

بسم الله الرحمن الرحيم



استفاده از یادگیری عمیق در تشخیص تکنیک‌های مقاعدسازی به کاررفته در میم‌ها

دانشجو: مهدیه نادری

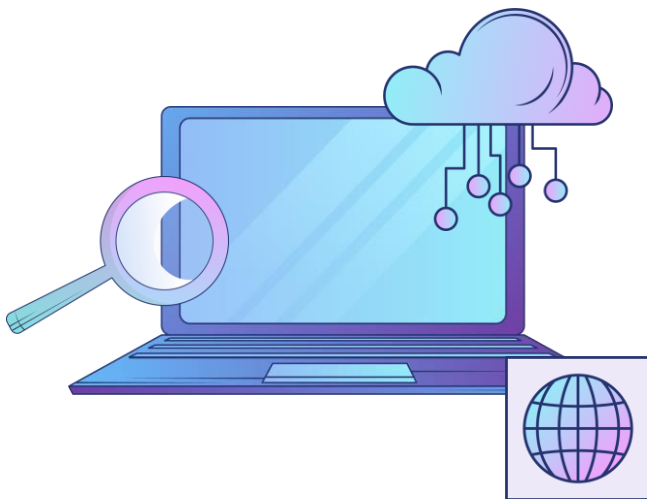
استاد راهنما: دکتر سید صالح اعتمادی

دانشکده مهندسی کامپیوتر

دانشگاه علم و صنعت

مهر ۱۴۰۳

فهرست مطالب



- مقدمه
- کارهای پیشین و مرتبط
- روش‌های پیشنهادی
- نتایج و تفسیر آن‌ها
- نتیجه‌گیری و کارهای آینده

اهمیت موضوع

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

• نیاز به تحلیل تکنیک‌های متقاعدسازی و پروپاگاندا در میم‌ها

○ گسترش شبکه‌های اجتماعی چند رسانه‌ای

○ افزایش تولید محتوای کاربرمحور

○ تاثیر گذاری روش‌های متقاعد سازی بر افکار عمومی

■ انتشار اطلاعات نادرست

■ دستکاری اجتماعی

• ضرورت توسعه‌ی روش‌های خودکار

○ دستیابی راحت‌تر به حجم بالای اطلاعات

○ جلوگیری از خطای انسانی، عدم دقت و ناپایداری

○ مقیاس پذیر برای حجم بزرگ دیتا

○ شناسایی سریع‌تر تغییرات در شبکه‌های اجتماعی





تعریف مسئله

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

۴

• تسک چهارم مسابقه Semeaval 2024 :

- تشخیص چندزبانه تکنیک‌های متقاعدسازی در میم‌ها
- مجموعه داده‌ی آموزشی به زبان انگلیسی، مجموعه داده‌ی تست به زبان انگلیسی، مقدونی شمالی و بلغاری

• Subtask 1:

- داده‌ی متنی میم‌ها
- تسک طبقه‌بندی چندلیبلی در حوزه‌ی پردازش زبان طبیعی
- ۲۰ لیل دارای رابطه‌ی سلسله‌مراتبی
- معیار ارزیابی hierarchical-F1

• Subtask 2a: مورد پژوهش قرار نگرفته‌است.

• Subtask 2b:

- داده‌ی چند رسانه‌ای میم‌ها (متن و تصویر)
- تسک طبقه‌بندی باینری چندرسانه‌ای (میم دارای تکنیک متقاعدسازی هست یا خیر؟)
- معیار ارزیابی F1-macro

استفاده از یادگیری عمیق در تشخیص تکنیک‌های متقاعدسازی به کاررفته در میم‌ها | مهدیه نادری

