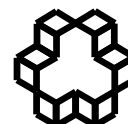


فرم ارائه پروژه کارشناسی دانشکده مهندسی کامپیوتر



تأسیس ۱۳۰۷
دانشگاه صنعتی خواجه نصیرالدین طوسی
دانشکده مهندسی کامپیوتر

مشخصات استاد راهنما :

گروه آموزشی	نام و نام خانوادگی استاد راهنما	مرتبه علمی		
		استاد	دانشیار	استادیار
اینترنت اشیا	دکتر کعبه یعقوبی			✓

عنوان پروژه: مدل تشخیص احساسات چندوجهی: ادغام سیگنال‌های صوتی، متنی و تصویری

Multimodal Emotion Recognition Model: Integration of Audio, Textual, and Visual Signals

(تمام قسمت‌ها به دقت تایپ شود)

نوع پروژه و سقف نمره	تعداد واحد	محل انجام پروژه
طراحی و ساخت ^۱ و کار عملی	۲۰	امضاء سرپرست آزمایشگاه
شبیه سازی یا طراحی	۱۹	
مطالعاتی	۱۷	

توضیحات :

۱۹۵۷۱
۸۱۵۱

امضاء استاد راهنما :

مشخصات دانشجو

ردیف	نام و نام خانوادگی دانشجو	شماره دانشجویی	گرایش	تعداد واحد گذرانده	معدل
۱	ارش نبئی	۹۹۳۱۱۴۳		۱۲۷	۱۴.۹۷
۲					
۳					

یادآوری می‌شود امضاء دانشجو به منزله ثبت نام پروژه بوده و عواقب ناشی از عدم پی‌گیری روند پروژه به عهده شخص دانشجو خواهد بود.

۱۹۵۷۱
۸۱۵۱

امضاء دانشجو :

۱۹۵۷۱
۸۱۵۱

امضاء دانشجو:

پروژه فوق‌الذکر در جلسه گروه مورخ به تصویب رسید. این پروژه براساس شیوه نامه حمایت از پروژه های کارشناسی عملی و تجربی مصوبه ۹۵/۰۱/۱۷ هیات رئیسه دانشگاه از نوع الف (پروژه های که موضوع اصلی آن ساخت، پیاده سازی عملی و یا کار آزمایشگاهی است) تعیین شد: ○
از نوع ب (پروژه های که موضوع اصلی آن ساخت نبوده اما برای انجام یا تایید بخشی از ایده ها یا نتایج آن، نیاز به ساخت، پیاده سازی عملی یا کار آزمایشگاهی است) تعیین شد: ○

امضاء مدیر گروه :

تاییدیه آموزش دانشکده مهندسی کامپیوتر

با توجه به واحدهای گذرانده و وضعیت آموزشی دانشجو، اخذ پروژه از نظر این اداره در ترم بلا مانع است.

امضاء سرپرست آموزش دانشکده مهندسی کامپیوتر

- تعریف پروژه:

در دنیای امروزی، تعاملات انسانی به طور گسترده‌ای به کمک فناوری‌های مبتنی بر هوش مصنوعی تحلیل می‌شوند. یکی از مهم‌ترین جنبه‌های تعاملات انسانی، درک احساسات است که در کاربردهایی مانند ربات‌های اجتماعی، خدمات مشتریان، مراقبت‌های بهداشتی و تحلیل رسانه‌های اجتماعی بسیار مهم است.
پروژه‌ی حاضر با هدف تشخیص احساسات انسانی از سه منبع مختلف (صدا، متن، تصویر) به صورت همزمان طراحی خواهد شد. برخلاف روش‌های سنتی که تنها یک نوع داده (مثلاً متن یا صدا) را تحلیل می‌کنند، این پروژه از مدل‌های یادگیری چندوجهی (Multimodal Learning) [۱] استفاده می‌کند تا اطلاعات مختلف را با هم ترکیب کرده و به دقت بالاتری در شناسایی احساسات برسد.

^۱ هر آنچه که بعنوان یک محصول یا بسته نرم افزاری توسط داوران شناخته شود.

در میان انواع مدل‌های چندوجهی، Transformer-Based Architectures بسیار موثر می‌باشد. این مدل از مکانیسم‌های توجه برای تراز و ادغام ویژگی‌ها از روش‌های مختلف استفاده می‌کند و اطمینان حاصل می‌کند که مرتبط ترین اطلاعات گرفته می‌شود. همچنین روش Lightweight Adaptive Networks استخراج ویژگی و حذف نویز را بهینه می‌کند و این روش برای سناریوهایی با داده‌های محدود مناسب می‌باشد [۲]. برای این منظور، از Attention Mechanism جهت وزندهی پویا به هر منبع اطلاعاتی استفاده خواهد شد تا سیستم بتواند تشخیص دهد که در هر لحظه کدام منبع (تصویر، متن یا صوت) نقش بیشتری در تعیین احساس دارد.

