

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

۱ تحلیل و خلاصه مقاله موری 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: برایند احساس مثبت است یا منفی؟ (نمیتواند جواب دهد)

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

۱ <https://arxiv.org/abs/2203.01054>

۲ <https://arxiv.org/abs/2310.18025>

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

(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:

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

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

- تمام تعاریف، تسک‌ها و صورت‌بندی‌های مسئله یکپارچه شوند
- روش‌های موجود به صورت 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.

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

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 چهارتایی (Quad)

۷. ASQP: چهارتایی (جنبه، دسته، نظر، احساس)

۱-۳ روش‌های حل (Solution Paradigms) و تأثیر

Survey روشن‌های 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 boundaries مدلسازی بهتر
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 T5-based methods: مثل	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 aspect opinion مربوط به هر

۱-۴ 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 نیز (مثلاً هتل) کار کند.	کلمات متفاوت: "خوشمزه" در رستوران ≠ "تمیز" در هتل جنبهای متفاوت: "غذا" و "اتاق"	Domain Adaptation • Multi-task Learning • Data Augmentation •
Cross-Lingual Transfer	داده برچسبدار فقط برای انگلیسی هست، ولی فارسی/عربی/چینی نیز میخواهیم.	-	Multilingual BERT (mBERT) Translation-based methods • Zero-shot cross-lingual transfer

۱-۵ چالش‌های باز (Open Challenges)

Survey این چالش‌های حل نشده را شناسایی کرده است:

ردیف	چالش	توضیح
۱	Implicit Aspects	جنبهایی که صریحاً در متن نیست ولی باید استخراج شود: "گرونه" ← قیمت
۲	Implicit Opinions	نظری که صریح نیست: "۳ ساعت منتظر موندم" ← 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/دامنهای مختلف

۱-۶ نتیجه‌گیری و روندهای پژوهشی (Research Trends)

Instruction Learning • LLMs و survey ها در زمان نگارش هنوز به طور 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 چالش‌های مطرح شده در حوزه‌ی ABSA ارائه نمی‌دهد و نقش اصلی آن بیشتر تشخیص و صورت‌بندی دقیق مسئله است تا ارائه روش‌های حل، اما با تحلیل نظاممند ادبیات موجود، روندهای اصلی پژوهشی این حوزه را به روشنی شناسایی و جهت‌گیری تحقیقات آینده را مشخص می‌کند.
- بر اساس نتایج این Survey، مسیر تکامل پژوهش‌های ABSA به سمت موارد زیر حرکت کرده است:
 - ✓ نخست، پژوهش‌ها از تسک‌های تکی و مجزا به سوی تسک‌های ترکیبی و استخراج هم‌زمان چند مؤلفه (sentiment و opinion aspect) پیش‌رفته‌اند تا تحلیل دقیق‌تری از متن حاصل شود.
 - ✓ دوم، رویکردهای مولد (Generative Approaches) به‌ویژه مدل‌های Seq2Seq مانند T5، به عنوان جایگزینی برای روش‌های برچسب‌گذاری سنتی مطرح شده‌اند و امکان یکپارچه‌سازی تسک‌ها را فراهم کرده‌اند.
 - ✓ سوم، نیاز به چارچوب‌های یکپارچه (Unified Frameworks) که بتوانند با یک مدل واحد چندین زیرتسک ABSA را پوشش دهند، به عنوان یک جهت‌گیری 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 می‌کند که پس از انتشار این Survey، به سرعت مورد توجه پژوهشگران قرار گرفت و امکان یادگیری تسک‌ها از طریق توضیح زبانی را فراهم ساخته است.
 - ✓ در نهایت، Survey به استفاده روزافزون از مدل‌های زبانی بزرگ (LLM-based Methods) مانند GPT اشاره می‌کند که افکار جدیدی را برای حل مسئله‌ی ABSA، به‌ویژه در سناریوهای کم‌داده و چنددامنه، گشوده‌اند.
 - ✓ در مجموع، این مقاله با ترسیم یک نقشه راه روشن، زمینه‌ی نظری لازم برای ظهور روش‌هایی مانند InstructABSA و مطالعات مبنی بر LLM‌ها را فراهم کرده است.