مهر ۱۴۰۳

طبقه بندی چند لیبلی سلسله مراتبی

مقدمه

کارهای
پیشین و
مرتبط

روش های
پیشنهادی

نتایج و
تفسیر آن ها

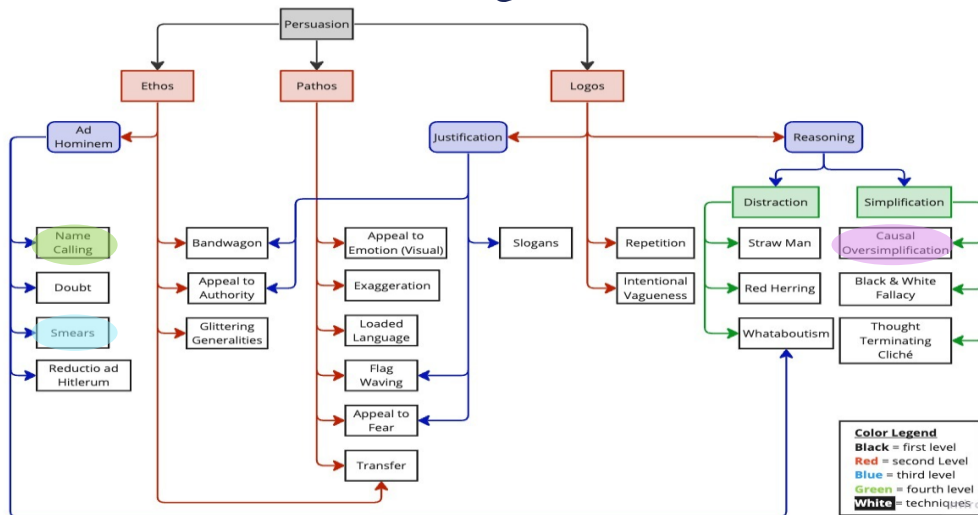
نتیجه گیری
و کارهای
آینده

مثال: فرض میکنیم لیبل درست برای ورودی داده شده Name Calling باشد.

سیستم A لیبل Smears و سیستم B لیبل Oversimplification را پیشبینی می کند.

با اینکه هر دو پیشبینی اشتباه است سیستم A پاداش جزئی بیشتری دریافت می کند زیرا لیبلی که پیشبینی کرده

تعداد گره والد و یا جد مشترک بیشتری با لیبل صحیح دارد.





راهنمای مدرن در پردازش چندوجهی

مقدمه

۱. مدل‌های ترکیبی متن و تصویر:

ویژگی‌های تصویری و متنی را به‌طور همزمان ترکیب می‌کنند. مانند VisualBERT و LXMERT

۲. مدل‌های چندوجهی:

از تعاملات پیچیده بین داده‌های متنی و تصویری برای یادگیری استفاده می‌کنند. مانند ViLT و CLIP

۳. مدل‌های تفکیکی متن و تصویر:

ابتدا ویژگی‌های متنی و تصویری را به‌طور جداگانه استخراج کرده و سپس آن‌ها را ترکیب می‌کنند.

BERT + CNN: استفاده از BERT برای پردازش متن و یک شبکه عصبی کانولوشن برای استخراج ویژگی‌های تصویری.

۴. مدل‌های یادگیری عمیق مبتنی بر توجه:

استفاده از مکانیسم توجه برای تمرکز بر ویژگی‌های مهم متن و تصویر

Multimodal Transformer: مدل مبتنی بر توجه که ویژگی‌های متنی و تصویری را با یکدیگر ترکیب می‌کند.

Unified Vision-Language Pre-training (UVLP): از مکانیسم توجه برای یادگیری مشترک متن و تصویر

استفاده می‌کند و توانایی درک عمیق‌تری از ارتباطات بین داده‌ها دارد.

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



کارهای پیشین

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

• تسک سوم Semeval 2023 :

- شناسایی ژانر، چارچوب‌بندی و تکنیک‌های متقاعدسازی در اخبار آنلاین در یک محیط چندزبانه (NLP)
- داده‌ها و زبان‌ها: مقالات خبری به زبان‌های انگلیسی، فرانسوی، آلمانی، ایتالیایی، لهستانی و روسی

• Subtask 1: طبقه‌بندی برای شناسایی نوع مقاله (نظر، گزارش، یا طنز)

- لیبل‌ها: سه نوع ژانر "گزارش"، "نظر"، "طنز"
- معیار ارزیابی: F1-macro
- مدل برتر: XLM-RoBERTa

