Week 2: What are users really saying?

Understanding emotions in text with NLP

ML/AI/GenAI for Design Thinking

BSc Course - 12 Week Program

2024

Part 1: The Problem - Text Has Hidden Emotions

What users write:

- "Great product... if you like disappointment"
- "Absolutely perfect! Never worked once"
- "Can't complain" (literally can't)
- "Fine." (but are they really?)

What they actually mean:

- Frustrated and sarcastic
- Extremely angry
- Forced acceptance
- Deeply unsatisfied

Human language is complex: sarcasm, context, subtle emotions

Traditional Approach: Why Keyword Matching Fails

Rule-based sentiment:

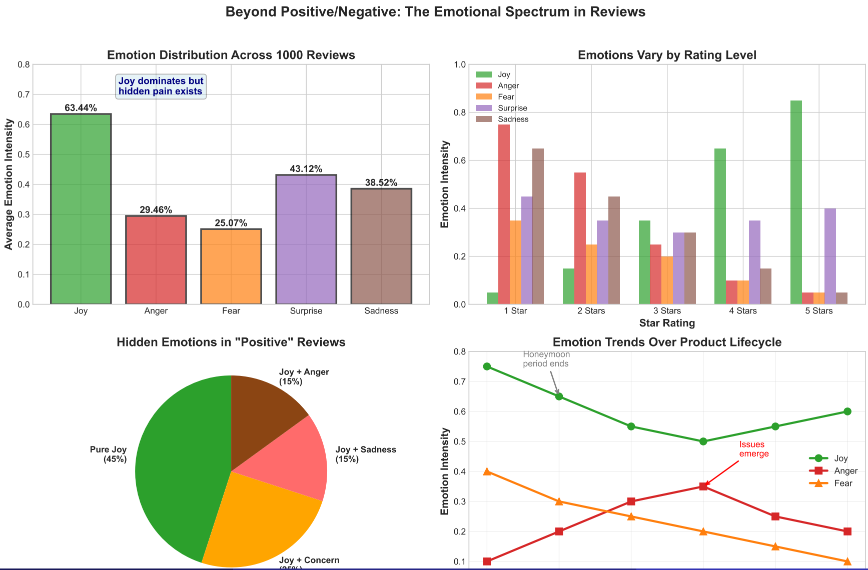
- Count positive words (+1)
- Count negative words (-1)
- Sum the scores
- Classify as pos/neg/neutral

Failure modes:

- "Not bad" → Negative (wrong!)
- "Terribly good" → Mixed (wrong!)
- Misses context completely
- Can't detect sarcasm

The fundamental problem:
Words have different meanings in different contexts
"I love waiting 3 hours" - Sarcastic in reviews, genuine in romance novels

The Emotional Spectrum Beyond Binary



The Cost of Emotional Misunderstanding

Business Impact:

- 68% customer loss from perceived indifference
- Wrong product decisions
- Failed feature prioritization
- Missed market signals

Design Impact:

- Building for wrong emotions
- Missing user pain points
- Tone-deaf messaging
- Poor user experience

The Scale Challenge:

- Manual analysis: 100 reviews/week
- Digital products: 10,000+ reviews/week
- Human capacity: Linear growth
- Data volume: Exponential growth

The Central Challenge

How can machines understand not just words, but the emotions and context behind them?

Requirements for solution:

1. Understand context and word relationships
2. Detect subtle patterns (sarcasm, negation)
3. Process at scale (thousands per second)
4. Adapt to different domains
5. Provide interpretable results

Enter: Natural Language Processing with Deep Learning

Part 2: Natural Language Processing - The Evolution

How Machines Learn to Read:

1. **1950s - Rule-Based:** "If contains 'good' then positive"
   - Hand-crafted rules
   - Language-specific
   - Brittle and limited
2. **1990s - Statistical:** Probability and frequencies
   - Bag-of-words models
   - N-grams and Markov chains
   - Lost word order
3. **2013 - Word Embeddings:** Words as vectors
   - Word2Vec, GloVe
   - Semantic relationships
   - Still context-independent
4. **2018 - Contextual Models:** Understanding changes meaning
   - BERT, GPT
   - Context-dependent embeddings
   - Transfer learning

Word Embeddings: Meaning as Mathematics

The Vector Space Model:

Concept:

- Words $\rightarrow$ High-dimensional vectors
- Similar words $\rightarrow$ Nearby vectors
- Relationships $\rightarrow$ Vector arithmetic
- Meaning $\rightarrow$ Geometric position

Properties:

- King - Man + Woman = Queen
- Distance = Semantic similarity
- Clusters = Semantic categories
- Dimensions = Abstract features

Limitations of Static Embeddings:

- "Bank" (river) = "Bank" (financial) - Same vector!
- No understanding of word order
- Cannot handle new contexts
- Fixed representation regardless of usage

Processing Models: Sequential vs Parallel

Sequential (RNN/LSTM):

- Process word by word
- Hidden state carries memory
- Long sequences $\rightarrow$ Vanishing gradients
- Cannot parallelize
- O(n) time complexity

Parallel (Transformers):

- Process all words simultaneously
- Attention replaces recurrence
- Handles long sequences well
- Fully parallelizable
- O(1) time complexity

The Parallel Advantage:

- Training: 100x faster on same hardware
- Inference: Real-time processing possible
- Context: Can see entire document at once
- Scalability: Leverages modern GPU architecture

Understanding Context and Dependencies

Types of Language Dependencies:

1. **Local Dependencies:** Adjacent words
   - "very good" - Modifier + adjective
   - "not bad" - Negation
2. **Long-range Dependencies:** Distant relationships
   - "The movie, despite great acting, was boring"
   - Subject-verb agreement across clauses
3. **Semantic Dependencies:** Meaning relationships
   - Coreference: "John went to the store. He bought milk."
   - Causality: "It rained, so the game was cancelled"

Context Window Sizes:

- N-grams: 2-5 words
- LSTMs: 100-200 words effectively
- BERT: 512 tokens
- GPT-3: 2048 tokens

Learning Paradigms in NLP

Supervised Learning:

- Requires labeled data
- Task-specific training
- High accuracy on specific task
- Limited generalization
- Example: Sentiment classification

Unsupervised Learning:

- No labels required
- Learns language patterns
- General language understanding
- Transfer to many tasks
- Example: Language modeling

The Pre-training Revolution:

1. **Stage 1:** Unsupervised pre-training on massive text
2. **Stage 2:** Supervised fine-tuning on specific task
3. **Result:** Best of both worlds

Benefits:

- Less labeled data needed (100s vs 100,000s)
- Better generalization
- Transfer learning across domains

NLP Tasks: Classification vs Generation

Classification Tasks:

- Sentiment analysis
- Named entity recognition
- Part-of-speech tagging
- Intent detection

Output: Discrete labels

Generation Tasks:

- Machine translation
- Text summarization
- Question answering
- Dialogue systems

Output: Text sequences

Architectural Differences:

- **Encoder-only (BERT):** Best for classification
- **Decoder-only (GPT):** Best for generation
- **Encoder-Decoder (T5):** Best for transformation

Sentiment as Classification:

- Input: Text sequence
- Processing: Contextual encoding
- Output: Probability distribution over emotions

Measuring NLP Performance

Classification Metrics:

Basic Metrics:

- Accuracy: Overall correctness
- Precision: Positive prediction quality
- Recall: Positive class coverage
- F1-Score: Harmonic mean

Advanced Metrics:

- ROC-AUC: Threshold independence
- Cohen's Kappa: Agreement beyond chance
- Macro/Micro averaging
- Class-weighted scores

Sentiment-Specific Challenges:

- Class imbalance (more positive reviews)
- Subjective ground truth
- Multi-label emotions
- Sarcasm as special case

