CIS530 HW2: Jonathon Delemos

2.1 All-complex (4 points)

Results

- Train Set:
  - Precision: .418
  - Recall: 1.00
  - F-score: .589
- Dev Set:
  - Precision: .43275
  - Recall: 1.00
  - F-score: .604

2.2 Word-length baseline (6 points)

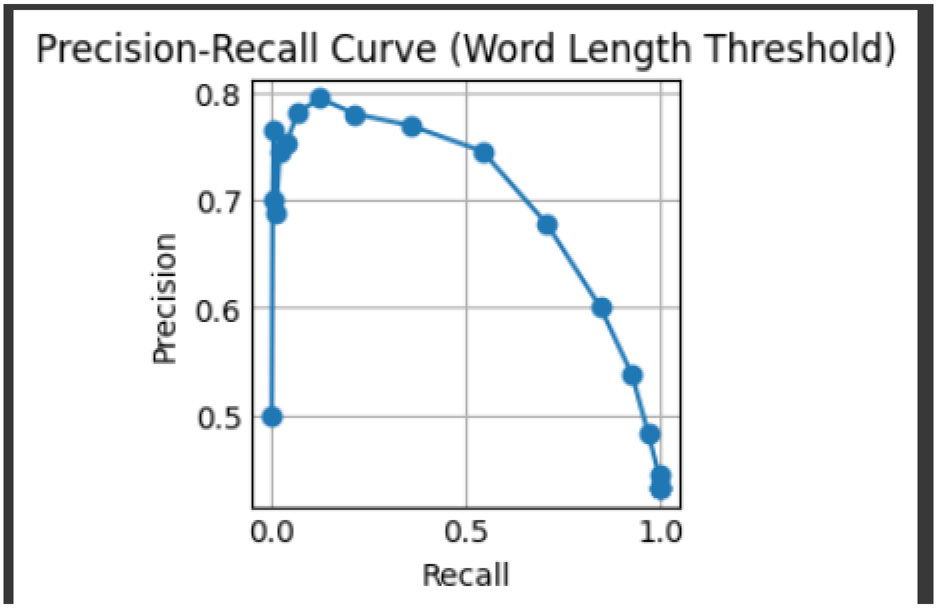
Results

- Train Set:
  - Precision: .605
  - Recall: .866
  - F-score: .712
- Dev Set:
  - Precision: .600
  - Recall: .844
  - F-score: .701

Threshold Analysis

- Range of thresholds tested: 1-20
- Best threshold: 7

Precision-Recall Curve



P-R Curve Analysis

The curve is skewed towards the left. This indicates that while minimizing recall we see the highest levels of precision. Inversely, if we maximize recall, we see diminishing returns in precision. This aligns with the expected relationship between precision and recall.

2.3 Word-frequency baseline (6 points)

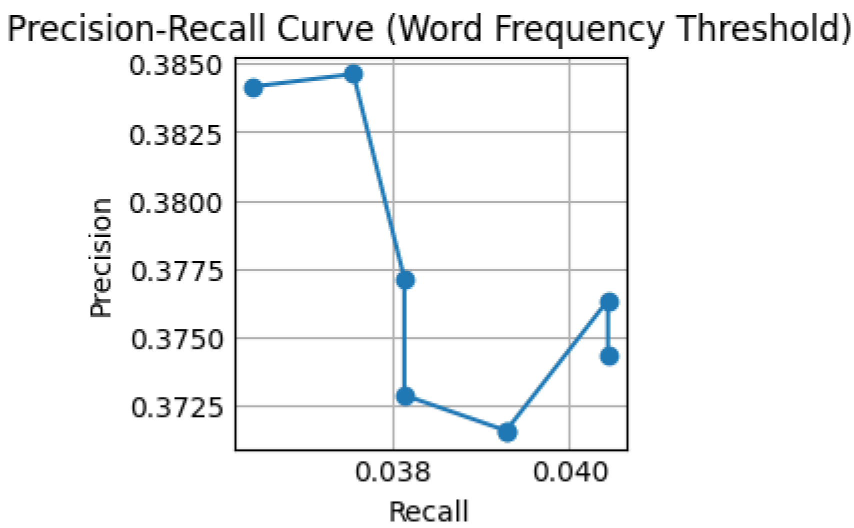
Results

- Train Set:
  - Precision: .326
  - Recall: .038
  - F-score: .068
- Dev Set:
  - Precision: .378
  - Recall: .040
  - F-score: .073

Threshold Analysis

- Range of thresholds tested: 0-3999
- Best threshold: 15

Precision-Recall Curve



P-R Curve Analysis

All words have a frequency higher than zero. The data somewhat resembles all_complex which makes sense. The f_scores are low which is to be expected. This demonstrates that most words have a high frequency, and that doesn't accurately classify them as complex or not.

Discussion

The data would suggest that word length would be a more accurate measure of the word complexity. Frequency feature measurement would need to be modified for it to be effective.

