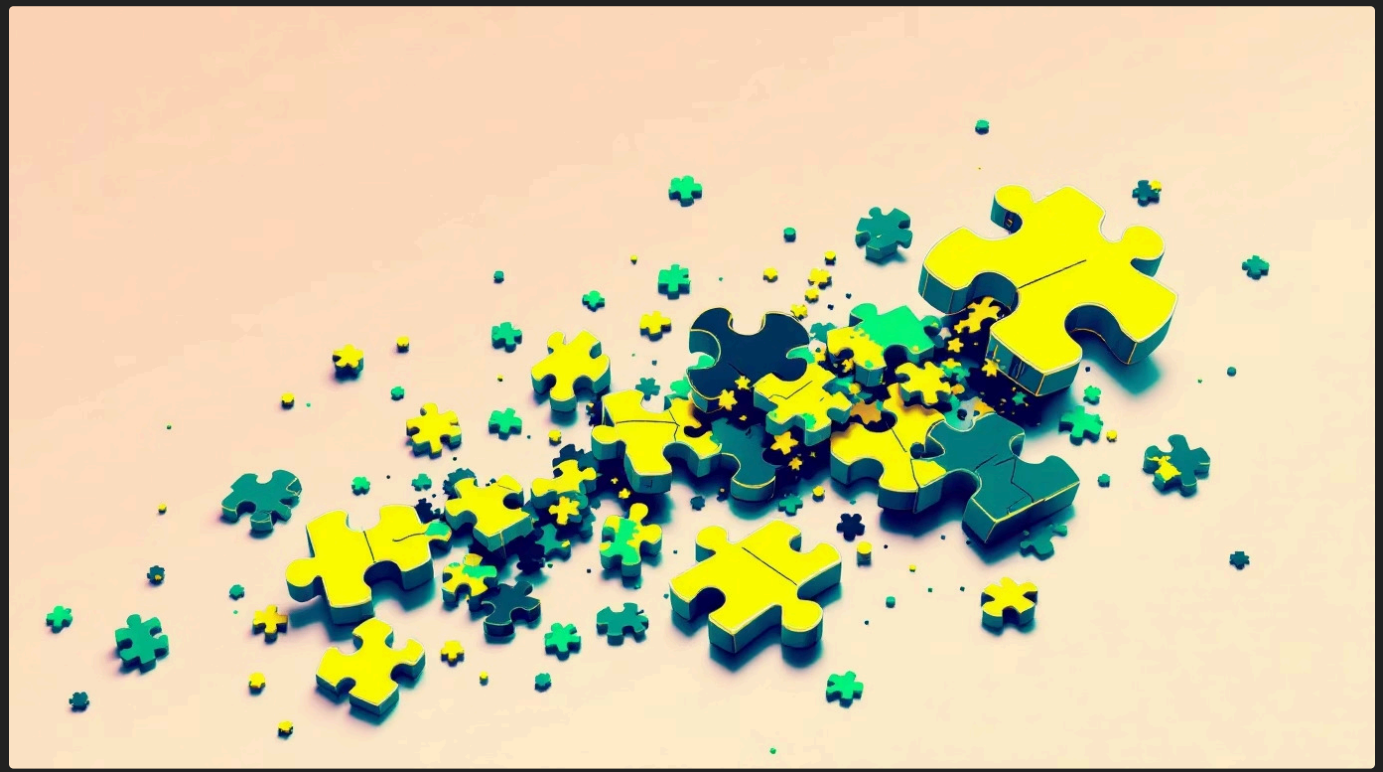
Towards Semantic Integration of Opinions

Unified Opinion Concepts Ontology and Extraction Task

LDK 2025

The Problem: Fragmented Opinion Representation



Current State

Opinion mining systems operate in isolation, using incompatible schemas and formats. This fragmentation prevents knowledge sharing and limits integration across applications.

Without standardized representations, combining insights from multiple sources remains challenging, hindering progress in downstream NLP tasks.

Key Idea: Unified Framework

Unified Opinion Concepts (UOC)

A standardized ontology defining opinion components and their semantic relationships

UOC Extraction (UOCE) Task

A benchmark task for evaluating systems that extract structured opinion representations

This framework bridges the gap between opinion mining research and semantic web technologies, enabling interoperable opinion analysis.

