Sentiment Classification Methods Using Logistic Regression

1. Introduction

Text classification is a fundamental task in natural language processing with applications ranging from sentiment analysis to document categorization. This study focuses on evaluating and comparing two popular feature extraction techniques—Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF)—within the context of logistic regression classifiers. The primary objective is to determine which feature representation method yields superior performance and under what conditions.

2. Methodology

2.1 Feature Extraction Methods

Two text representation methods were compared in this study:

- **Bag of Words (BoW)**: A simple representation that counts word occurrences in documents, disregarding grammar and word order but retaining multiplicity.

- **Term Frequency-Inverse Document Frequency (TF-IDF)**: A numerical statistic reflecting the importance of a word in a document relative to a corpus, combining term frequency with inverse document frequency to reduce the impact of common words.

2.2 Classification Model

The study employed logistic regression as the classification algorithm with various hyperparameter configurations:

- Regularization strength (C) values ranging from 10^-3 to 10^2
- Class weighting: None and balanced
- L1 penalty (Lasso regularization)
- Solver: liblinear (suitable for smaller datasets)

2.3 Evaluation Metrics

Performance was evaluated using multiple metrics:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion matrices

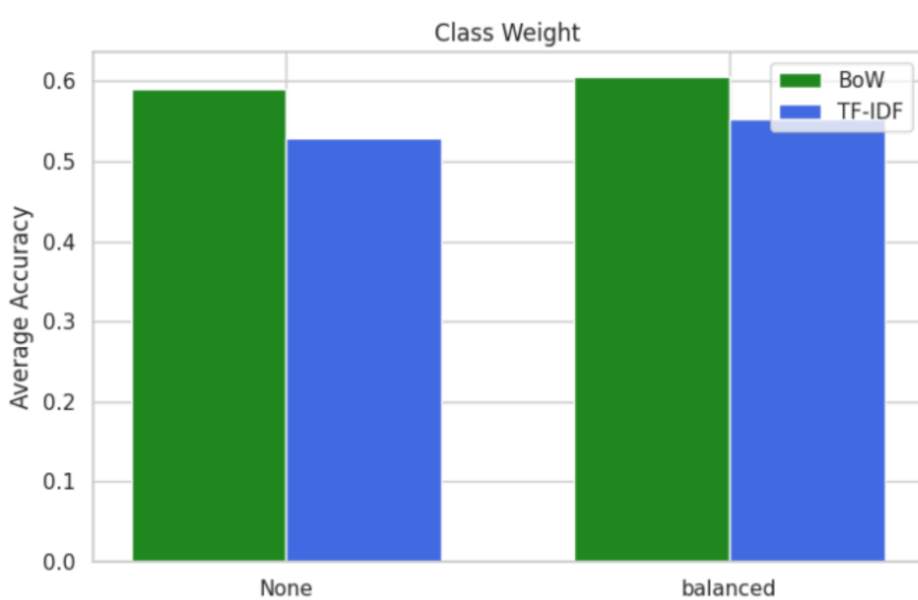
3. Results and Analysis

3.1 Overall Performance Comparison

BoW consistently outperformed TF-IDF across most experimental conditions. The overall average accuracy was approximately 0.59 for BoW compared to 0.54 for TF-IDF.

3.2 Effect of Class Weighting

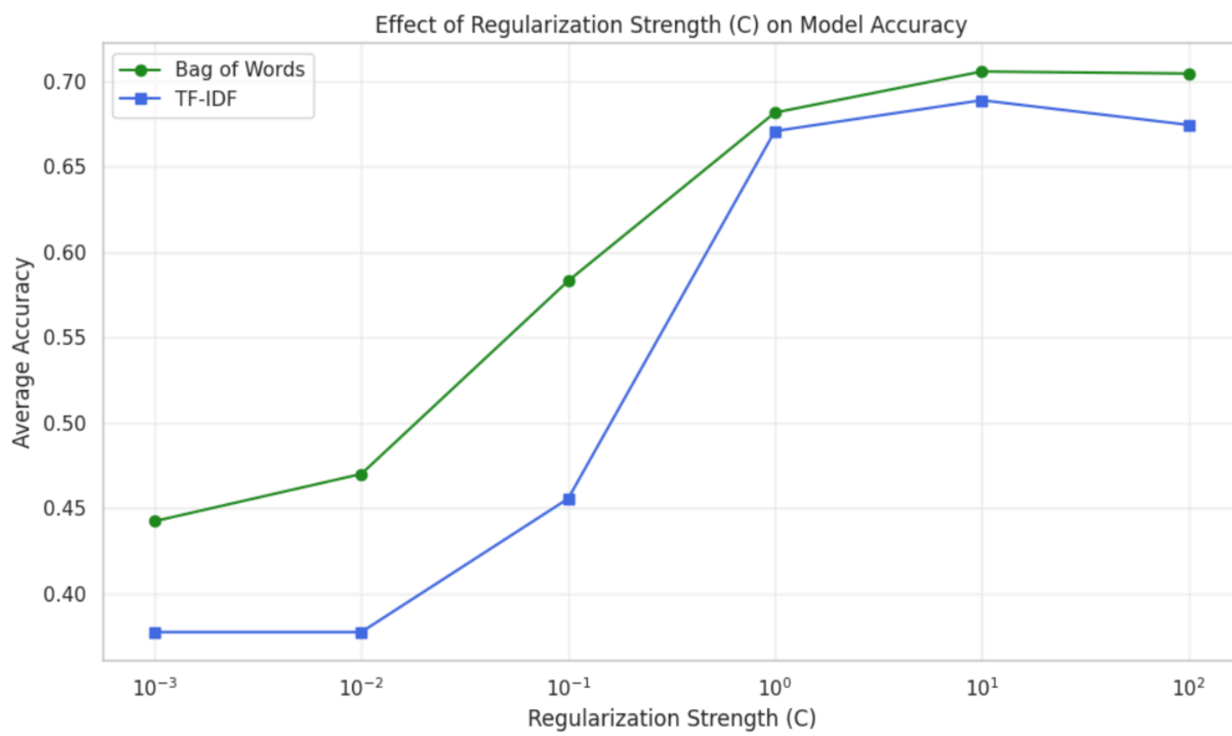
Class weighting had a minimal positive impact on both feature extraction methods. With balanced class weights, BoW achieved an accuracy of approximately 0.60, compared to 0.59 without class weights. TF-IDF showed a similar marginal improvement from 0.52 to 0.54.



