

بررسی مقالات و ارائه‌ای از مسئله ABSA with Instruction Learning

در راستای تهیه ارائه‌ای از مبحث ABSA with Instruction Learning، مقالات مرتبط مندرج در جدول ذیل در این حوزه بررسی شده و در ادامه خلاصه‌ای از مفاد آنها تشریح گردیده است:

ردیف	عنوان	نوع-سال انتشار	نویسندگان
۱	A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges ¹	مقاله مروری-۲۰۲۲	Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, Wai Lam
۲	Large Language Models for Aspect-Based Sentiment Analysis ²	مقاله پژوهشی-۲۰۲۳	Paul F. Simmering, Paavo Huoviala
۳	InstructABSA: Instruction Learning for Aspect Based Sentiment Analysis ³	مقاله پژوهشی-۲۰۲۴	Kevin Scaria, Himanshu Gupta, Siddharth Goyal, Saurabh Sawant, Swaroop Mishra, Chitta Baral

۱ تحلیل و خلاصه مقاله مروری شماره یک: A Survey on ABSA

۱-۱ مقدمه و هدف

تحلیل احساسات چیست؟

تحلیل احساسات (Sentiment Analysis) یکی از مهم‌ترین حوزه‌های پردازش زبان طبیعی است که هدفش شناسایی نظرات و احساسات افراد از متن است.

مثال ساده: "این فیلم عالی بود" ← احساس: مثبت

مشکل تحلیل احساسات سنتی: تحلیل احساسات سنتی فقط یک برچسب کلی به کل متن می‌دهد. اما در دنیای واقعی، یک متن می‌تواند چند نظر یا جنبه متفاوت داشته باشد. به طور مثال:

متن: "غذای این رستوران عالی بود ولی سرویس **افتضاح** بود"

تحلیل سنتی: برآیند احساس مثبت است یا منفی؟ (نمی‌تواند جواب دهد)

حل مشکل روش‌های سنتی: Aspect-Based Sentiment Analysis (ABSA)

1 <https://arxiv.org/abs/2203.01054>

2 <https://arxiv.org/abs/2310.18025>

3 <https://aclanthology.org/2024.naacl-short.63.pdf>

ABSA (تحلیل احساسات مبتنی بر جنبه) مشکل رویکردهای سنتی را حل کرده و از یک متن، جنبه‌ها (aspects) را استخراج کرده و نظر / احساس نسبت به هر جنبه را تعیین می‌کند. مثلاً برای جمله "غذای این رستوران عالی بود ولی سرویس افتضاح بود"، جنبه‌ها و احساس‌های متفاوتی استخراج می‌شود:

جنبه (Aspect)	نظر (Opinion)	احساس (Sentiment)
غذا	عالی	مثبت ✓
سرویس	افتضاح	منفی ✗

در نتیجه، هدف این مقاله مروری ارائه یک نقشه جامع از حوزه ABSA است؛ به‌طوری که:

- تمام تعاریف، تسک‌ها و صورت‌بندی‌های مسئله یکپارچه شوند
- روش‌های موجود به‌صورت سیستماتیک دسته‌بندی شوند
- چالش‌های باز و مسیرهای آینده پژوهش مشخص گردد

این مقاله بیشتر نقش چارچوب نظری و مرجع پایه را در ادبیات ABSA دارد.

۱-۲ عناصر اصلی ABSA و طبقه‌بندی تسک‌های آن

این مقاله چهار عنصر اصلی جملات را در ABSA را به شرح جدول زیر معرفی می‌کند که ورودی‌ها و خروجی‌های هر تسک ABSA ترکیبی از این عناصر خواهد بود.

عناصر جمله	توضیح	مثال
Aspect Term	کلمه/عبارت خاص در متن	"غذا"، "سرویس"
Aspect Category	دسته‌بندی کلی جنبه	FOOD#QUALITY, SERVICE#GENERAL
Opinion Term	کلمه‌ای که نظر رو بیان می‌کند	"عالی"، "افتضاح"، "سریع"
Sentiment Polarity	قطبیت احساس	مثبت، منفی، خنثی

مقاله در ادامه یک مرور سیستماتیک و جامع روی تمام تسک‌های ABSA و روش‌های حل و چالش‌های موجود ارائه کرده و تسک‌های ABSA را مشتمل بر موارد زیر طبقه‌بندی کرده است. این طبقه‌بندی باعث شده پژوهش‌ها قابل مقایسه شوند و مشخص شود هر روش دقیقاً کدام مسئله را بررسی و رفع می‌کند.