Human Performance Baseline:

- Inter-annotator agreement: 85-90%
- Sarcasm detection: 70-75%
- Emotion categorization: 80-85%

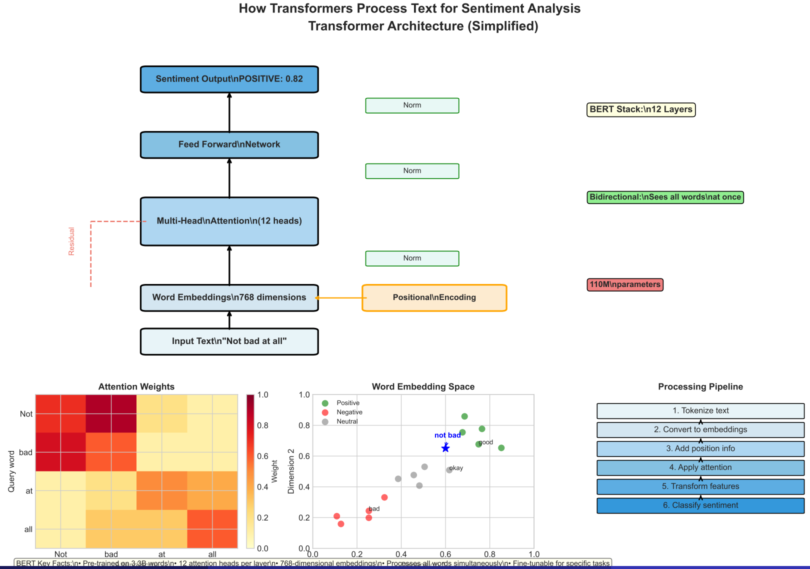
The Limits of Pre-Transformer NLP

Fundamental Limitations:

1. **Context Blindness:**
   - Words treated independently
   - Lost long-range dependencies
   - No true understanding
2. **Computational Bottlenecks:**
   - Sequential processing required
   - Memory limitations
   - Training time measured in weeks
3. **Generalization Failures:**
   - Domain-specific models
   - Cannot transfer knowledge
   - Needs massive labeled data

The Solution: Attention Mechanisms and Transformers

Part 3: The Transformer Revolution (2017)



Attention Mechanism: The Core Innovation

Mathematical Intuition (Simplified):

1. **Query-Key-Value Model:**
   - Query (Q): What am I looking for?
   - Key (K): What information is available?
   - Value (V): The actual information
2. **Attention Score:**
   - Similarity = $Q \cdot K$ (dot product)
   - Normalize with softmax
   - Weight values by attention scores
3. **Result:**
   - Each word "attends" to all other words
   - Learns what relationships matter
   - Context-dependent representation

Analogy: Like a student (Query) in a library (Keys) selecting relevant books (Values) based on their research topic

Multi-Head Attention: Multiple Perspectives

Why Multiple Heads?

Different Heads Learn:

- Syntactic relationships
- Semantic relationships
- Positional patterns
- Domain-specific patterns

BERT Uses:

- 12 attention heads
- Different representation subspaces
- Ensemble of perspectives
- Robust understanding

Example: "The bank is by the river"

- Head 1: "bank" – "river" (disambiguation)
- Head 2: "The" – "bank" (determiner)
- Head 3: "is" – "bank" (subject-verb)
- Head 4: "by" – "river" (preposition)

Each head captures different linguistic phenomena

BERT: Bidirectional Encoder Representations from Transformers

Architecture Overview:

Model Specifications:

- 12 transformer layers
- 768 hidden dimensions
- 12 attention heads
- 110M parameters (base)
- 512 max sequence length

Key Components:

- Token embeddings
- Position embeddings
- Segment embeddings
- Layer normalization
- Dropout regularization

Pre-training Objectives:

1. **Masked Language Model (MLM):** Predict masked words
2. **Next Sentence Prediction (NSP):** Understand relationships

