Word2Vec on Federal Reserve Text Data: A Domain-Specific Embedding Analysis

JUIYU CHENG

January 9, 2026

1 Introduction

Word embeddings capture semantic relationships between words by representing them as dense vectors in a continuous space. The classic example "king - man + woman ≈ queen" demonstrates how vector arithmetic can reveal analogical relationships. In this project, we apply Word2Vec to Federal Reserve text data to see whether the model can learn relationships specific to monetary policy.

Our corpus consists of:

- **FOMC Statements**: Official policy announcements
- **FOMC Minutes**: Detailed meeting summaries
- **Fed Speeches**: Remarks by Powell, Yellen, Bernanke, and other officials

2 Data Collection and Parsing

2.1 Data Sources

We collected text data from the Federal Reserve's official website, including documents from three Fed Chairs (Bernanke, Yellen, Powell) and various FOMC communications. The raw data was organized into subdirectories:

- `./data/fed_text/fomc/` – Contains statements, minutes, and press conference transcripts
- `./data/fed_text/speeches/` – Contains speeches organized by speaker

2.2 Parsing Pipeline

Our parsing pipeline, implemented in Apache Spark with Scala, performs the following steps:

1. **Recursive File Discovery**: Traverse all subdirectories to find `.txt` files
2. **Metadata Extraction**: Parse filenames to extract date, document type, speaker, and meeting ID
3. **Paragraph Segmentation**: Split documents by double newlines, filtering paragraphs with fewer than 50 characters
4. **DataFrame Creation**: Store records with schema: `(doc_type, date, speaker, meeting_id, source_file, text)`

Corpus Statistics:

- Total Documents (training units): 350
- Document types: statements, minutes, speeches
- Speakers: Powell, Yellen, Bernanke, FOMC (collective)

2.3 Fed-Specific Tokenization

Standard tokenization loses important Fed terminology. We implemented custom tokenization with:

1. **Manual Phrase Mappings** (50+ phrases):
   - "federal funds rate" → `federal_funds_rate`
   - "quantitative easing" → `quantitative_easing`
   - "balance sheet" → `balance_sheet`
   - "forward guidance" → `forward_guidance`
2. **Percentage Normalization**:
   - "2%" → `2_percent`
   - "25 basis points" → `25_basis_points`
3. **Acronym Preservation**: FOMC, PCE, CPI, GDP, QE, QT, MBS, IOER, IORB, etc.

3 Model Training

3.1 Hyperparameter Grid Search

We trained thirteen Word2Vec model variants using Spark MLlib, exploring vector sizes (100, 200, 300), window sizes (5, 10, 15), and minimum counts (3, 5, 10). The top 7 configurations by analogy score are shown below:

Model	Vector Size	Window	Min Count	Vocab Size	Analogy Score
v100_w15_mc5	100	15	5	7,016	**0.5100**
v100_w10_mc3	100	10	3	9,009	0.4769
v100_w10_mc5	100	10	5	7,016	0.4768
v100_w5_mc5	100	5	5	7,016	0.4347
v200_w15_mc5	200	15	5	7,016	0.3955
v200_w10_mc5	200	10	5	7,016	0.3817
v300_w10_mc5	300	10	5	7,016	0.3347

Table 1: Word2Vec hyperparameter configurations ranked by analogy score. All models used 15 iterations and seed 42. Best model (bold) uses 100-dim vectors with window size 15.

Model Selection: We selected the best model based on "analogy score"—the average top-3 cosine similarity across three required analogies:

1. `quantitative_easing` − `asset_purchases` + `balance_sheet`
2. `tapering` − `asset_purchases` + `balance_sheet`
3. `tightening` − `hikes` + `cuts`

This criterion directly measures the model's ability to capture semantic relationships in monetary policy domain, rather than just vocabulary coverage. For each analogy, we report the average cosine similarity of the top-3 predicted terms (excluding query words), which reduces sensitivity to single outliers.

4 Results: Policy Analogies

All analogies were computed using the best model: `v100_w15_mc5` (100-dimensional vectors, window size 15, min count 5), which achieved the highest analogy score of 0.5100.