❖ تسک‌های تکی (Single Element)

۱. ATE: فقط استخراج جنبه

۲. OTE: فقط استخراج نظر

۳. ATSC: فقط تعیین احساس

❖ تسک‌های جفتی (Pair)

۴. AOPE: جفت (جنبه، نظر)

۵. ASPE: جفت (جنبه، احساس)

❖ تسک‌های سه تایی (Triplet)

۶. ASTE: سه تایی (جنبه، نظر، احساس)

❖ تسک‌های چهار تایی (Quad)

۷. ASQP: چهار تایی (جنبه، دسته، نظر، احساس)

۳-۱ روش‌های حل (Solution Paradigms) و تأثیر Pre-training

Survey روش‌های مختلف حل ABSA را به چند دسته تقسیم کرده است:

ردیف	نام روش	توضیح	مزایا
۱	Sequence Labeling	ایده: هر کلمه یک برچسب میگیرد. مثل NER	ساده و قابل فهم ابزارهای آماده زیاد
۲	Span-based Methods	ایده: به جای برچسب‌گذاری هر کلمه، محدوده شروع و پایان هر عنصر را پیدا کن.	دقت بالاتر برای عبارات چند کلمه‌ای مدل‌سازی بهتر boundaries
۳	MRC (Machine Reading Comprehension)	ایده: تبدیل ABSA به سوال-جواب	استفاده از دانش مدل‌های QA انعطاف‌پذیری بالا
۴	Seq2Seq (Sequence-to-Sequence)	ایده: ورودی متن، خروجی هم متن. مدل یاد میگیرد که خروجی را تولید کند مثل: T5-based methods	یکپارچگی: همه تسک‌ها با یک فرمت استفاده از semantics برچسب‌ها InstructABSA نیز از این پارادایم استفاده میکند
۵	Graph-based Methods	ایده: استفاده از گراف وابستگی نحوی جمله	مدل‌سازی روابط syntactic تشخیص بهتر opinion مربوط به هر aspect

۴-۱ سناریوهای عملی

Survey بررسی کرده است که ABSA در دنیای واقعی چه چالش‌هایی دارد:

عنوان	مشکل	چالش	راه حل
Cross-Domain Transfer	داده فقط برای یک دامنه (مثلاً رستوران) داریم، ولی می‌خواهیم روی دامنه دیگر (مثلاً هتل) کار کند.	<ul style="list-style-type: none"> کلمات متفاوت: "خوشمزه" در رستوران \neq "تمیز" در هتل جنبه‌های متفاوت: "غذا" و "اتاق" 	<ul style="list-style-type: none"> Domain Adaptation Multi-task Learning Data Augmentation
Cross-Lingual Transfer	داده برچسب‌دار فقط برای انگلیسی هست، ولی فارسی/عربی/چینی نیز می‌خواهیم.	-	<ul style="list-style-type: none"> استفاده از Multilingual BERT (mBERT) Translation-based methods Zero-shot cross-lingual transfer

۵-۱ چالش‌های باز (Open Challenges)

Survey این چالش‌های حل نشده را شناسایی کرده است:

ردیف	چالش	توضیح
۱	Implicit Aspects	جنبه‌ای که صریحاً در متن نیست ولی باید استخراج شود: "گروه" \leftarrow قیمت
۲	Implicit Opinions	نظری که صریح نیست: "۳ ساعت منتظر موندم" \leftarrow نظر منفی
۳	Multi-aspect Sentences	جملاتی با چند جنبه و احساس متفاوت "غذا عالی بود ولی قیمتش زیاده و سرویس هم متوسط بود"
۴	Aspect-Opinion Alignment	تشخیص اینکه کدام نظر به کدام جنبه مربوطه
۵	Resource Scarcity	کمبود داده برچسب‌دار در زبان‌ها/دامنه‌های مختلف

۶-۱ نتیجه‌گیری و روندهای پژوهشی (Research Trends)

- Instruction Learning و LLM ها در زمان نگارش survey هنوز به‌طور جدی بررسی نشده‌اند