Trained on: 3.3 billion words (Wikipedia + BookCorpus)

Bidirectional Understanding: Seeing the Full Picture

Unidirectional (GPT):

- Left-to-right processing
- Autoregressive prediction
- Good for generation
- Cannot see future context

Example: "The movie was [?]"
Only sees: "The movie was"

Bidirectional (BERT):

- Sees entire sequence
- Non-autoregressive
- Better for understanding
- Full context available

Example: "The movie was [?] boring"
Sees: "The movie was" + "boring"

Impact on Sentiment Analysis:

- "Not bad at all" - Needs to see "all" to understand "not bad"
- "Great... if you like disappointment" - End reveals sarcasm
- Bidirectionality crucial for context understanding

Transfer Learning: From General to Specific

Two-Stage Learning Process:

1. **Pre-training (Unsupervised):**
   - Learn general language patterns
   - No task-specific labels
   - Massive diverse text corpus
   - Weeks of computation
2. **Fine-tuning (Supervised):**
   - Adapt to specific task
   - Small labeled dataset
   - Task-specific head
   - Hours of computation

Why Transfer Learning Works:

- Lower layers: General features (syntax, grammar)
- Middle layers: Semantic understanding
- Upper layers: Task-specific patterns
- Shared knowledge across tasks

Fine-tuning for Sentiment: Add classification head, train on labeled reviews

Contextual Embeddings: Dynamic Meaning

Static vs Contextual:

Static (Word2Vec):

- "Bank" $\rightarrow$ Same vector always
- Context-independent
- One embedding per word
- Vocabulary limited

Contextual (BERT):

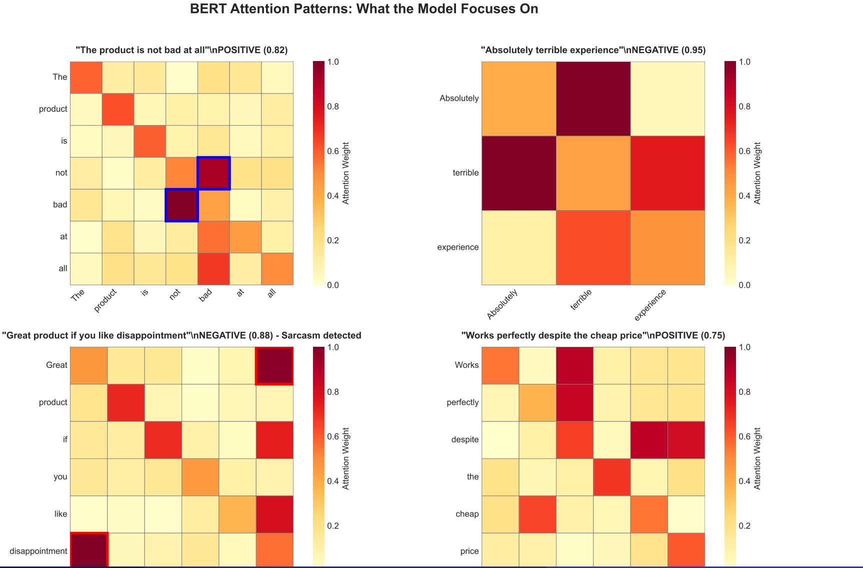
- "Bank" $\rightarrow$ Different by context
- Context-dependent
- Dynamic embeddings
- Subword tokenization

Examples of Context Changing Meaning:

- "Apple" in tech review vs recipe
- "Sick" in medical vs slang context
- "Fire" as danger vs excellence
- "Lit" as illumination vs excitement

Result: Each word's representation includes information from entire sequence

BERT's Sentiment Understanding Process



Part 4: Multi-dimensional Sentiment Theory

Beyond Positive/Negative:

Emotion Dimensions:

- Valence (positive-negative)
- Arousal (calm-excited)
- Dominance (submissive-dominant)