4.1 Analogy 1: QE Regime Shift

$$\texttt{quantitative\_easing} - \texttt{asset\_purchases} + \texttt{balance\_sheet} \approx \texttt{?} \tag{1}$$

Intuition: Quantitative easing involves asset purchases that expand the balance sheet. Removing the "purchasing" aspect while keeping the balance sheet focus should yield terms related to *holding* or *maintaining* assets.

Top 10 Results:

1. `arvind`: 0.5736
2. `krishnamurthy`: 0.5551
3. `wayne`: 0.4983
4. `rgensen`: 0.4894
5. `qe`: 0.4792 ✓
6. `passmore`: 0.4772
7. `vissing`: 0.4649
8. `gauti`: 0.4588
9. `diana`: 0.4533
10. `hancock`: 0.4496

Analysis: The analogy captured "qe" (abbreviation for quantitative easing) at rank 5 with improved similarity (0.4792). The top results include economist names (Krishnamurthy, Vissing-Jørgensen) who are prominent QE researchers. The larger window size (15) captures broader context, improving similarity scores significantly compared to smaller window models.

4.2 Analogy 2: Tapering Process

$$\texttt{tapering} - \texttt{asset\_purchases} + \texttt{balance\_sheet} \approx \texttt{?} \tag{2}$$

Intuition: Tapering is the gradual reduction of asset purchases. Removing asset purchases and adding balance sheet focus should yield terms about shrinking or normalizing the balance sheet.

Top 10 Results:

1. `runoff`: 0.6532 ✓
2. `phases`: 0.5987
3. `normalization`: 0.5810 ✓
4. `commence`: 0.5433
5. `proceed`: 0.5392
6. `normalizing`: 0.5276 ✓
7. `transition`: 0.5199
8. `gradual`: 0.5060
9. `soon`: 0.5031
10. `outlined`: 0.4826

Analysis: **Excellent result!** The model correctly identified "runoff" (0.6532) as the top result, with "normalization" (0.5810) at rank 3. These are precisely the terms used in Fed communications to describe balance sheet reduction after tapering. The improved window size produces even higher similarity scores (runoff: 0.6532 vs previous 0.5307). Additional relevant terms like "phases," "transition," and "gradual" reflect the Fed's measured approach to policy normalization.

4.3 Analogy 3: Policy Stance Reversal

$$\texttt{tightening} - \texttt{hikes} + \texttt{cuts} \approx \texttt{?} \tag{3}$$

Intuition: Tightening policy involves rate hikes. Replacing hikes with cuts should reverse the stance, yielding dovish/accommodative terms.

Top 10 Results:

1. `forceful`: 0.4342
2. `cumulative`: 0.3584
3. `prompt`: 0.3370
4. `dimensions`: 0.3224
5. `commenting`: 0.3103
6. `affects`: 0.3086
7. `differences`: 0.3068
8. `ecb`: 0.3038
9. `define`: 0.3034
10. `responses`: 0.3028

Analysis: This analogy remains challenging. The top result "forceful" (0.4342) may relate to aggressive policy action but is not directly "easing" or "dovish." The word "ecb" (European Central Bank) at rank 8 suggests cross-central-bank comparisons in easing contexts. This analogy is inherently difficult because "tightening" and "easing" are both policy stances that appear in similar contexts, making the vector arithmetic less discriminative.

5 Nearest Neighbor Analysis

To validate the embeddings, we examined nearest neighbors for key terms:

Macro Terms:

- `inflation`: Should neighbor "prices," "target," "2_percent"
- `unemployment`: Should neighbor "labor," "jobs," "employment"
- `labor_market`: Should neighbor "employment," "wages," "jobs"

Policy Tools:

- `balance_sheet`: Should neighbor "assets," "holdings," "securities"
- `asset_purchases`: Should neighbor "securities," "treasury," "mbs"

Stance Terms:

- `tightening`: Should neighbor "restrictive," "hikes," "higher"
- `easing`: Should neighbor "accommodative," "cuts," "stimulus"

6 Synonym and Antonym Analysis

6.1 Task 2: Synonym Pairs

We computed cosine similarity for Fed-relevant synonym pairs (different roots, similar meaning):

Best Result:

- **Words**: `employment` and `jobs`
- **Similarity Score**: 0.4894

Top 5 Synonym Pairs by Similarity:

Word A	Word B	Similarity
employment	jobs	0.4894
unemployment	jobless	0.4745
contraction	recession	0.4385
forward_guidance	communication	0.4307
inflation	prices	0.4112

Table 2: Best synonym pairs found by the model

Analysis: The model correctly identified labor market synonyms as having the highest similarity. "Employment/jobs" (0.4894) and "unemployment/jobless" (0.4745) both capture the dual mandate focus on labor markets. The "forward_guidance/communication" pair (0.4307) also scored highly, confirming the model understands Fed communication strategies. Overall synonym similarities improved with the larger window size.

6.2 Task 3: Antonym Pairs with High Similarity

Word2Vec often assigns high similarity to antonyms because they appear in similar contexts. We tested:

Best Result:

- **Words**: `up` and `down`
- **Similarity Score**: 0.5794

Top 5 Antonym Pairs by Similarity:

Word A	Word B	Similarity
up	down	0.5794
inflation	deflation	0.3959
increase	decrease	0.3897
slow	fast	0.3718
easing	tightening	0.3584

Table 3: Antonym pairs with highest similarity scores

Analysis: As expected, antonyms show positive similarity due to distributional semantics— "up" and "down" both describe directional movement and appear in similar contexts ("rates went up/down"). The Fed-specific pair "easing/tightening" scored 0.3584, confirming the model learned these opposing policy stances appear in similar contexts. Interestingly, the larger window size (15) produced *lower* antonym similarities compared to smaller windows, suggesting the model better distinguishes directional semantics with broader context.

7 Failure Cases and Analysis

7.1 Failure 1: Temporal Regime Mixing

Problem: The corpus spans multiple policy regimes—QE-era (2008-2014) vs. hiking cycle (2022+). Terms like "accommodation" meant different things when rates were at zero vs. 5%.

Evidence: Neighbors of "accommodation" include both QE-related terms (asset purchases) and rate-related terms, showing semantic confusion.

Why It Failed: Word2Vec learns one embedding per word; it cannot represent polysemy across time periods.

7.2 Failure 2: Doc-Type Ambiguity

Problem: FOMC statements use forward-looking, hedged language ("the Committee expects"), while minutes are retrospective with detailed discussion.

Evidence: "inflation" neighbors differ by document type—statements associate with "target" and "objective," while minutes associate with "participants" and "projection."

Why It Failed: The unified model averages over these doc-type differences.

7.3 Failure 3: Sparse Acronyms

Problem: Policy-critical acronyms like IORB, SRF, ON RRP appear rarely (<50 times) and only in recent documents.

What We Tried: Lowering min_count to 3; phrase detection to merge "on rrp" → "on_rrp"

Why It Partially Failed: ∼50 occurrences are insufficient for stable embeddings; Word2Vec needs hundreds of diverse contexts.

7.4 Failure 4: Training Data

One additional perspective worth highlighting is that FOMC communications can be viewed as an independent variable, while market reactions (rates, equities, volatility, expectations) act as dependent variables.

Without explicitly analyzing how markets respond to FOMC announcements, we are missing an important dimension of how these communications function in practice. Incorporating market reaction data could strengthen the interpretation of the embeddings by linking semantic structure to economic impact.

8 Conclusion

We successfully built a Word2Vec model on Federal Reserve text data that captures meaningful relationships between monetary policy terms. These results reflect distributional similarity rather than strict semantic equivalence, consistent with the Word2Vec training objective.

Key findings:

1. Fed-specific tokenization (phrase detection, acronym preservation) significantly improves embedding quality
2. **Model selection by analogy score** (vs. anchor coverage) yields better downstream performance—the best model `v100_w15_mc5` achieved 0.5100 average analogy score
3. Larger window size (15) captures broader policy context, improving analogy and synonym results while reducing spurious antonym similarity
4. The model learns expected relationships: employment$\sim$jobs (0.49), runoff$\sim$tapering (0.65), forward_guidance$\sim$communication (0.43)
5. Temporal mixing and sparse terms remain challenging for static Word2Vec embeddings