- با توجه به Survey بودن مقاله، بیشتر نقش «تشخیص مسئله» دارد تا «حل مسئله»
- با وجود آن که این مقاله مروری راه حل عملی جدیدی برای چالش های مطرح شده در حوزه ی ABSA ارائه نمی دهد و نقش اصلی آن بیشتر تشخیص و صورت بندی دقیق مسئله است تا ارائه روش های حل، اما با تحلیل نظام مند ادبیات موجود، روندهای اصلی پژوهشی این حوزه را به روشنی شناسایی و جهت گیری تحقیقات آینده را مشخص می کند.
- بر اساس نتایج این Survey، مسیر تکامل پژوهش های ABSA به سمت موارد زیر حرکت کرده است:
 - ✓ نخست، پژوهش ها از تسک های تکی و مجزا به سوی تسک های ترکیبی و استخراج هم زمان چند مؤلفه (sentiment و opinion, aspect) پیش رفته اند تا تحلیل دقیق تری از متن حاصل شود.
 - ✓ دوم، رویکردهای مولد (Generative Approaches) به ویژه مدل های Seq2Seq مانند T5، به عنوان جایگزینی برای روش های برچسب گذاری سنتی مطرح شده اند و امکان یکپارچه سازی تسک ها را فراهم کرده اند.
 - ✓ سوم، نیاز به چارچوب های یکپارچه (Unified Frameworks) که بتوانند با یک مدل واحد چندین زیر تسک ABSA را پوشش دهند، به عنوان یک جهت گیری کلیدی شناسایی شده است؛ مسیری که بعدها در مدلهایی مانند InstructABSA تحقق یافته است.
 - ✓ چهارم، مقاله به ظهور Instruction Learning به عنوان یک رویکرد نوظهور اشاره می کند که پس از انتشار این Survey، به سرعت مورد توجه پژوهشگران قرار گرفت و امکان یادگیری تسک ها از طریق توضیح زبانی را فراهم ساخته است.
 - ✓ در نهایت، Survey به استفاده روزافزون از مدل های زبانی بزرگ (LLM-based Methods) مانند GPT اشاره می کند که افق های جدیدی را برای حل مسئله ی ABSA، به ویژه در سناریوهای کم داده و چنددانه، گشوده اند.
 - ✓ در مجموع، این مقاله با ترسیم یک نقشه راه روشن، زمینه ی نظری لازم برای ظهور روش هایی مانند InstructABSA و مطالعات مبتنی بر LLM ها را فراهم کرده است.

۲ تحلیل و خلاصه مقاله شماره دو: LLM for ABSA

۲-۱ مقدمه و هدف

تحلیل احساسات مبتنی بر جنبه (ABSA)، برای ارائه بینش در مورد متون دیجیتالی، مانند بررسی محصولات یا بحث‌های انجمن‌ها استفاده شده و یک قابلیت کلیدی برای زمینه‌های علوم اجتماعی دیجیتال، علوم انسانی و تحقیقات بازار می‌باشد.

مدل‌های زبانی بزرگ (Large Language Models - LLMs) مثل GPT-3.5 و GPT-4 قابلیت‌های بی‌سابقه‌ای در پردازش متن دارند. این مدل‌ها عمومی بوده و می‌توانند نقش‌های مختلفی را از جمله نقش مدل‌های تخصصی ایفا کنند. اما آیا LLM های عمومی مثل GPT می‌توانند با مدل‌های تخصصی ABSA رقابت کنند؟ چراکه ABSA برای تحقیقات بازار و تحلیل نظرات مشتریان خیلی مهم بوده و اگر LLM ها بتوانند این کار را انجام بدهند، دیگر نیازی به آموزش مدل‌های تخصصی نخواهد بود.

پس هدف اصلی این مقاله بررسی این پرسش است که آیا مدل‌های زبانی بزرگ (LLMs) مانند GPT-3.5 و GPT-4 می‌توانند بدون طراحی مدل‌های تخصصی، مسئله‌ی Aspect-Based Sentiment Analysis (ABSA) را با دقت قابل رقابت با روش‌های state-of-the-art (SOTA) حل کنند یا خیر.

