Complaints Are All You Need

Toward LLM Generative Classification for Automotive Customer Complaints

Philip Menchaca, Prithvijit Dasgupta, Tyler Allen

Business Problem

Complaint Analysis Process Has Limitations

- Expert-labeled classification of complaints identify potential faulty systems
- Expert labeling is time-consuming and costly
- Cannot easily detect emerging issues

Data

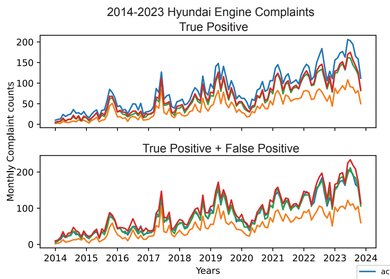
National Highway Traffic Safety Administration

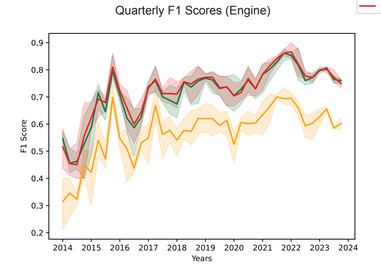
- NHTSA-labeled complaint and recall data from November 2023 (1.9 million complaints)
- Chose: Ford, GM, Tesla, Hyundai, Toyota, Honda
- Selected recall-prone component systems:
  - air bags, seat belts, service brakes, fuel system, gasoline, power train, visibility, electrical system, vehicle speed control, wheels, engine, suspension
- Class imbalance in labels

Methodology

Find Patterns & Test LLM Classification Performance

- Complaint revealed times of spikes
  - Confirmed with STA/LTA analysis
  - Times of spikes provide pattern to estimate LLM performance
- 3 prompts to test performance in different scenarios (FLAN-T5 Large)
  - Unconstrainted generation with no labels
  - Constrained generation with specific labels
  - Partially constrained generation on simulated emerging issues with withheld labels



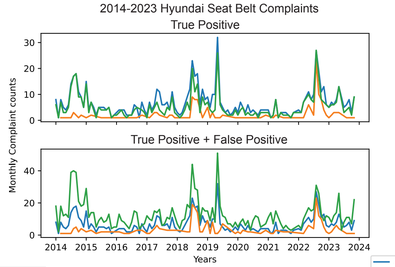


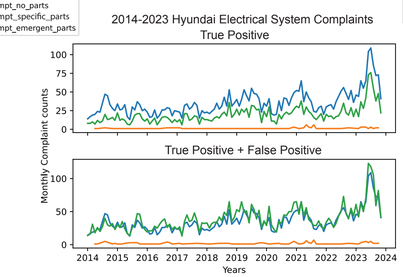
Mean Average Error of Monthly Counts of Labels

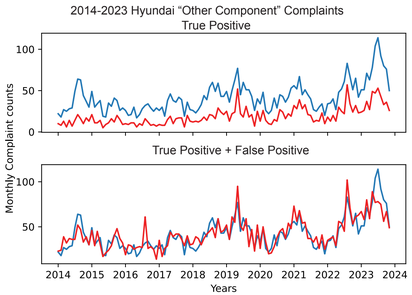
No Labels	Specific Labels	Emergent Labels
16.335	6.820	6.480

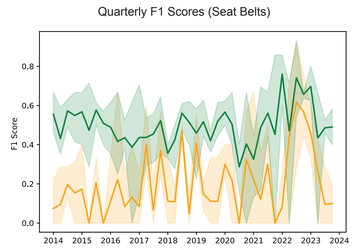
General Accuracy

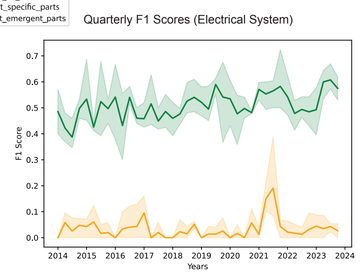
No Labels	Specific Labels	Emergent Labels
0.203	0.5895	0.594

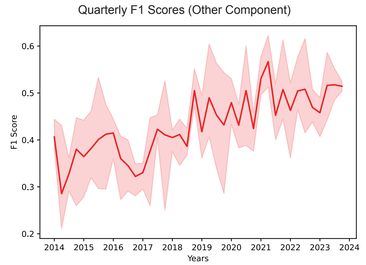












Findings

LLM Generated Labels Follow Trends

- Mediocre accuracy but reveals trends without training or requiring expert labeling
- Adding specific components to the prompt improves performance
- Able to follow trends for withheld labels ("other component")
- Performance appears to be component-specific for the data we tested
- Even when wrong, it uncovers trends, making the model potentially suitable for real-time analysis
- The F1 scores (per component) suggest that the performance improves as more customers complain about a specific component

Future Work

Are Complaints Really All You Need?

- Generalize to all components and all OEMs
- Investigate interpretability of LLM model
- Augment LLM with expert input

References

- https://nhtsa.gov/nhtsa-datasets-and-apis
- https://huggingface.co/google/flan-t5-large
- Plaza-del-Arco, Flor Miriam, Debora Nozza, and Dirk Hovy. 2023. "Leveraging Label Variation in Large Language Models for Zero-Shot Text Classification." arXiv.Org. July 24, 2023. https://arxiv.org/abs/2307.12973v1.