



تحليل احساسات

آرش اصغری (۴۰۴۰۲۹۱۴)

ارائه درس پردازش زبان طبیعی

استاد غیاثی راد

معرفی حوزه

- یکی از زیرشاخه‌های اصلی پردازش زبان طبیعی با هدف تعیین نگرش و احساسات واقعی نویسنده پشت کلمات
- دارای سه سطح تحلیل:
 - تحلیل قطبیت: دسته‌بندی به مثبت، منفی و خنثی
 - تشخیص نوع احساس: تشخیص حالات روحی، مانند خشم، شادی و ...
 - بررسی دقیق: تشخیص شدت احساسات در کنار نوع حس
- استفاده از ابزارهای NLP به عنوان زیربنا انجام تحلیل‌های متنی

کاربردهای عملی تحلیل احساسات



بازارهای مالی

پیش‌بینی قیمت سهام با تحلیل
اخبار و پست‌های شبکه‌های
اجتماعی.



سیاست و افکار عمومی

سنجش محبوبیت کاندیداها و
واکنش به سیاست‌گذاری‌ها
(Political Analysis).



کسب‌وکار و شهرت برند

درک بازخورد مشتریان برای بهبود
محصولات و افزایش درآمد. (مثال:
تحلیل نظرات آمازون و دیجی‌کالا).

به‌طور کلی در هر کجا که نیاز به فهم بازخورد کاربران می‌باشد، مدل‌های تحلیل احساسات نقش حیاتی دارند.

رویکردهای پیشنهادی:

در این ارائه، نتایج و متدولوژی‌های سه مقاله کلیدی^{1 2 3} در حوزه تحلیل احساسات مورد بررسی قرار گرفته است. در ادامه این رویکردها به ترتیب از ضعیف‌ترین مورد تا بهترین عملکرد به طور خلاصه معرفی می‌شوند.

-
1. K. L. Tan, C. P. Lee, and K. M. Lim, "A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research," *Applied Sciences*, vol. 13, no. 7, p. 4550, Apr. 2023.
 2. D. Anggraini, S. Rahmawati, and R. Kurniawan, "Natural Language Processing For Automatic Sentiment Analysis In Social Media Data," *International Journal of Intelligent Engineering and Systems*, vol. 1, no. 2, pp. 16-19, 2024.
 3. L. Huang, "Deep Learning for Text Sentiment Analysis: A Survey," in *Proceedings of the 2nd International Conference on Machine Learning and Automation*, 2025, doi: 10.54254/2755-2721/104/20241153.

روش‌های مبتنی بر لغت‌نامه:

استفاده از لغت‌نامه‌های پیش‌تعریف شده (مانند SentiNetWord) که به هر کلمه یک امتیاز مثبت یا منفی می‌دهند.

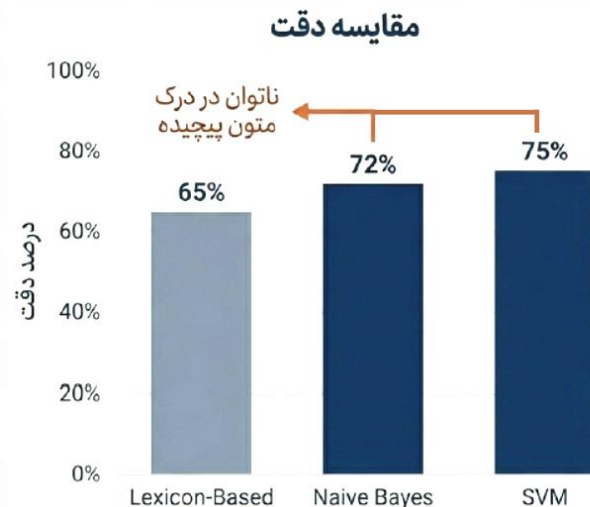
امتیاز نهایی متن از مجموع امتیاز کلمات آن به دست می‌آید.

- مزایا: پیاده‌سازی آسان و بدون نیاز به آموزش مدل
- نقاط ضعف: عدم درک کنایه، زبان عامیانه و کلمات با معانی چندگانه در بافت‌های مختلف

یادگیری ماشین سنتی:

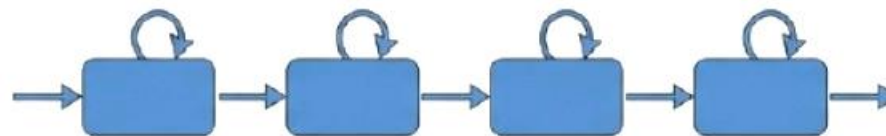
رویکرد تبدیل متن به ویژگی‌های عددی (مانند TF-IDF یا N-gram) و آموزش الگوریتم‌های ریاضی روی داده‌های برچسب‌گذاری شده است.