• Subtask 2: شناسایی چارچوب‌ها به صورت چندکلاسه

- لیبل‌ها: ۱۴ چارچوب مانند "اقتصادی"، "اخلاقی"، و "سیاسی"
- معیار ارزیابی: F1-micro
- مدل‌های برتر: mBERT و ALBERT

• Subtask 3: شناسایی تکنیک‌های متقاعدسازی هر پاراگراف به صورت چندلیلی

- لیبل‌ها: ۲۳ تکنیک متقاعدسازی مختلف
- مدل برتر: XLM-RoBERTa large
- استراتژی‌ها: بررسی پیش‌پردازش و تنظیم آستانه اطمینان
- معیار ارزیابی: F1-micro



مدل‌های زبانی

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

مدل GPT-2

- معماری ترنسفورمر و مدل زبان خودنظارتی
- چندزبانه بودن: اگرچه عمدتاً برای زبان انگلیسی طراحی شده است، اما قابلیت تولید متن در زبان‌های دیگر را نیز داراست.
- مقیاس‌پذیری: با افزایش ابعاد مدل و داده‌های آموزشی، کیفیت تولید متن به طرز قابل توجهی افزایش می‌یابد.
- پیشرفته بودن: بهبود با استفاده از تکنیک‌های پیشرفته‌ای مانند fine tune و یادگیری انتقالی

مدل XLM-Roberta

- معماری ترنسفورمر
- پشتیبانی از چند زبان: به طور همزمان از بیش از ۱۰۰ زبان مختلف پشتیبانی می‌کند
- آموزش بر روی داده‌های متنوع: آموزش دیده روی متون چند زبانه
- یادگیری بین زبانی: این مدل به طور خاص برای تسهیل یادگیری بین زبانی طراحی شده است و از دانش یک زبان برای بهبود عملکرد در زبان‌های دیگر استفاده می‌کند.



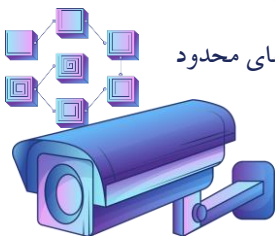
مدل‌های بینایی کامپیوتر عمیق

مدل VGG-16:

- عمق شبکه: یادگیری ویژگی‌های پیچیده تر با ۱۶ لایه‌ی قابل آموزش (لایه‌های کانولوشن و Fully Connected)
- مورد استفاده در در تسک‌های انتقال یادگیری و Fine-tuning
- عملکرد بالا در شناسایی اشیاء: عملکرد بسیار خوب در مسابقات معتبر شناسایی اشیاء مانند ImageNet

مدل Vision Transformer (ViT):

- معماری مبتنی بر ترنسفورمر و عدم نیاز به CNN
- تجزیه تصویر به تکه‌ها: تصاویر به تکه‌های کوچک (پچها) تقسیم می‌شوند و هر تکه به‌عنوان یک ورودی به مدل داده می‌شود. این روش به مدل کمک می‌کند تا ویژگی‌های محلی و کلی را همزمان یاد بگیرد.
- نیازمند داده‌های آموزشی بسیار زیاد برای بهترین عملکرد و نداشتن عملکرد بهینه در داده‌های محدود
- کاهش پارامترها و پیچیدگی
- توسعه‌پذیری: قابلیت ترکیب با سایر معماری‌ها



مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



مجموعه داده‌ها Subtask1

از اجتماع مجموعه داده‌های آموزشی، اعتبارسنجی و توسعه برای آموزش مدل‌های پیشنهادی استفاده شد.

