشخیص احساس در مقایسه با روش‌های موجود دیگر می‌باشد. دامی و همکاران (1396) در پژوهشی به استخراج جنبه برای تحلیل احساسات با شبکه عصبی کانولوشن بسیار عمیق پرداختند. در این مقاله به بررسی و ارایه یک رویکرد یادگیری بسیار عمیق جهت استخراج جنبه برای تحلیل احساسات پرداخته شده است. نتایج ارزیابی حاصل از طبقه بندی گروهی همراه با مدل تعبیه کلمات مبتنی بر یادگیری بسیار عمیق برای تحلیل احساسات نشان داد که روش پیشنهادی در مقایسه با روشهای پایه عملکرد به مراتب بهتری دارد.

**اهداف و کاربرد مورد انتظار**

در این پژوهش با استفاده از یادگیری عمیق به تشخیص احساس از گفتار زبان فارسی خواهیم پرداخت. برای مثال می­توان حالت­های عاطفی افراد و میزان اثربخشی احساس بر صدا را شناسایی کرد. از کاربردهای منحصر به فرد این تحقیق می­توان به تولید متا دیتا به منظور بهینه­سازی فرآیند جست و جو و مکانیزه کردن سامانه­های آرشیو استفاده کرد. در اموری که ارتباط متقابل میان انسان و ماشین وجود دارد، تشخیص حس گفتار امری ضروری به نظر می­رسد؛ زیرا با تشخیص احساس پاسخ این سیستم ها به کاربر مناسبتر خواهد بود. از جمله­ی این کاربردها ایجاد موتور جست و جوی احساسی و دسته بندی­های نامه­های الکترونیک و پیام های صوتی بر اساس حس گوینده است. از دیگر کاربردهای آن نیز می­توان به سیستم ترجمه­ی خودکار اشاره کرد که در آن حالت احساسی فرد نقش مهمی در ارتباط طرفین و انتقال مفاهیم و معانی دارد. همچنین در کابین خلبان، مشخص شده است که سیستم های تشخیص گفتار آموزش دهنده داده­شده با گفتار تأکیدی، کارایی بهتری نسبت به سیستم­های تشخیص گفتار آموزش داده شده با گفتار معمولی دارد. تشخیص حس از گفتار در مراکز مخابراتی نیز مفید است؛ هدف اصلی در این کاربردها تشخیص احساس ناامیدی یا عصبانیت گوینده است. به علاوه پزشکان با استفاده از سیستم تشخیص حس گفتار می­توانند برخی از بیماری­های خاص مثل پارکینسون و ... را تشخیص دهند. همچنین تشخیص حس از گفتار در حوزه­ی سرگرمی نیز برای تولید اسباب بازی­های هوشمند و بازی­های رایانه­ای کاربرد فراوانی دارد.

**روش پیشنهادی**

در این تحقیق به منظور تشخیص احساس از گفتار از روش یادگیری عمیق با استفاده از الگوریتم­های شبکه عصبی کانولوشن استفاده خواهد گردید. در این تحقیق ما با استفاده از روشی مانند طیف سنجی شبکه عصبی کانولوشن به تشخیص صدا خواهیم پرداخت. طیف سنجی نوعی ارائه از گفتار در طول زمان و فرکانس می­باشد. فیلتر­های شبکه کانولوشن دو بعدی در تشخیص ورودی­های ویژگی دو بعدی کمک می­کنند. چنین ویژگی قادر نخواهد بود تا در زمینه تشخیص از متن به کار رود. طیف سنجی­ها، که حاوی اطلاعات اضافی هستند، در تلاش برای بهبود شناخت احساسات می­باشند و مهارت­های مهمی را ارائه می­دهند. به منظور جمع آوری داده­ها یعنی صداهای وروری در این تحقیق که به صورت صداهای فارسی هستند، از دیتابیس فارسی گفتار احساس[[2]](#footnote-2) استفاده خواهد شد. این دیتاست تنها دیتاست معتبر گفتار زبان فارسی می­باشد که در دسترس هست.

**جدول زمانبندی**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **زمان(ماه)**  **شرح فعالیت** | اول | دوم | سوم | چهارم | پنجم | ششم |
| مطالعات كتابخانه‌اي | ◼ | ◼ | ◼ | ◼ | ◼ | ◼ |
| جمع آوری اطلاعات |  | ◼ | ◼ |  |  |  |
| تجزيه و تحليل داده‌ها |  |  | ◼ | ◼ |  |  |
| نتيجه‌گيري |  |  |  | ◼ | ◼ |  |
| نگارش پايان‌نامه |  |  |  |  | ◼ | ◼ |

**منابع**

احمدی، سحر. (1396). تشخیص احساس از روی سیگنال گفتار با استفاده از ویژگی های آمارگان مرتبه بالا. پایان نامه کارشناسی ارشد. دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان.

