

Sentiment Analysis System Using LangChain and Arize

TAMUHack Team

TEAM@TAMU.EDU

*Department of Computer Science and Engineering
Texas A&M University
College Station, TX 77840, USA*

Editor: TAMU Hackathon

Abstract

We present an integrated system for analyzing customer car reviews using LangChain for orchestration and Arize for monitoring. The system processes natural language reviews to extract sentiment, categorize feedback, and provide actionable insights. By combining LangChain's structured chains with OpenAI's language models and Arize's observability tools, we achieve robust analysis with comprehensive monitoring capabilities. Our implementation demonstrates effective sentiment classification, consistent output formatting, and scalable review processing while maintaining production reliability.

Keywords: LangChain, Arize, sentiment analysis, MLOps, monitoring, OpenAI

1 Introduction

Our system addresses the challenge of analyzing customer car reviews at scale while maintaining high accuracy and operational visibility. The implementation leverages LangChain's structured prompting capabilities with OpenAI's language models and Arize's monitoring to create a robust pipeline for sentiment analysis and insight extraction.

2 System Architecture

2.1 LangChain Implementation

The core of our system uses LangChain's chain architecture to structure interactions with OpenAI's models. We implement a ReviewAnalyzer class that processes reviews through carefully crafted prompt templates and Pydantic models for consistent output formatting. The system uses zero-temperature inference to ensure deterministic outputs and employs structured prompting to guide the model toward specific analysis patterns.

2.2 Data Processing Pipeline

Our implementation processes reviews through multiple stages:

1. Initial review collection and preprocessing
2. Categorization by vehicle aspects (performance, safety, etc.)
3. Sentiment analysis using LangChain chains
4. Structured output generation with Pydantic validation
5. Frontend visualization using spider charts and interactive components

2.3 Monitoring and Observability

We integrate Arize's Phoenix framework for comprehensive monitoring:

- Distributed tracing for request flows
- Performance metrics collection
- Model output validation
- Error tracking and alerting

The monitoring setup includes:

- Real-time performance dashboards
- Custom metric tracking for model latency and throughput
- Automated alerting thresholds for error rates
- Historical trend analysis for model drift detection
- Integration with existing DevOps tooling

2.4 OpenAI Integration

The system leverages OpenAI's language models through a structured interface, ensuring consistent output formatting using Pydantic models:

- Custom ReviewAnalysis schema for structured outputs
- Zero-temperature inference for deterministic results
- Prompt templating for consistent model interactions
- Error handling with graceful fallbacks
- Rate limiting and request batching for efficiency

2.5 Arize Monitoring

We implement comprehensive monitoring using Arize's Phoenix framework:

- Model performance tracking across different review categories
- Latency and throughput metrics for each analysis stage
- Input/output distribution monitoring
- Custom dashboards for sentiment analysis accuracy
- Automated reporting for stakeholder updates

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor.

Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$\int_0^\infty e^{-\alpha x^2} dx = \frac{1}{2} \sqrt{\int_{-\infty}^\infty e^{-\alpha x^2} dx \int_{-\infty}^\infty e^{-\alpha y^2} dy} = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}}$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$\sum_{k=0}^{\infty} a_0 q^k = \lim_{n \rightarrow \infty} \sum_{k=0}^n a_0 q^k = \lim_{n \rightarrow \infty} a_0 \frac{1 - q^{n+1}}{1 - q} = \frac{a_0}{1 - q}$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-p \pm \sqrt{p^2 - 4q}}{2}$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 \Phi}{\partial t^2}$$

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Here is a citation Chow and Liu (1968).

3 Results and Analysis

Our system demonstrates robust performance across several key metrics:

3.1 Sentiment Analysis Accuracy

The implementation shows strong performance in categorizing reviews:

- Precise positive/negative mention counting
- Contextual sentiment understanding
- Representative quote extraction
- Category-specific insight generation

3.2 Frontend Visualization

The system presents analysis results through:

- Interactive spider charts showing category performance
- Detailed review breakdowns with sentiment scores
- Quote highlighting for key insights
- Real-time analysis updates

3.3 Scalability and Performance

The architecture handles varying loads efficiently:

- Batch processing capabilities
- Structured output caching
- Optimized API interactions
- Robust error handling

4 Future Work

Planned improvements include:

- Enhanced multi-language support
- More granular sentiment analysis
- Advanced bias detection
- Expanded monitoring metrics
- Real-time analysis streaming
- Integration with additional data sources

Acknowledgments and Disclosure of Funding

This work was developed during the Texas A&M Hackathon 2024. We acknowledge the support of the OpenAI, LangChain, and Arize teams in providing the tools and documentation that made this implementation possible. The project received no external funding and was created purely as an educational exercise.

Appendix A.

In this appendix we prove the following theorem from Section 6.2:

Theorem *Let u, v, w be discrete variables such that v, w do not co-occur with u (i.e., $u \neq 0 \Rightarrow v = w = 0$ in a given dataset \mathcal{D}). Let N_{v0}, N_{w0} be the number of data points for which $v = 0, w = 0$ respectively, and let I_{uv}, I_{uw} be the respective empirical mutual information values based on the sample \mathcal{D} . Then*

$$N_{v0} > N_{w0} \Rightarrow I_{uv} \leq I_{uw}$$

with equality only if u is identically 0. ■

Appendix B.

Proof. We use the notation:

$$P_v(i) = \frac{N_v^i}{N}, \quad i \neq 0; \quad P_{v0} \equiv P_{v(0)} = 1 - \sum_{i \neq 0} P_{v(i)}.$$

These values represent the (empirical) probabilities of v taking value $i \neq 0$ and 0 respectively. Entropies will be denoted by H . We aim to show that $\frac{\partial I_{uv}}{\partial P_{v0}} < 0 \dots$

Remainder omitted in this sample. See <http://www.jmlr.org/papers/> for full paper.

References

Chow, C. K. and Liu, C. N. (1968) Approximating discrete probability distributions with dependence trees, *IEEE Transactions on Information Theory*, (3), 462–467