



پروژه تشخیص احساسات با داده‌های چندوجهی

هدف پروژه

هدف این پروژه طراحی و پیاده‌سازی یک سیستم تشخیص احساسات است که با بهره‌گیری از یادگیری ماشین و یادگیری عمیق، و با استفاده از دو مدالیته‌ی متن و گفتار، احساس گوینده را پیش‌بینی کند. در این مسئله، هر نمونه داده شامل متن گفتار^۱ و سیگنال صوتی بوده و مدل باید با ترکیب هر دو، طبقه‌بندی احساسات را انجام دهد. تمرکز پروژه بر ساخت یک خط لوله‌ی انتها به انتها^۲ از پیش‌پردازش و استخراج ویژگی تا آموزش و ارزیابی است، به گونه‌ای که مدل نهایی عملکرد قابل قبولی روی داده‌های ارزیابی ارائه دهد.

جزئیات پروژه

در این پروژه انتظار می‌رود یک مدل تشخیص احساسات بر پایه‌ی داده‌های چندوجهی^۳ ارائه شود. به طوری که برای هر نمونه، متن گفتار و سیگنال صوتی به صورت متناظر در دسترس بوده و خروجی سیستم یک Emotion Label باشد. از نظر پردازش متن، دست شما برای استفاده از مدل‌های از پیش‌آموزش‌دیده^۴ باز است و می‌توانید بسته به طراحی خود از وزن‌های آماده استفاده کنید، آن‌ها را ثابت^۵ نگه دارید یا با راهبردهای ریزتنظیم^۶ برای مسئله‌ی تشخیص احساسات بهینه‌سازی کنید، با این حال لازم است نحوه‌ی استفاده و دلیل انتخاب این راهبردها در گزارش شفاف توضیح داده شود. برای مدالیته‌ی صوت نیز استفاده از انتقال یادگیری^۷ مجاز است، زیرا داده‌های گفتاری معمولاً نسبت به تغییرات گوینده، لهجه، شدت صدا، نویز محیط، کیفیت ضبط و شرایط واقعی حساس‌اند و بهره‌گیری از نمایش‌های آموخته‌شده می‌تواند به تعمیم پذیری^۸ کمک کند. با این وجود، صرفاً تکیه بر یک مدل آماده‌ی «بدون طراحی و بدون آموزش/انطباق» کافی نیست و شما باید معماری کلی و روش‌شناسی پروژه را خودتان طراحی کنید، اجزای لازم برای تطبیق با وظیفه را تعریف کنید و فرآیند آموزش و ارزیابی را به صورت قابل بازتولید^۹ ارائه دهید. ارزیابی باید روی تقسیم‌بندی استاندارد داده انجام شود و عملکرد با معیارهای مناسب گزارش و تحلیل گردد، علاوه بر شاخص‌هایی مانند Accuracy / Precision / Recall / F1، ارائه‌ی تحلیل‌های تکمیلی مانند Confusion Matrix و تحلیل الگوهای خطا می‌تواند کیفیت گزارش را افزایش دهد.

^۱ Speech Transcript

^۲ End-to-End Pipeline

^۳ Multimodal

^۴ Pretrained Language Models

^۵ Freezing

^۶ Fine-tuning Strategies

^۷ Transfer Learning

^۸ Robustness / Generalization

^۹ Reproducibility

دادگان

برای این پروژه، دادگان باید به گونه‌ای باشد که برای هر نمونه هر دو مدالیت‌های متن و صوت در دسترس باشد و علاوه بر آن برچسب احساس نیز ارائه شود؛ به عنوان نمونه، دادگان‌هایی مانند MELD و IEMOCAP این ویژگی را دارند و شامل گفتار به همراه متن متناظر و برچسب گذاری احساسی هستند. ارزیابی بر اساس ۷ برچسب احساسی انجام می‌شود (Anger, Disgust, Fear, Joy, Neutral, Sadness, Surprise) و بنابراین دادگان انتخابی شما نیز باید همین برچسب‌ها را داشته باشد یا امکان نگاشت^{۱۰} دقیق آن‌ها به این مجموعه فراهم باشد. همچنین باید توجه داشت که در مرحله‌ی تحویل، مدل روی نمونه‌های آزمون تصادفی بررسی می‌شود. بنابراین انتخاب دادگان و آماده‌سازی آن باید به گونه‌ای انجام شود که مدل نهایی تعمیم‌پذیری و پایداری مناسبی در برابر تغییرات واقعی مانند تفاوت گوینده و شرایط ضبط داشته باشد، و در گزارش نیز اختلاف تعریف احساسات و عدم توازن کلاس‌ها^{۱۱} بین دادگان‌ها شفاف توضیح داده شود.

ملاحظات اجرایی و تحویل پروژه

در تحویل نهایی، انتظار می‌رود تمامی اجزای پروژه به صورت کامل و قابل اجرا ارائه شوند، شامل کدها، فایل‌های تنظیمات و مستندات لازم برای بازتولید نتایج. گزارش پروژه باید تصمیمات طراحی، روش‌شناسی، معماری مدل، فرایند آموزش و تنظیم ابرپارامترها^{۱۲} و نتایج ارزیابی را به صورت شفاف پوشش دهد و علاوه بر گزارش معیارهای اصلی، تحلیل خطا و تفسیر نتایج را نیز ارائه کند.

موارد امتیازی

۱. ارائه‌ی یک واسط سبک برای تست مدل مانند UI ساده یا API جهت بارگذاری ورودی و مشاهده خروجی

۲. اجرای اعتبارسنجی متقابل^{۱۳} برای گزارش پایدارتر نتایج.

۳. انجام Ablation Study کامل: مقایسه‌ی فقط متن، فقط صوت، و ترکیبی و تحلیل اثر هر کدام

۴. دستیابی به عملکرد بالاتر از یک آستانه‌ی مشخص و دریافت امتیاز وزندار مبتنی بر رتبه‌بندی.

سالم و موفق باشید.

¹⁰ Label Mapping

¹¹ Class Imbalance

¹² Hyperparameters

¹³ Cross-Validation