نویسندگان این مقاله به‌طور خاص می‌خواهند سه رویکرد را با هم مقایسه کنند:

۱. استفاده از LLM ها در حالت zero-shot/few-shot با prompt engineering

۲. fine-tuning یک LLM عمومی (GPT-3.5) روی داده‌ی ABSA

۳. مقایسه این روش‌ها با مدل تخصصی InstructABSA به‌عنوان SOTA قبلی

لازم به ذکر است، تمرکز مقاله روی تسک مشترک استخراج جنبه + تعیین قطبیت احساس در دیتاست استاندارد SemEval-2014 است که شامل نقد و بررسی‌هایی در مورد لپ‌تاپ و رستوران می‌باشد.

۲-۲ روش کار

نویسندگان عملکرد GPT-3.5 (۱۷۵ میلیارد پارامتر) و GPT-4 (۱.۷ تریلیون پارامتر) را در مقایسه با InstructABSA (۲۰۰ میلیون پارامتر)، در تنظیمات مختلف ذیل ارزیابی می‌کنند:

- zero-shot (بدون هیچ مثالی، فقط با prompt)

- few-shot (با ۶ مثال در prompt)

- fine-tuned GPT-3.5 (آموزش روی کل دیتاست (۵۷۵۹ نمونه))

برای افزایش قابلیت استفاده‌ی عملی، خروجی مدل‌ها به صورت JSON ساخت یافته تولید می‌شود. همچنین مقاله فقط به دقت بسنده نمی‌کند و یک تحلیل خطا (Error Analysis) و یک تحلیل اقتصادی (Cost-Performance Trade-off) هم ارائه می‌دهد، که در مقالات ABSA کمتر دیده می‌شود.

۲-۳ نتایج کلیدی

مدل	حالت	F1-Score	مقایسه با InstructABSA
GPT-4	Zero-shot	~۵۵	ضعیف تر ✗
GPT-4	Few-shot (6 examples)	~ ۷۸	تقریباً برابر
GPT-3.5	Fine-tuned	۸۳.۸	۵.۷٪ بهتر ✓
InstructABSA	Fine-tuned	۷۸.۱	- baseline -

جدول مقایسه F1-Score

جنبه	Prompt Engineering	Fine-tuning
نیاز به داده	کم (چند مثال)	زیاد (هزاران نمونه)
عملکرد	خوب	عالی
انعطاف پذیری	بالا (تغییر آسان)	پایین
هزینه اجرا	بالا	پایین
هزینه توسعه	پایین	بالا

جدول مقایسه Fine-tuning با Prompt Engineering

طبق جداول فوق، مهم ترین نتیجه‌ی مقاله این است که:

- fine-tuned GPT-3.5 به $F1 = 83.8\%$ می‌رسد و رکورد SOTA رو میشکنند (این مقدار ۵.۷٪ بهتر از InstructABSA است) اما با هزینه‌ای به مراتب بالاتر و مدلی با حدود ۱۰۰۰ برابر پارامتر بیشتر!
- اما GPT-4 در حالت zero-shot عملکرد ضعیف تری از انتظار دارد. چراکه بدون مثال، LLM ها نمیتوانند قواعد labeling را یاد بگیرند
- GPT-4 با few-shot به نتایج قابل قبول و در سطح مدل‌های تخصصی می‌رسد

- prompt engineering برای zero/few-shot خیلی مهم است، اما نقش آن بعد از fine-tuning تقریباً از بین می‌رود
- fine-tuned models نیازی به prompt پیچیده ندارند

این مقاله به‌خوبی نشان می‌دهد که:

- LLMها توانایی ذاتی انجام ABSA را دارند (مثلاً برای تحقیقات بازار) اما برای رسیدن به دقت بالا، هنوز به fine-tuning نیاز است
- روش‌هایی مانند InstructABSA از نظر نسبت دقت به هزینه، همچنان بسیار رقابتی‌اند و برای مقیاس‌های بزرگ کارآمدتر هستند
- Instruction Learning با مدل‌های کوچک‌تر یک راه‌حل مهندسی شده و اقتصادی‌تر است
- بنابراین، طبق توصیه محققان این مقاله:
 - اگر داده داریم ← Fine-tuning (بهترین عملکرد، هزینه کمتر)
 - اگر داده نداریم (برای زبان‌های کم منبع) ← Few-shot prompting (شروع سریع)