اسماعیلیان، زینب. (1392). تشخیص نوع احساس از روی گفتار با استفاده از ویژگی‌های زمان-فرکانسی. پایان نامه کارشناسی ارشد. دانشگاه آزاد اسلامی واحد شاهرود.

دامی، سینا. امیررضا عرفانی راد. (1396). استخراج جنبه برای تحلیل احساسات با شبکه عصبی کانولوشن بسیار عمیق، *چهارمین کنفرانس بین المللی مطالعات نوین در علوم کامپیوتر و فناوری اطلاعات*، مشهد، دانشگاه صنعتی سجاد.

حریمی، علی. (1393). تشخیص احساس به کمک پردازش چند بعدی گفتار. رساله دکتری. دانشگاه سمنان.

یوسفی نژاد، رسول. باقر نایینی، بابک. شفیعیان، معصومه. (1394). تشخیص احساس از سیگنال گفتار با استفاده از موجگ بیونیک. نشریه عملی ترویجی صوت و ارتعاش. 9 (3): 27-36.

Chen LS, Tao H, Huang TS, Miyasato T, Nakatsu R. (1998). “Emotion recognition from audiovisual information.” *InMultimedia Signal Processing, IEEE Second Workshop*. 6(5): 83-88.

Gaurav, M. (2017). “Performance Analysis of spectral and prosodic features and their fusion for emotion recognition in speech.” *IEEE. SLT*. 37(4): 78-90.

Hassan, A., Damper, R. (2012). “Classification of emotional speech using 3DEC hierarchical classifier.” *speech communication* 54. 903-916.

Hubner, D., Vlasenko, B., Bock, R. (2015). “The Performance of The Speaking Rate Parameter in emotion recognition from speech.” *IEEE, International conference on Multimedia and Expo Workshops*.

Li X. (2017). “SPEech Feature Toolbox (SPEFT) Design and Emotional Speech Feature Extraction” (Doctoral dissertation, Faculty of Graduate School, Marquette University).

Nicholson J., Takahashi K., Nakatsu R. (2016). “Emotion recognition in speech using neural networks.” Neural computing & applications. 9(4): 290-6.

Pan Y., Shen P, Shen L. (2012). “Feature extraction and selection in speech emotion recognition.” *Proceeding of the onlinepresent*. 7(3): 64-9.

Yu F., Chang E., Xu Y.Q., Shum H.Y. (2015). “Emotion detection from speech to enrich multimedia content.” *Advances in multimedia information processing*. 550-7.

Yang, B., Lugger, M. (2010). “Emotion recognition from speech signals using new harmony features.” *signal processing* 90. 1415-1423.

Yao, J., Zhang, Y. (2018). “Bionic Wavelet Transform: A New Time-Frequency Method Based on an Auditory Model.” *Proc. IEEE, Trans.Biomedical Eng*. 48(8): 856-863.

|  |
| --- |
| **امضاء دانشجو: رسول مومنی زاده تاريخ: 26/09/1398**  **توجه:** كليه حقوق مادی و معنوی مرتبط بر نتايج تحقيق اين پايان نامه ( اعم از چاپ مقاله، ارائه به بخش صنعت و ...) متعلق به دانشگاه اراک است و انتشار نتايج حاصل از آن تابع مقررات دانشگاهی است و با موافقت استاد راهنما صورت می‌گيرد. |
| امضاء استاد راهنما: سرکار خانم دکتر مریم مومنی تاريخ: 26/09/1398 |
| امضاء استاد مشاور: تاريخ: / /1398 |

## تصويب پيشنهاديه پايان نامه كارشناسي ارشد

موضوع پايان نامه آقاي / خانم رسول مومنی زاده دانشجوي دوره كارشناسي ارشد رشته **مهندسی کامپیوتر**  با شماره دانشجويي 39713161021 تحت عنوان: **تشخیص احساس سیگنال گفتار با استفاده از ویژگی­های فرکانسی-زمانی و الگوریتم­های یادگیری عمیق**

## در شورای گروه آموزشي مهندسی کامپیوتر در تاريخ / / 1398 بررسی و به تصويب رسيد.

## امضا مدير گروه کامپیوتر

## موضوع پايان نامه اخیر در شوراي تحصيلات تکميلی دانشکده فنی مهندسی در تاريخ / / 1398 بررسی و به تصويب رسيد.

## امضا معاون آموزشی دانشكده فنی و مهندسی

1. Automatic emotion recognition [↑](#footnote-ref-1)
2. PESD [↑](#footnote-ref-2)