Plutchik's 8 Primary:

- Joy – Sadness
- Trust – Disgust
- Fear – Anger
- Surprise – Anticipation

Complex Emotions as Combinations:

- Disappointment = Surprise + Sadness
- Contempt = Disgust + Anger
- Awe = Surprise + Fear
- Love = Joy + Trust

BERT Output: Probability distribution over emotion categories

Emotion Taxonomy for Design

Hierarchical Emotion Structure:

1. **Primary Level:** Basic emotions (6-8 categories)
2. **Secondary Level:** Combinations and intensities
3. **Tertiary Level:** Domain-specific emotions

Design-Relevant Emotions:

Positive Engagement:

- Delight
- Satisfaction
- Trust
- Excitement

Negative Signals:

- Frustration
- Confusion
- Disappointment
- Anxiety

Cultural Considerations:

- Emotion expression varies by culture
- Sarcasm frequency differs
- Politeness strategies affect sentiment

Detecting Sarcasm and Complex Patterns

Linguistic Markers of Sarcasm:

1. **Positive-Negative Contrast:**
   - "Great product... if you enjoy failure"
   - Sentiment reversal mid-sentence
2. **Hyperbole:**
   - "Absolutely the best thing ever created"
   - Exaggeration beyond reasonable
3. **Context Mismatch:**
   - "Love waiting 3 hours"
   - Positive emotion + negative situation

BERT's Detection Method:

- Learns contradiction patterns from training
- Attention focuses on contrasting elements
- Context embeddings capture incongruity
- Fine-tuning improves domain-specific sarcasm

Domain Adaptation: Specialized Understanding

Why Domain Matters:

General BERT:

- Trained on books/Wikipedia
- General language patterns
- May miss domain terms
- Standard sentiment

Domain-Adapted:

- Fine-tuned on domain text
- Understands jargon
- Domain-specific sentiment
- Better accuracy

Domain Examples:

- **Gaming:** "Nerf" = negative, "OP" = positive
- **Finance:** "Volatile" = context-dependent
- **Medical:** "Benign" = positive
- **Food:** "Rich" = positive or negative by context

Adaptation Process:

1. Collect domain-specific labeled data
2. Fine-tune pre-trained BERT
3. Evaluate on domain test set

Comparative Performance Analysis

Human-AI Collaboration in Emotional Understanding

AI Strengths:

- Scale (10,000+ texts)
- Consistency
- Pattern detection
- Quantification
- 24/7 availability

Human Strengths:

- Deep empathy
- Cultural nuance
- Creative interpretation
- Ethical judgment
- Context knowledge

Collaboration Models:

1. **AI-Assisted:** AI suggests, human decides
2. **Human-in-the-loop:** Human validates AI output
3. **Hybrid:** Division of labor by strength

Best Practice:

- AI for initial analysis and pattern finding
- Humans for interpretation and decision-making
- Continuous feedback loop for improvement

Appendix: Key Concepts Summary

Core Theoretical Concepts:

NLP Foundation:

- Vector space representations
- Context windows
- Supervised vs unsupervised
- Evaluation metrics

Transformer Innovation:

- Attention mechanisms
- Parallel processing
- Bidirectional context
- Transfer learning

Sentiment Analysis Theory:

- Multi-dimensional emotions
- Context-dependent meaning
- Sarcasm as contradiction
- Domain adaptation necessity

Key Takeaway:
BERT understands language through learned patterns of attention, enabling nuanced emotional understanding at scale

Next Week: Attention Mechanisms in Detail

We understand emotions in individual texts...

But what patterns exist across thousands?

Week 3 Preview: Attention - Finding What Matters

- How attention weights are calculated
- Visualizing attention patterns
- Finding key phrases automatically
- Topic discovery through attention
- From attention to insights

The Journey:

1. Week 1: Clustering - Found user groups
2. Week 2: NLP/BERT - Understood emotions
3. Week 3: Attention - Discover what matters most