۲-۴ تحلیل اقتصادی

مدل	هزینه برای ۱۰۰۰ نمونه	F1-score
InstructABSA	~\$0.05	78.1
GPT-3.5 Fine-tuned	~\$0.36	83.8
GPT-4 Few-shot	~\$15.02	~78

اگر فقط به دقت نگاه کنیم، Fine-Tuned GPT-3.5 بهترین نتیجه را دارد. اما اگر هزینه به ازای دقت را در نظر بگیریم:

- InstructABSA مقرون به‌صرفه‌تر از GPT-4 است
 - InstructABSA همچنان برای کاربردهای صنعتی با منابع محدود گزینه‌ی بسیار مناسبی است
- این تحلیل نشان می‌دهد که بهترین مدل از نظر دقت، لزوماً بهترین انتخاب عملی نیست.

۳ تحلیل و خلاصه مقاله شماره سه: InstructABSA

۳-۱ مقدمه و هدف

مشابه با مقاله مروری اول، این مقاله نیز تسک‌های ABSA را به شرح جدول زیر بررسی کرده و با پیاده سازی روشهای Instruction based و با یک مدل و رویکرد واحد، سعی نموده همه تسک‌ها را بهبود داده و به بعضی از چالش‌های مطرح شده در survey فوق پاسخ بدهد:

ردیف	تسک	نام کامل	ورودی	خروجی
۱	ATE	Aspect Term Extraction	جمله	لیست جنبه‌ها
۲	ATSC	Aspect Term Sentiment Classification	جمله + جنبه	احساس (مثبت/منفی/خنثی)
۳	ASPE	Aspect Sentiment Pair Extraction	جمله	جفت‌های (جنبه، احساس)
۴	AOOE	Aspect Oriented Opinion Extraction	جمله + جنبه	عبارت نظر
۵	AOPE	Aspect Opinion Pair Extraction	جمله	جفت‌های (جنبه، نظر)
۶	AOSTE	Aspect Opinion Sentiment Triplet Extraction	جمله	سه تایی (جنبه، نظر، احساس)
۷	ACOSQE	Aspect Category Opinion Sentiment Quad Extraction	جمله	چهار تایی کامل

همانگونه که در نمونه‌های واقعی جدول ذیل ذکر شده است، تحلیل جنبه و قطبیت آن اهمیت و کاربرد بسیار بالایی در کسب و کار خواهد داشت:

نمونه‌های واقعی	کاربردها	اهمیت
بررسی نظرات رستوران‌ها در Google Maps	تحلیل نظرات مشتریان در کسب و کارها	درک دقیق نظرات کاربران
نظرسنجی محصولات در آمازون	پایش برند در شبکه‌های اجتماعی	شناسایی نقاط قوت/ضعف محصول
تحلیل توییت‌ها درباره یک فیلم در توییتر	بهبود کیفیت خدمات و محصولات	بهبود تصمیم‌گیری کسب و کار
ارزیابی خدمات بانک‌ها توسط مشتریان	سیستم‌های توصیه گر هوشمند	انعطاف‌پذیری با حوزه‌های جدید

۳-۲ ایده اصلی روش

به جای Fine-tuning معمولی (مدل Tk-Instruct)، از روش Instruction Learning که یادگیری مبتنی بر دستورالعمل است استفاده می‌شود که ساختار هر instruction در دو نسخه شماره یک و دو اینگونه است:

- **InstructABSA-1**: تعریف دستورالعمل + ۲ مثال مثبت
- **InstructABSA-2**: تعریف دستورالعمل + ۲ مثال مثبت + ۲ مثال منفی + ۲ مثال خنثی

پس با تعریف دستورالعمل، مدل را به صورت انعطاف پذیر با حوزه یا نیاز خاص تطبیق می دهیم. نوآوری اینجاست: استفاده هم زمان از مثال های مثبت، منفی و خنثی داخل instruction مثال دستورالعمل: "برای جمله زیر، احساس نسبت به جنبه {جنبه} را شناسایی کن: {متن}"

۳-۳ نتایج کلیدی

تسک	dataset	InstructABSA	SOTA قبلی	بهبود
ATE	Rest14	92.76	87.07	✓ +5.69%
ATE	Lapt14	92.30	87.93	✓ +4.37%
ATSC	Rest15	84.50	74.91	✓ +9.59%
ASPE	Lapt14	79.34	75.97	✓ +3.37%