Train set	7000
Validation set	500
Dev set	1500
English test set	1500
Bulgarian test set	436
North Macedonian test set	259

مقدمه

کارهای
پیشین و
مرتبط

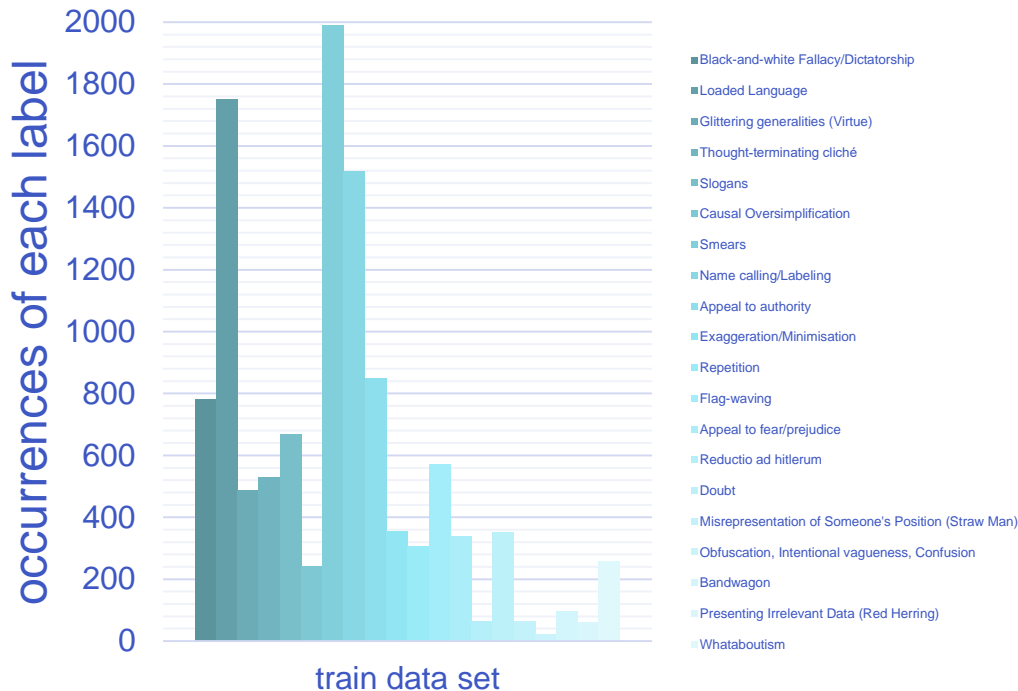
روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



عدم توازن داده‌ها در Subtask1



مقدمه

کارهای
پیشین و
مرتبط

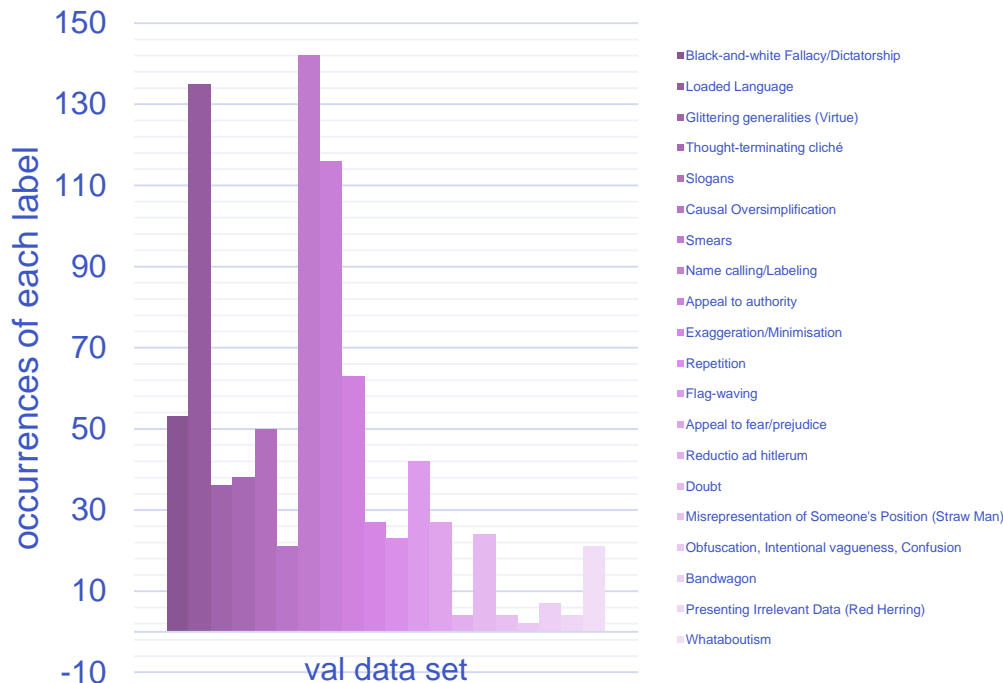
روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