- الگوریتم‌های رایج: Naive Bayes، SVM و Logistic Regression
- ویژگی: این روش‌ها انعطاف‌پذیرتر از لغت‌نامه هستند و می‌توانند با رفتار کاربران در پلتفرم‌های مختلف سازگار شوند، اما برای درک متون پیچیده کافی نیستند.



یادگیری عمیق:

- شبکه‌های پیچشی (CNN): برای استخراج ویژگی‌های محلی و الگوهای خاص در متن (مانند ترکیبات خاص کلمات) استفاده می‌شود.
- حافظه طولانی کوتاه‌مدت (LSTM): به دلیل ساختار حافظه‌دار، می‌تواند توالی کلمات و روابط میان آن‌ها را در جملات طولانی درک کند.
- تفاوت: این مدل‌ها به طور خودکار ویژگی‌ها را استخراج می‌کنند و نیاز به مهندسی دستی ویژگی‌ها را کاهش می‌دهند.



مدل‌های مبتنی بر ترنسفورمر:

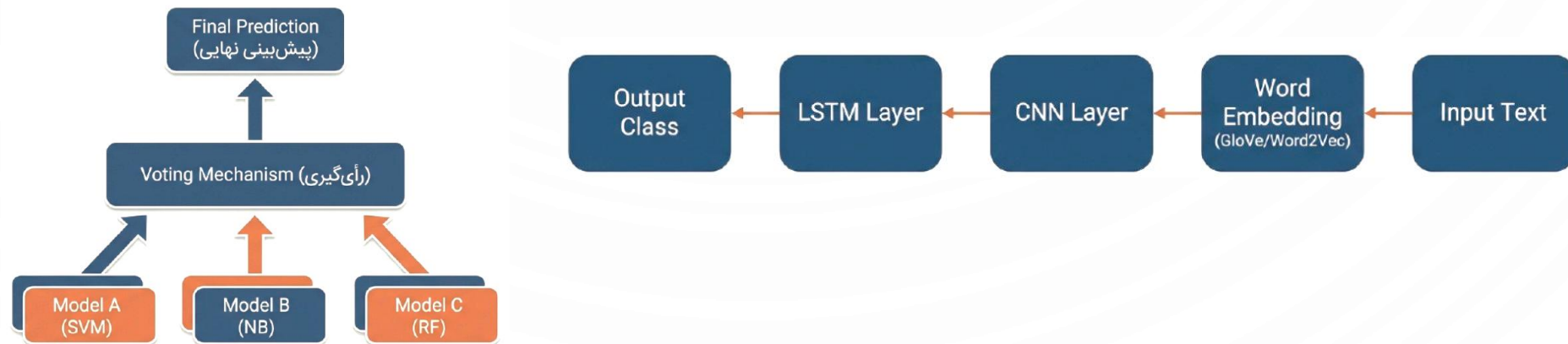
استفاده از مکانیزم توجه (Attention) برای درک دوطرفه معنای کلمات. BERT کلمات را بر اساس کلمات قبل و بعد از آنها در کل جمله تحلیل می‌کند.

- ویژگی کلیدی: توانایی بسیار بالا در درک بافتار (Context) و تفاوت‌های ظریف معنایی
- عملکرد: در حال حاضر به عنوان استاندارد طلایی (State-of-the-art) در اکثر وظایف encoding شناخته می‌شود



مدل‌های ترکیبی و یادگیری جمعی:

- مدل‌های ترکیبی (مانند BERT-BiLSTM-CRF): ترکیب قدرت درک زبانی BERT با قدرت تحلیل توالی LSTM برای رسیدن به دقت بالاتر در استخراج ویژگی‌های احساسی.
 - یادگیری جمعی (Ensemble): ترکیب پیش‌بینی‌های چندین مدل مختلف و تصمیم‌گیری بر اساس رأی‌گیری اکثریت
- هدف: پوشش دادن نقاط ضعف یک مدل توسط مدل‌های دیگر و دستیابی به بالاترین دقت ممکن