3.1 Naive Bayes (4 points)

Results

- Train Set:
  - Precision: .469
  - Recall: .968
  - F-score: .632
- Dev Set:
  - Precision: .495
  - Recall: .979
  - F-score: .657

3.2 Logistic Regression (4 points)

Results

- Train Set:
  - Precision: .726
  - Recall: .693
  - F-score: .709
- Dev Set:
  - Precision: .725
  - Recall: .658
  - F-score: .689

3.3 Discussion on NB and LR (6 points)

Model Comparison

The logistic regression model outperformed the Naive Bayesian model in this testing exercise. The F Scores were better for both the training and dev set in the logistic regression model.

Performance Analysis

The difference in performance could be explained by the mathematical model - the logistic regression model is most likely better suited for this specific task. It’s a discriminative objective, so perhaps this model outperformed due to the nature of the exercise.

4.1 Own model's performance (6 points)

Dev Set Performance

- Precision: .4913
- Recall: .952
- F1-score: .648

4.2 Own model's analysis (15 points)

Model Description

The model I built finds the average number of synonyms. It determines words to be complex if the number of synonyms are less than the average number in the wordset. I estimated the less synonyms a word had, the more unique and therefore complex the word might be.

Features Description

My model implemented a synonym feature that determined complexity based on the number of synonyms. If a word had less than the average number of synonyms, it was determined to be complex. Length didn't factor in.

Feature Selection Justification

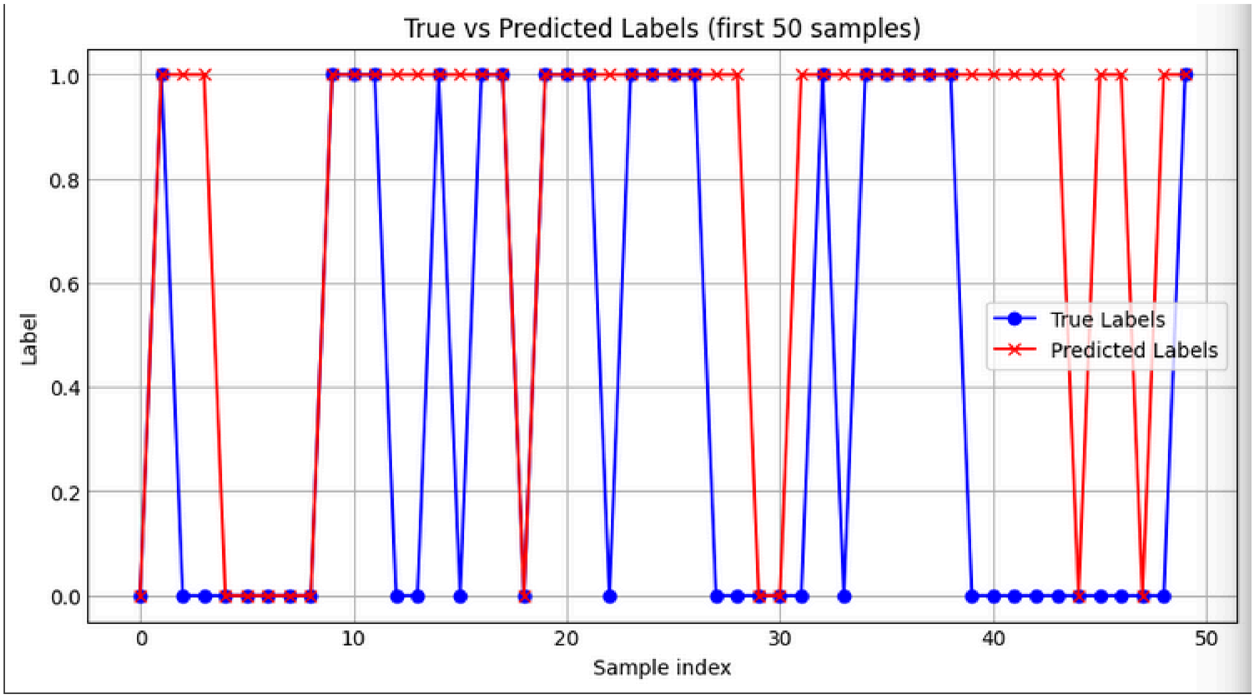
My argument: short obscure words can be complex. Examples of models included the length and frequency of the words. My thought was the number of relevant synonyms would determine the complexity of the word in the word set. Although it didn't produce the greatest f score, I do believe that the model was clever in its application. It could operate in an alternative way, strictly based on the number of adjacent synonym nodes.

Performance Examples

- Examples of words the model classified correctly:
  - True Complex: renewable, 'coup', 'combed', 'plummeted',
  - True Simple: hammer, 'continues','exceed','conventional', 'surprisingly',
- Examples of words the model classified incorrectly:
  - False Complex: ‘sure’, go-ahead, 'retailers', 'sled'
  - False Simple: 'mollusks'

Error Analysis

- Category 1: Conjugated Words
  - Examples: Showings
  - Potential reason: Conjugated words don't inherit their lemma's synonyms.
- Category 2: Simple objects that don't have many synonyms
  - Examples: Sled
  - Potential reason: Some objects don't have very many synonyms. What's another
  - word for sled, taboggan? It's a well known word with limited synonyms.


True vs Predicted Labels (first 50 samples)
Label
Sample index
True Labels
Predicted Labels
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50