عدم توازن داده‌ها در Subtask1



مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

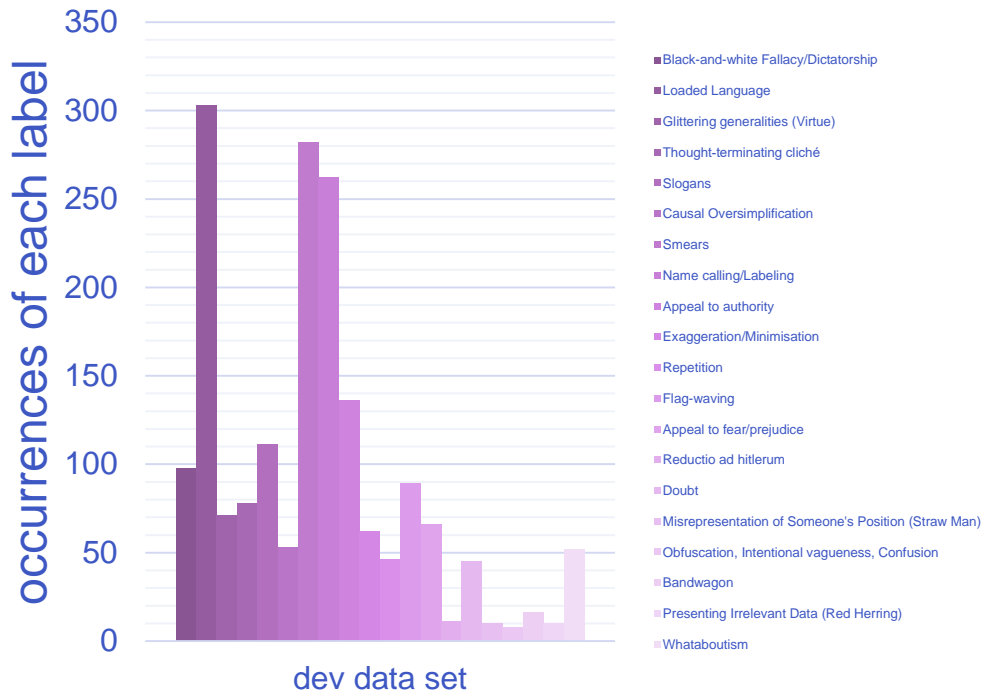
۱۲

مهر ۱۴۰۳

استفاده از یادگیری عمیق در تشخیص تکنیک‌های متقاعدسازی به کاررفته در میم‌ها | مهدیه نادری



عدم توازن داده‌ها در Subtask1



مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

۱۳

مهر ۱۴۰۳

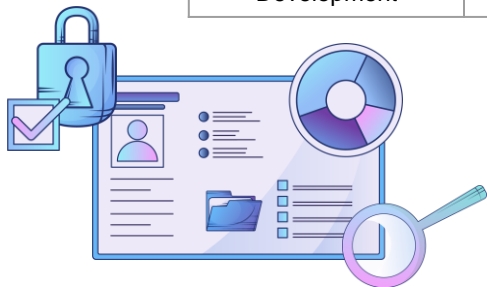
استفاده از یادگیری عمیق در تشخیص تکنیک‌های متقاعدسازی به کاررفته در میم‌ها | مهدیه نادری



مجموعه داده‌ها Subtask 2b

- ID: شناسه منحصر به فرد هر میم
- متن: محتوای متنی موجود در میم
- تصویر: نام فایل تصویر مربوط به میم
- برچسب: Propagandistic ، Non-Propagandistic