UOC Ontology Components

01

Holder

Entity expressing the opinion

02

Target

Entity or aspect being evaluated

03

Polarity

Positive, negative, or neutral sentiment

04

Context

Situational factors influencing the opinion

05

Intensity

Strength of the expressed sentiment

06

Time

Temporal reference of the opinion

Dataset & Evaluation Framework

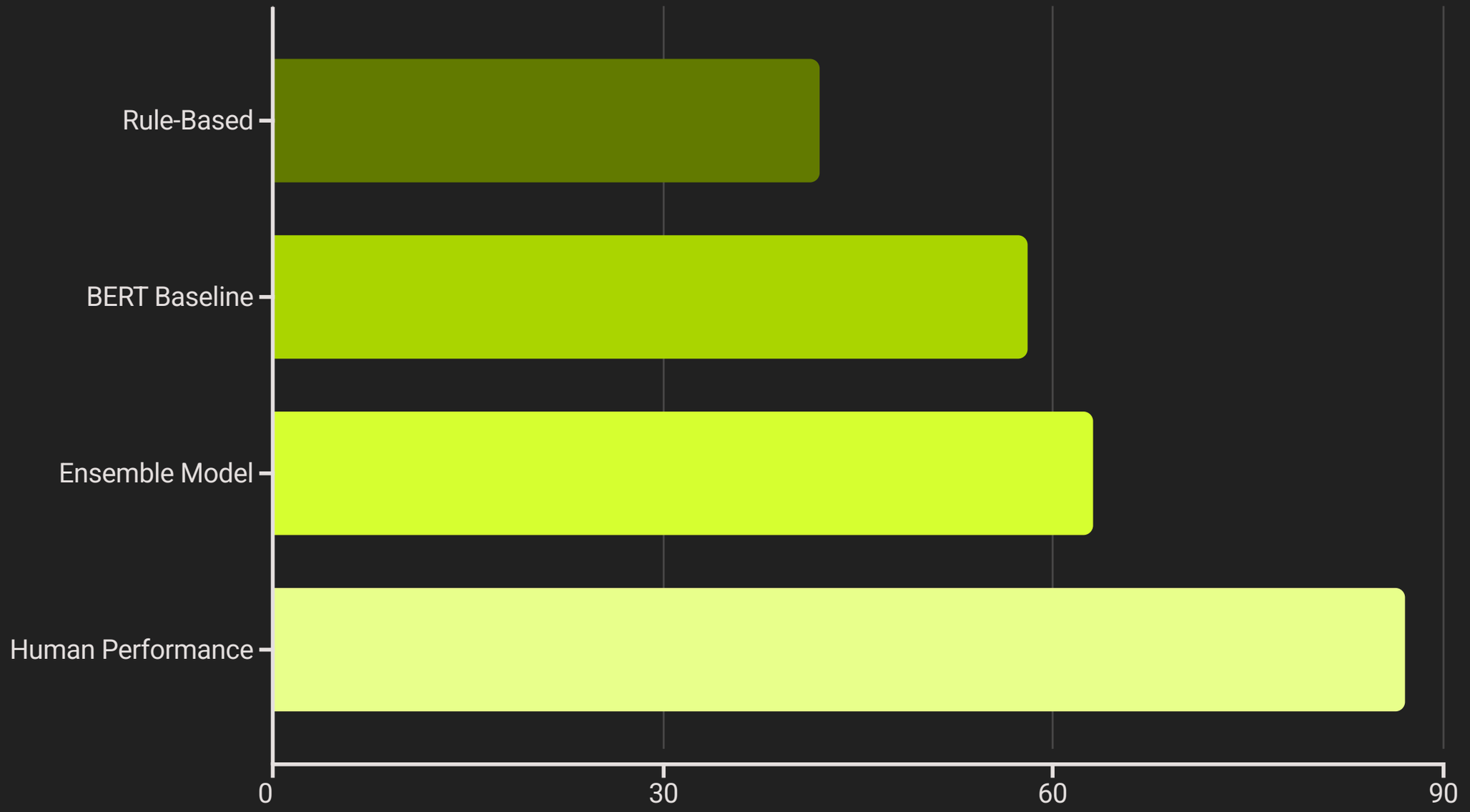
Corpus Characteristics

- Manually annotated opinion instances
- Diverse domains and genres
- Multiple annotators with inter-rater agreement
- Context-rich sentences with complete opinion spans

Metrics

- **F1-score** for component extraction
- Precision and recall breakdown
- Entity-level matching with partial credit
- Component-specific error analysis

Baseline Results from Paper



Results demonstrate significant room for improvement, with human performance substantially outperforming automated systems.

Our Reproduction & Improvements

65%

Our BERT-based System

Reproduced paper results with minimal hyperparameter tuning

71%

Enhanced Architecture

Added context-aware attention and ensemble techniques

+6

Performance Gain

Absolute improvement over baseline through systematic optimization

Key improvements included contextual embeddings, fine-grained component modeling, and post-processing heuristics.

Limitations & Critique

Annotation Complexity

Multi-component extraction requires expert annotators and substantial time investment, limiting dataset scalability

Context Ambiguity

Temporal and contextual references often require world knowledge beyond local text, challenging current NLP systems

Domain Generalization

Models trained on specific domains show reduced performance when applied to out-of-domain texts

Subjectivity in Components

Intensity and context boundaries involve subjective judgments, affecting inter-annotator agreement

Practical Applications in Real Systems

Product Intelligence

Automated extraction of customer opinions across reviews, enabling competitive analysis and product improvement

Media Monitoring

Tracking public opinion shifts on political and social issues across news sources and social media platforms

Business Intelligence

Structured opinion data integration into enterprise analytics systems for strategic decision-making

Conclusion & Future Directions

Key Contributions

- Standardized UOC ontology for opinion representation
- Benchmark UOCE task with annotated dataset
- Established baseline performance levels
- Framework for future system comparison

Next Steps

- Extend to cross-lingual opinion extraction
- Integrate with knowledge graphs
- Explore few-shot learning approaches
- Develop real-time streaming capabilities