جدول مقایسه F1-Score و Accuracy

- این نتایج با مدل T5 با M ۲۰۰ پارامتر به دست آمده است، در حالی که مدل های قبلی B ۱.۵ پارامتر داشتند (۷ برابر بزرگتر)
- GPT2-med با ۱.۵ میلیارد پارامتر، ضعیف ترین است. این نشون میده پارامتر بزرگتر همیشه بهتر نیست.
- InstructABSA نشان می دهد که با طراحی هوشمند prompt و استفاده از مثال های متنوع، می توان با مدل های کوچک به نتایج بهتر از مدل های بزرگ رسید.
- اگر مثال هایی با برچسب اشتباه به مدل بدهیم، عملکرد مدل ۱۰٪ افت می کند. این نشان می دهد مدل واقعاً از مثال ها یاد می گیرد و طراحی درست instruction می تواند به اندازه ی معماری پیچیده مهم باشد
- کارایی نمونه (Sample Efficiency): به صورت میانگین در تسک های مختلف ABSA، با ۵۰٪ داده، نتایج رقابتی با روش های دیگر حاصل می شود. این یعنی نیاز به داده برچسب خورده کمتر شده، هزینه annotation کاهش پیدا کرده و برای زبان هایی با داده کم (مثل فارسی!) نیز مفید خواهد بود
- مدل محدودیت هایی هم مشتمل بر موارد زیر دارد:
- ✓ Exposure Bias: در تسک های پیچیده تر (AOPE, AOSTE) عملکرد مدل ضعیف تر است
- ✓ مدل فقط روی دیتاست های SemEval تست شده است
- ✓ فقط روی زبان انگلیسی تست شده است

۴ جمع‌بندی و روند تکاملی مقالات

مقاله مروری سال ۲۰۲۲ نقطه‌ی شروع این مسیر پژوهشی محسوب می‌شود. این مقاله با ارائه یک طبقه‌بندی جامع از تسک‌ها، عناصر و روش‌های ABSA، نشان می‌دهد که این حوزه با پراکندگی شدید در صورت‌بندی مسئله و راه‌حل‌ها مواجه است.

Survey تأکید می‌کند که استفاده از مدل‌های مجزا برای هر زیرتسک، پیچیدگی مدل‌سازی و هزینه‌ی داده را افزایش داده و کمبود داده و نیاز به چارچوب‌های یکپارچه و روش‌های مولد به‌عنوان یک خلأ اساسی مطرح می‌شود. در این مرحله، مقاله بیشتر نقش تشخیص مسئله و تعیین جهت تحقیقات آینده را دارد تا ارائه‌ی راه‌حل عملی.

مقاله سال ۲۰۲۳ در ادامه‌ی طبیعی مسیر ترسیم‌شده توسط Survey، این پرسش را مطرح می‌کند که آیا مدل‌های زبانی بزرگ عمومی می‌توانند بدون طراحی خاص برای ABSA، این مسئله را حل کنند یا خیر. نتایج این مطالعه نشان می‌دهد که:

- LLM‌ها در حالت zero-shot و few-shot توانایی پایه‌ای حل ABSA را دارند اما برای رسیدن به عملکرد state-of-the-art، fine-tuning ضروری است
- مدل fine-tuned GPT-3.5 میتواند رکورد جدیدی ($F1=83.8$) ثبت کند و حتی عملکرد بهتری از روش‌های تخصصی مانند InstructABSA دارد، اما این بهبود ۵.۷٪ با هزینه ۱۰۰۰ برابر پaramتر بیشتر به دست آمده است!
- این مقاله عملاً سقف توان بالقوه‌ی LLM‌ها را نشان می‌دهد و مسئله‌ی مهم trade-off بین دقت و هزینه را برجسته می‌کند.