Data Set/Label	Propagandistic	Non-Propagandistic
Train	۸۰۰	۴۰۰
Validation	۱۰۰	۵۰
Development	۲۰۰	۱۰۰



مجموعه داده تست زبان انگلیسی : ۶۰۰ داده
مجموعه داده تست زبان بلغاری: ۱۰۰ داده
مجموعه داده تست زبان مقدونیه شمالی: ۱۰۰ داده

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

پیش پردازش داده‌ها

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

- پردازش تصاویر: تغییر اندازه به ۲۲۴×۲۲۴ در حالت RGB و نرمال‌سازی.
- تجزیه و تحلیل زبانی: استفاده از ابزار NLTK برای توکن‌سازی، حذف کلمات توقف و ریشه‌یابی.
- پاکسازی متن: حذف نویزها با استفاده از API OpenAI
- اصلاح دستی داده‌ها: مداخله انسانی در موارد خطای استخراج متن.
- نمونه:

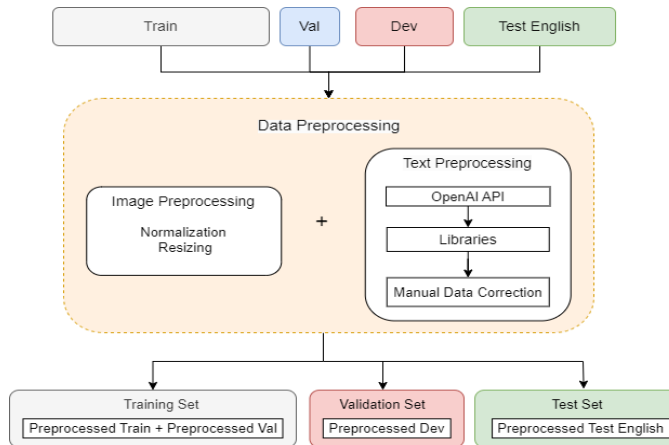
```
{ "id": "25064",  
  "text": "@:\\nDer",  
  "image": "prop_meme_4499.png",  
  "label": "propagandistic" }
```





Subtask 2b در API OpenAI

- از API سرویس OpenAI برای پیش‌پردازش متون میم‌ها استفاده شده است. ابتدا دستورالعمل خاصی برای استخراج متن اصلی و حذف اطلاعات اضافی مانند نام کاربری و کاراکترهای اضافی تعریف می‌شود. سپس از طریق یک درخواست HTTP به API OpenAI، درخواست ارسال شده و متن پیش‌پردازش شده از پاسخ دریافت می‌گردد. این متن پردازش شده سپس در یک فایل ذخیره می‌شود تا به‌عنوان داده تمیز شده برای مراحل بعدی مدل‌سازی و تحلیل استفاده شود.



مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



روش استفاده شده در Subtask1

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

- ترکیب یادگیری نظارت‌شده و تنظیم دقیق در مدل‌های GPT-2 و XLM-Roberta
- استفاده از داده‌های ترکیبی آموزش و اعتبارسنجی و توسعه دهنده
- سورت کردن داده‌ها بر اساس طول متن: داده‌ها با طول مشابه در یک بچ و تسریع فرایند آموزش
- مقابله با عدم توازن کلاس‌ها با BCE و class-weight
- بهینه‌سازی نزول گرادیان با AdamW
- تنظیم ابرپارامترها با جستجوی شبکه‌ای و اعتبارسنجی متقابل

پارامتر	مقدار
تعداد دوره آموزشی	۵
اندازه دسته آموزشی	۱۶
اندازه دسته اعتبارسنجی	۱۶
نرخ یادگیری	۰,۰۰۱



روش‌های استفاده شده در Subtask 2b

- استخراج ویژگی‌های متنی با GPT-2 و XLM-RoBERTa
- استخراج ویژگی‌های تصویری با ViT و VGG
- ایجاد نمایه چندوجهی از ترکیب ویژگی‌های متنی و تصویری
- لایه فیوژن چندوجهی برای ادغام ویژگی‌ها
- لایه طبقه‌بندی دودویی برای پیش‌بینی وجود تکنیک‌های متقاعدسازی