۲ تحلیل و خلاصه مقاله شماره دو: LLM for ABSA

۱-۲ 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 LLM‌های عمومی مثل GPT میتوانند با مدل‌های تخصصی ABSA رقابت کنند؟

چراکه ABSA برای تحقیقات بازار و تحلیل نظرات مشتریان خیلی مهم بوده و اگه LLM‌ها 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.

نویسندهای این مقاله به طور خاص می‌خواهند سه رویکرد را با هم مقایسه کنند:

۱. استفاده از LLM‌ها در حالت zero-shot/few-shot
۲. یک LLM عمومی (GPT-3.5) روی داده‌ی ABSA fine-tuning
۳. مقایسه این روش‌ها با مدل تخصصی InstructABSA به عنوان SOTA قبلی

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

۲-۲ روش کار

نویسندهای این مقاله 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:

- (prompt با هیچ مثالی، فقط zero-shot)
- (prompt با ۶ مثال در few-shot)

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

٣-٢ نتایج کلیدی

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

جدول مقایسه F1-Score

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

جدول مقایسه Prompt Engineering با

طبق جداول فوق، مهم‌ترین نتیجه‌ی مقاله این است که:

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

- تقریباً از fine-tuning خیلی مهم است، اما نقش آن بعد از zero/few-shot prompt engineering بین می‌رود
- نیازی به prompt پیچیده ندارند fine-tuned models
- این مقاله به خوبی 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 بزرگ‌کارآمدتر هستند
 - Instruction Learning با مدل‌های کوچک‌تر یک راه حل مهندسی‌شده و اقتصادی‌تر است
 - بنابراین، طبق توصیه محققان این مقاله:
 - اگر داده داریم \leftarrow Fine-tuning (بهترین عملکرد، هزینه کمتر)
 - اگر داده نداریم (برای زبان‌های کم منبع) \leftarrow Few-shot prompting (شروع سریع)

۴-۲ تحلیل اقتصادی

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

- اگر فقط به دقت نگاه کنیم، GPT-3.5 Fine-Tuned بهترین نتیجه را دارد. اما اگر هزینه به ازای دقت را در نظر بگیریم:
- InstructABSA از GPT-4 مفروض به صرفه‌تر 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

۱-۳ مقدمه و هدف

مشابه با مقاله مژوی 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 تسک‌های را بهبود داده و به بعضی از چالش‌های مطرح شده در survey فوق پاسخ بدهد:

ردیف	تسک	نام کامل	ورودی	خر裘ی
۱	ATE	Aspect Term Extraction	جمله	لیست جنبه‌ها
۲	ATSC	Aspect Term Sentiment Classification	جمله + جنبه	احساس (مثبت/منفی/ختنی)
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	سه‌تایی (جنبه، نظر، احساس)
۷	ACOSQE	Aspect Category Opinion Sentiment Quad Extraction	جمله	چهارتایی 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	در ک دقيق نظرات کاربران
نظرسنجی محصولات در آمازون	پایش برندهای اجتماعی 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 توصیه‌گر هوشمند	انعطاف‌پذیری با حوزه‌های جدید

۲-۳ 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 انعطاف‌پذیر با حوزه یا نیاز خاص تطبیق می‌دهیم.

نوآوری اینجاست: استفاده همزمان از مثال‌های مثبت، منفی و خنثی داخل instruction

مثال دستورالعمل: "برای جمله زیر، احساس نسبت به جنبه {جنبه} را شناسایی کن: {متن}"