مقاله InstructABSA که در سال ۲۰۲۴ منتشر شده است، پاسخی عملی و هدفمند به هر دو جریان قبلی است:

- از یک سو، نیاز به چارچوب‌های یکپارچه که در Survey 2022 مطرح شده بود
 - و از سوی دیگر، تمایل به استفاده از صورت‌بندی مولد که در مطالعات LLM-based برجسته شده است
- InstructABSA با بهره‌گیری از Instruction Learning و یک مدل نسبتاً کوچک T5، نشان می‌دهد که می‌توان:

- همه‌ی زیرتسک‌های ABSA را با یک مدل واحد پوشش داد
 - نیاز به داده را کاهش داده (Sample Efficiency) و با مدلی ۷ برابر کوچک‌تر از روش‌های قبلی، به نتایج State-of-the-Art دست یافته است.
 - بدون نیاز به LLM های بسیار بزرگ و با دقتی نزدیک یا حتی بالاتر از روش‌های قبلی
- این مقاله را می‌توان نقطه‌ی تعادل میان نظریه، عملکرد و هزینه در مسیر تکامل ABSA دانست.

به‌صورت زمانی، ادبیات ABSA از تعریف و شناسایی مسئله (2022)، به آزمون توان حداکثری LLM ها (2023) و در نهایت به ارائه‌ی یک راه‌حل یکپارچه، کارآمد و مهندسی‌شده مبتنی بر (2024) Instruction Learning حرکت کرده است. با این حال، چالش Exposure Bias در تسک‌های پیچیده‌تر مانند استخراج سه‌تایی و چهارتایی عناصر همچنان باقی است.

جنبه	Survey (2022)	InstructABSA (2024)	LLMs for ABSA (2023)
نوع	مروری	پژوهشی	پژوهشی
هدف	طبقه‌بندی حوزه	روش جدید	ارزیابی LLMs
مدل	-	Tk-Instruct (200M)	GPT-3.5/4 (175B+)
رویکرد	-	Instruction Tuning	Prompting + Fine-tuning
Sample Efficiency	-	✓ بالا	✗ پایین (برای few-shot)
هزینه	-	✓ خیلی کم	✗ بالا

جدول مقایسه مقالات بررسی شده

همچنین در ادامه چالش‌های مطرح شده حوزه ABSA در مقاله مروری ۲۰۲۲ و نحوه رفع و پاسخ به آنها در دو مقاله دیگر، در جدول ذیل ارائه شده است که طبق مقایسه صورت گرفته، انتخاب بین InstructABSA و LLMs بستگی به trade-off بین هزینه و کارایی دارد. برای کاربردهای صنعتی با حجم بالا، InstructABSA مقرون‌به‌صرفه‌تر است، در حالی که برای prototyping سریع یا زبان‌های کم‌منبع، LLMs با few-shot prompting گزینه بهتری هستند.

جدول مقایسه‌ای راه‌حل‌های InstructABSA و LLMs در رفع چالش‌های مطرح ABSA

چالش	توضیح چالش	راه‌حل InstructABSA	راه‌حل LLMs (GPT)	برنده
کمبود داده	نیاز به داده برچسب‌دار زیاد برای آموزش	با ۵۰٪ داده به نتایج رقابتی میرسد. مثال‌های مثبت/منفی/خنثی در	Few-shot Prompting با ۶ مثال عملکرد قابل قبول. برای بهترین نتیجه،	InstructABSA

	fine-tuning روی کل داده لازم است	instruction به یادگیری سریع تر کمک میکنند		
 LLMs	GPT از دانش عمومی pre-training استفاده کرده و جنبه‌های ضمنی را بهتر تشخیص میدهد	وابسته به وجود مثال‌های instruction در implicit عملکرد متوسط	جنبه‌هایی که صریحاً در متن نیستند (مثلاً "گروه" ← قیمت)	Implicit Aspects
 مساوی	یک API برای همه تسک‌ها. فقط prompt عوض میشود	یک مدل واحد برای هر ۷ زیرتسک. فقط instruction عوض میشود	نیاز به مدل‌های جداگانه برای هر زیرتسک	Multi-task
 InstructABSA	مدل بزرگتر و هزینه بیشتر نیاز به API یا GPU های قوی	مدل کوچک و هزینه کمتر قابل اجرا روی یک GPU معمولی	منابع GPU و زمان inference	هزینه محاسباتی
 LLMs	عملکرد بهتر به دلیل دانش عمومی گسترده. کمتر وابسته به دامنه خاص	عملکرد خوب در cross-domain مثلاً آموزش روی Rest14 ، تست روی Lapt14 با افت معقول	انتقال مدل از یک دامنه (رستوران) به دامنه دیگر (لپ‌تاپ)	Cross-domain