پارامتر	مقدار
تعداد دوره آموزشی	۱۰
اندازه دسته آموزشی	۳۲
اندازه دسته اعتبارسنجی	۳۲
وزن کاهشی	۰,۰۰۱
نرخ یادگیری	۰,۰۰۱
بهترین آستانه	۰,۳۹

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



نتایج Subtask1

نتایج آزمایش مدل‌ها روی مجموعه تست به زبان انگلیسی

Model	Hierarchical F1	Hierarchical Precision	Hierarchical Recall
XLM-RoBERTa	0/42667	0/29247	0/77798
XLM-RoBERTa with best treshold	0/50573	0/47555	0/54001
GPT-2	0/40266	0/26986	0/79281
GPT-2 with best threshold	0/59727	0/52563	0/69152
Baseline	0/36865	0/47711	0.30036

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

۱۹



نتایج Subtask1

نتایج بهترین مدل XLM-RoBERTa و GPT-2 بر اساس Hierarchical F1 در مجموعه تست زبان‌های بلغاری و مقدونیه شمالی

Language	Model	Hierarchical F1	Hierarchical Precision	Hierarchical Recall
Bulgarian	XLM-RoBERTa	0/38433	0/38672	0/38197
	GPT-2	0/33070	0/20953	0/78418
	Baseline	0/28377	0/31881	0/25567
North Macedonian	XLM-RoBERTa	0/30863	0/18891	0/84256
	GPT-2	0/31222	0/28901	0/33948
	Baseline	0/30692	0/31403	0/30012

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



نتایج Subtask 2b

نتایج آزمایش مدل‌ها روی مجموعه اعتبارسنجی

Model	F1-macro	F1-macro Best Threshold
VGG+XLM-RoBERTa	0/58	0/63
VGG+GPT-2	0/71	0/76
ViT + XLM-RoBERTa	0/40	0/53
ViT + GPT-2	0/35	0/51

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



نتایج Subtask 2b

نتایج مدل VGG+GPT-2 (بهترین مدل) روی مجموعه تست زبان‌های انگلیسی، بلغاری و مقدونیه شمالی

Language	F1-macro	Baseline F1-macro	F1-micro	Baseline F1-micro
English	0/67398	0/25000	0/74000	0/33333
Bulgarian	0/51637	0/16667	0/74000	0/20000
North Macedonian	0/57653	0/09091	0/79000	0/10000

مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده



جمع بندی

مقدمه

کارهای
پیشین و
مرتبط

روش های
پیشنهادی

نتایج و
تفسیر آن ها

نتیجه گیری
و کارهای
آینده

○ Subtask1:

- استفاده از تکنیک انتخاب بهترین آستانه برای بهبود عملکرد مدل GPT-2 در زبان انگلیسی
- عملکرد بهتر مدل GPT-2 بر اساس معیار Hierarchical F-1 (مقدار ۰.۳۱۲۲۲)
- عدم کسب نتایج مطلوب برای هر دو مدل در زبان مقدونیه ای
- عملکرد برتر مدل XLM-RoBERTa در زبان بلغاری با دقت ۰.۳۸۴۳۳

○ Subtask2b:

- مدل VGG + GPT-2، بهترین عملکرد در داده های انگلیسی
- در مدل های ViT + XLM-RoBERTa و ViT + GPT-2، F1-macro به ترتیب ۰.۵۳ و ۰.۵۱
- عملکرد پایین تر در زبان های بلغاری و مقدونیه ای
- مدل GPT-2، بهبود عملکرد در چندین زبان
- چالش های داده های چندزبانه