۳-۳ نتایج کلیدی

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

جدول مقایسه F1-Score و Accuracy

- این نتایج با مدل T5 با M ۲۰۰ پارامتر به دست آمده است، در حالی که مدل‌های قبلی B ۱.۵ پارامتر داشتند (۷ برابر بزرگتر)
- GPT2-med با ۱.۵ میلیارد پارامتر، ضعیف‌ترین است. این نشون میده پارامتر بزرگتر همیشه بهتر نیست.
- نشان می‌دهد که با طراحی هوشمند prompt و استفاده از مثال‌های متنوع، می‌توان با مدل‌های InstructABSA کوچک به نتایج بهتر از مدل‌های بزرگ رسید.
- اگر مثال‌هایی با برچسب اشتباه به مدل بدھیم، عملکرد مدل ۱۰٪ افت می‌کند. این نشان می‌دهد مدل واقعاً از مثال‌ها یاد می‌گیرد و طراحی درست instruction می‌تواند به اندازه‌ی معماری پیچیده مهم باشد
- کارایی نمونه (Sample Efficiency): به صورت میانگین در تسک‌های مختلف ABSA، با ۵۰٪ داده، نتایج رقابتی با روش‌های دیگر حاصل می‌شود. این یعنی نیاز به داده برچسب‌خورده کمتر شده، هزینه annotation کاهش پیدا کرده و برای زبان‌هایی با داده کم (مثل فارسی!) نیز مفید خواهد بود
- مدل محدودیت‌هایی هم مشتمل بر موارد زیر دارد:
 - ✓ در تسک‌های پیچیده‌تر (AOPE, AOSTE) عملکرد مدل ضعیف‌تر است
 - ✓ مدل فقط روی دیتاست‌های SemEval تست شده است
 - ✓ فقط روی زبان انگلیسی تست شده است

۴ جمع‌بندی و روند تکاملی مقالات

مقاله موری 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 راه حل‌ها مواجه است.

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

مقاله سال ۲۰۲۳ در ادامه‌ی طبیعی مسیر ترسیم شده 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 fine-tuning ضروری است state-of-the-art
- مدل GPT-3.5 fine-tuned 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 پارامتر 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 2022 Survey مطرح شده بود.
 - و از سوی دیگر، تمایل به استفاده از صورت‌بندی مولد که در مطالعات LLM-based برگسته شده است.
- Instruction Learning با بهره‌گیری از InstructABSA یک مدل نسبتاً کوچک T5، نشان می‌دهد که می‌توان:

- همهی زیرتستک‌های 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 LLM‌های بسیار بزرگ و با دقیقی نزدیک یا حتی بالاتر از روش‌های قبلی
- این مقاله را می‌توان نقطه‌ی تعادل میان نظریه، عملکرد و هزینه در مسیر تکامل ABSA دانست.

به صورت زمانی، ادبیات ABSA از تعریف و شناسایی مسئله (2022)، به آزمون توان حداکثری LLM‌ها (2023) و در نهایت به ارائه‌ی یک راه حل یکپارچه، کارآمد و مهندسی شده مبتنی بر (2024) Instruction Learning حرکت کرده است. با این حال، چالش Exposure Bias در تستک‌های پیچیده‌تر مانند استخراج سه‌تایی و چهارتایی عناصر همچنان باقی است.

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

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 مقرن به صرفه‌تر است، در حالی که برای prototyping سریع یا زبان‌های کم منبع، LLMs با few-shot prompting گزینه بهتری هستند.

جدول مقایسه‌ای راه حل‌های ABSA و LLMs در رفع چالش‌های مطرح

برنده	راه حل (GPT)	راه حل	توضیح چالش	چالش
InstructABSA 	Few-shot Prompting با ۶ مثال عملکرد قابل قبول. برای بهترین نتیجه،	با ۵۰٪ داده به نتایج رقابتی میرسد. مثال‌های مثبت/امنی/خشنی در	نیاز به داده بر چسب دار زیاد برای آموزش	کمبود داده

	روی کل fine-tuning 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(مثال "گرونه" ← 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 گستردگی. کمتر وابسته به دامنه خاص	عملکرد خوب در-cross-domain مثال آموزش روی Rest14، تست روی Lapt14 با افت معقول	انتقال مدل از یک دامنه (رستوران) به دامنه دیگر (لپ‌تاپ)	Cross-domain