- روند نما یا مراحل انجام پروژه:

۱. انتخاب و آماده سازی دیتاست

استفاده از دیتاست‌های استاندارد چند وجهی مانند IEMOCAP و FER2013 پردازش داده‌ها شامل استخراج ویژگی‌های صوتی (Mel Spectrogram)، پردازش متنی (BERT) و استخراج ویژگی‌های تصویری (CNN)

۲. طراحی و پیاده سازی مدل‌های پردازش مستقل

طراحی ماژول پردازش صوت با استفاده از CNN
طراحی ماژول پردازش متن با استفاده از BERT
طراحی ماژول پردازش تصویر با استفاده از CNN

۳. طراحی ماژول ترکیب (Fusion Module)

پیاده سازی Attention Mechanism برای تعیین میزان اهمیت هر منبع
ترکیب خروجی‌های تصویر، صوت و متن برای تولید یک بردار ویژگی نهایی

۴. آموزش و ارزیابی مدل

آموزش مدل با استفاده از دیتاست‌های چند وجهی
مقایسه دقت با روش‌های قبلی (Unimodal vs Multimodal)
بررسی معیارهایی مثل Accuracy, F1 score, Weighted Accuracy, Weighted F1

۵. بهینه سازی و تست نهایی

بهینه سازی مدل با استفاده از تکنیک‌هایی مانند Regularization
تست مدل روی داده‌ها و بررسی عملکرد آن

- خروجی مشخص پروژه (این خروجی‌ها در جلسه دفاعیه ارائه خواهد شد):

۱. مدل نهایی برای تشخیص احساسات چند وجهی

این مدل قادر خواهد بود تا با استفاده از تصویر، صوت و متن به طور همزمان احساسات انسانی را تشخیص دهد.
از Fusion Mechanism برای ترکیب هوشمند داده‌ها استفاده خواهد کرد.

۲. مقایسه عملکرد مدل با روش‌های موجود

بررسی و مقایسه عملکرد مدل چندوجهی با مدل‌های تک وجهی (متن، صوت یا تصویر به تنهایی)
ارائه نمودارهای Accuracy، Recall و F1-score

۳. مستندات و گزارش نهایی پروژه

ارائه مستندات کامل درباره‌ی مراحل پیاده‌سازی، چالش‌ها، و تحلیل نتایج
نمودارهای عملکرد مدل

- مواد و تجهیزات مورد نیاز:

۱. سخت افزار

دسترسی به یک واحد پردازشی مناسب جهت یادگیری مدل (ترجیحا یک واحد GPU نسبتا قوی)
مقدار RAM کافی برای بارگذاری مدل و پردازش داده‌ها (حداقل ۱۶ گیگابایت)
فضای ذخیره سازی مناسب برای نگهداری مدل‌ها و داده‌ها

۲. زبان فریمورک‌های و کتابخانه‌های مورد نیاز

زبان برنامه نویسی مناسب (python)
استفاده از فریمورک های یادگیری و ابزار های پردازش داده (scikit ,OpenCV)

۳. ابزارهای اجرا و تست
یک ویرایشگر کد مانند vs code یا Notebook

۴. سرویس های ابری
امکان استفاده از پلتفرم های ابری برای اجرای مدل در صورت نیاز به توان پردازشی بیشتر (google cloud)

- برآورد مالی پروژه:

- منابع

- [1] Makhmudov, F., Kultimuratov, A., & Cho, Y.-I. Enhancing Multimodal Emotion Recognition through Attention Mechanisms in BERT and CNN Architectures. *Applied Sciences*, 14(10), 4199 (2024).
[2] Guo, X., Zhai, S. A novel adaptive lightweight multimodal efficient feature inference network ALME-FIN for EEG emotion recognition. *Cogn Neurodyn* 19, 24 (2025).

اسامی اساتید ارزیاب به ترتیب اولویت:

- ۱- .
۲-
۳-
۴-

مراحل انجام کار:

- ۱- تعریف و تأیید پروژه توسط استاد راهنما ، امضاء آذ توسط دانشجو
۲- اخذ تأییدیه وضعیت آموزشی از اداره آموزش دانشکده توسط دانشجو و تحویل فرم تکمیل شده به دفتر گروه ها
۳- طرح ، بررسی و تصویب پروژه در جلسه گروه
۴- ارسال فرم پروژه تصویب شده به آموزش جهت ثبت نام و سپس عودت به دفتر گروه ها جهت ثبت در سامانه گلستان و ضبط در سوابق تا زمان دفاع.