مقایسه نهایی:

| روش تحلیل | دقت | مناسب برای ... |
|--------------------------|-----------|--|
| لغت نامه | ۶۵٪ | تحلیل‌های سریع و ساده با منابع کم |
| یادگیری ماشین (NB، SVM) | ۷۲٪ - ۷۵٪ | مجموعه داده‌های کوچک و پردازش سریع |
| یادگیری عمیق (LSTM، CNN) | ۷۸٪ - ۸۱٪ | تحلیل جملات طولانی و داده‌های پیچیده |
| ترنسفورمر (BERT) | ۸۸٪ | بالاترین دقت در درک معنا و بافت متن |
| یادگیری جمعی و ترکیبی | تا ۹۸.۹٪ | پروژه‌هایی که دقت در آنها اولویت اول است |

چالش‌های باقی‌مانده:



چند زبانی

تمرکز بیشتر بر زبان انگلیسی
شده است و تفاوت‌های فرهنگی
در نظر گرفته نشده است.



تشخیص کنایه‌ها

مدل‌ها در تشخیص معنای باطنی
جملات و زبان محاوره مشکل
دارند.



تحلیل دقیق

مدل‌ها در تشخیص دقیق احساس
و شدت آن به اندازه کافی تکامل
نیافتند.



هزینه مجموعه‌داده‌ها

ساخت دیتاست‌های به اندازه کافی
بزرگ برای همه زبان‌ها توسط
انسان هزینه زیادی دارد.

مسیرهای پیش‌رو:

1. تحلیل احساسات دقیق فراتر از قطبیت
2. سبک‌سازی مدل‌ها برای اجرا بر روی موبایل
3. فراهم کردن امکان یادگیری بدون نظارت
4. تحلیل احساسات چندرسانه‌ای
5. بهبود تفسیرپذیری و شفاف‌سازی فرایند
6. شناسایی طعنه و ابهام‌ها
7. بهبود عملکرد برای همه فرهنگ‌ها

نتیجه گیری:

- تحلیل احساسات پلی حیاتی میان زبان غیرساختارمند انسان و تصمیم‌گیری‌های هوشمند تجاری و سیاسی است.
- این حوزه از روش‌های سنتی مبتنی بر لغت‌نامه و ماشین‌بردار پشتیبان به سمت مدل‌های پیشرفته یادگیری عمیق مانند BERT و LSTM حرکت کرده که دقت را به شکل چشم‌گیری افزایش داده‌اند.
- قدرت مدل‌های ترکیبی: مدل‌های Ensemble که چندین الگوریتم را با هم ترکیب می‌کنند، در حال حاضر بهترین عملکرد را در بنچمارک‌های جهانی دارند.
- با وجود پیشرفت‌ها، مدل‌های فعلی هنوز در درک زبان عامیانه، کنایه و ساختارهای پیچیده زبانی با محدودیت‌هایی روبرو هستند.
- آینده این حوزه در گرو ساخت مدل‌های تفسیرپذیرتر، سبک‌تر و چندزبانه است تا بتوان در دنیای واقعی و با منابع محدود، تحلیل‌های دقیق‌تری ارائه داد.

محدوده پوشش ارائه:

این ارائه بخش عمده‌ای از مباحث پایه و پیشرفته‌ی تحلیل احساسات را پوشش می‌دهد، از جمله:

- تعریف مسئله و سطوح مختلف تحلیل احساسات
- رویکردهای مبتنی بر لغت‌نامه، یادگیری ماشین، یادگیری عمیق و ترنسفورمرها
- بررسی چالش‌ها و مسیرهای آینده این حوزه
- در این ارائه تمرکز اصلی بر تحلیل احساسات متنی بوده و موضوعاتی مانند موارد زیر مورد بررسی قرار نگرفته‌اند:
- تحلیل احساسات چندرسانه‌ای (صوت و تصویر)
- تحلیل بر خط
- تحلیل و پیاده‌سازی صنعتی در مقیاس بزرگ

تحلیل احساسات کاربران فارسی‌زبان در شبکه اجتماعی توییتر و مقایسه با احساس واقعی آن‌ها، نمونه‌ای کاربردی از وظایف طبقه‌بندی و درک معنا بوده و به‌طور مستقیم زیرمجموعه حوزه تحلیل احساسات محسوب می‌شود.