کارهای آینده



مقدمه

کارهای
پیشین و
مرتبط

روش‌های
پیشنهادی

نتایج و
تفسیر آن‌ها

نتیجه‌گیری
و کارهای
آینده

- استفاده از داده‌های آموزشی بزرگ‌تر و متنوع‌تر: افزودن مجموعه‌های داده جدید برای بهبود یادگیری الگوها
- تنظیم دقیق‌تری از هایپرپارامترها: جستجوی پیشرفته برای بهینه‌سازی نرخ یادگیری و دیگر هایپرپارامترها
- روش‌های یادگیری نیمه‌نظارت‌شده: بهره‌گیری از داده‌های بدون برچسب برای افزایش دقت مدل
- تجزیه و تحلیل عمیق‌تر نتایج: استفاده از متریک‌های متنوع برای بررسی عملکرد و شناسایی نقاط ضعف
- پیش‌پردازش داده‌ها: مراحل پیچیده‌تری برای پاک‌سازی و نرمال‌سازی داده‌ها
- ترکیب مدل‌های مختلف: استفاده از چندین مدل از جمله ترنسفورمر و CNN برای افزایش دقت پیش‌بینی
- استفاده از تکنیک‌های نوآورانه: مانند یادگیری تقویتی یا یادگیری فعال برای بهبود عملکرد
- تحلیل خطاها: شناسایی و بررسی نمونه‌های نادرست پیش‌بینی‌شده برای بهبود مدل





با سپاس از توجه شما

M_naderi98@comp.iust.ir



مراجع

- Bakhshande, Fatemehzahra, and Naderi, Mahdieh. "CVcoders on SemEval-2024 Task 4." In Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), June 2024, Mexico. Association for Computational Linguistics, 1912–1918.
- Brown, Tom B., et al. "Language Models Are Few-Shot Learners (GPT-3)." ArXiv, 2020.
- Carion, N., et al. "End-to-End Object Detection with Transformers." *European Conference on Computer Vision (ECCV)*, 2020.
- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). "Unsupervised Cross-lingual Representation Learning at Scale."
- Conneau, A., et al. "Unsupervised Cross-lingual Representation Learning at Scale (XLM-R)." ArXivDeng, J., et al. (2009). "ImageNet: A large-scale hierarchical image database." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 248–255.
- Dimitrov, Dimitar, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. "SemEval-2021 Task 6: Detection of Persuasion Techniques in Texts and Images." *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2021)*, 2021, pp. 70–98.
- Dosovitskiy, A., et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." arXiv preprint arXiv:2010.11929.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. [Chapter 9: Convolutional Networks].
- Hsinhung, W. (n.d.). *GPT-2 detailed model architecture*. Medium.

مراجع

- Jowett, Garth S., and Victoria O'Donnell. *Propaganda & Persuasion*. 6th ed., SAGE Publications, 2019.
- Karpathy, Andrej. "Let's Reproduce GPT-2 (124M)." YouTube.
- Karpathy, Andrej. "nanoGPT" repo.
- Lample, G., & Conneau, A. (2019). "Cross-lingual Language Model Pretraining." ArXiv.
- M. Bach, T. Minervini, S. Pradhan, and E. Hovy, "Evolving Multimodal Models: Detection of Persuasion Techniques with Limited Labeled Data," ArXiv, 2023.
- OpenAI. (2023). *GPT-2*. GitHub. Archived from the original on March 11, 2023. Retrieved March 13, 2023, Propaganda, "SemEval 2023 Task 3: Persuasion Techniques in Texts and Images," 2023.
- Propaganda, "SemEval 2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes," 2024.
- Simonyan, K., & Zisserman, A. (2014). "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv preprint arXiv:1409.1556.
- Touvron, H., et al. "Training data-efficient image transformers & distillation through attention." arXiv preprint arXiv:2012.12877.
- Vaswani, Ashish, et al. "Attention Is All You Need." ArXiv, 2017.
- Better Language Models and Their Implications, OpenAI Blog, 2019.
- S. Kiritchenko, S. Matwin, R. Nock, and A. F. Famili, "Learning and Evaluation in the Presence of Class Hierarchies: Application to Text Categorization," in *Proceedings of the National Research Council Canada*, University of Ottawa, Canada, 2004.
- A. Kr. Ojha, S. Matwin, R. Waseem, and G. D. S. Martino, "SemEval-2023 Task 3: Detecting Persuasion Techniques in Texts and Images," *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, Association for Computational Linguistics, June 2023, pp. 591–605.