3.3 Effect of Regularization Strength

Regularization strength (C parameter) had a significant impact on model performance. Both methods showed improved performance as C increased from 10^-3 to 10^1, with BoW consistently outperforming TF-IDF across the entire range:

- At low regularization strengths (C=10^-3), both methods performed poorly, with BoW at 0.44 and TF-IDF at 0.38.
- Performance improved rapidly between C=10^-1 and C=10^0 for both methods.
- Optimal performance was achieved at C=10^1 for both methods (BoW: 0.71, TF-IDF: 0.69).
- At very high regularization (C=10^2), performance plateaued for BoW but slightly decreased for TF-IDF.

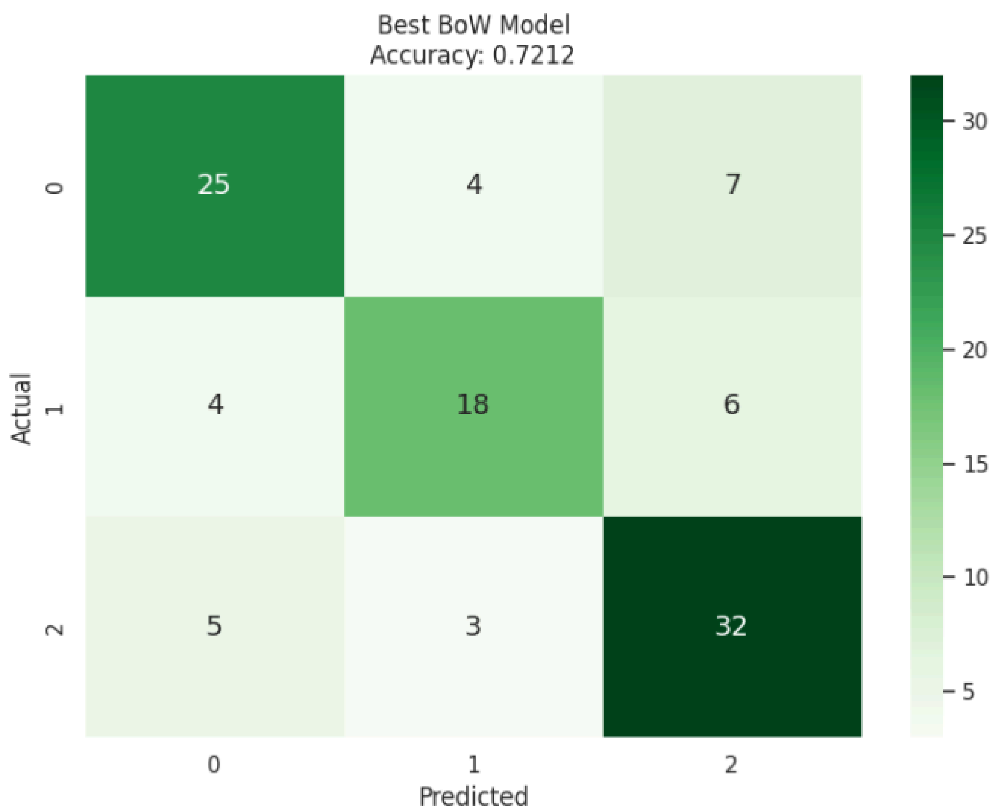


3.4 Confusion Matrix Analysis

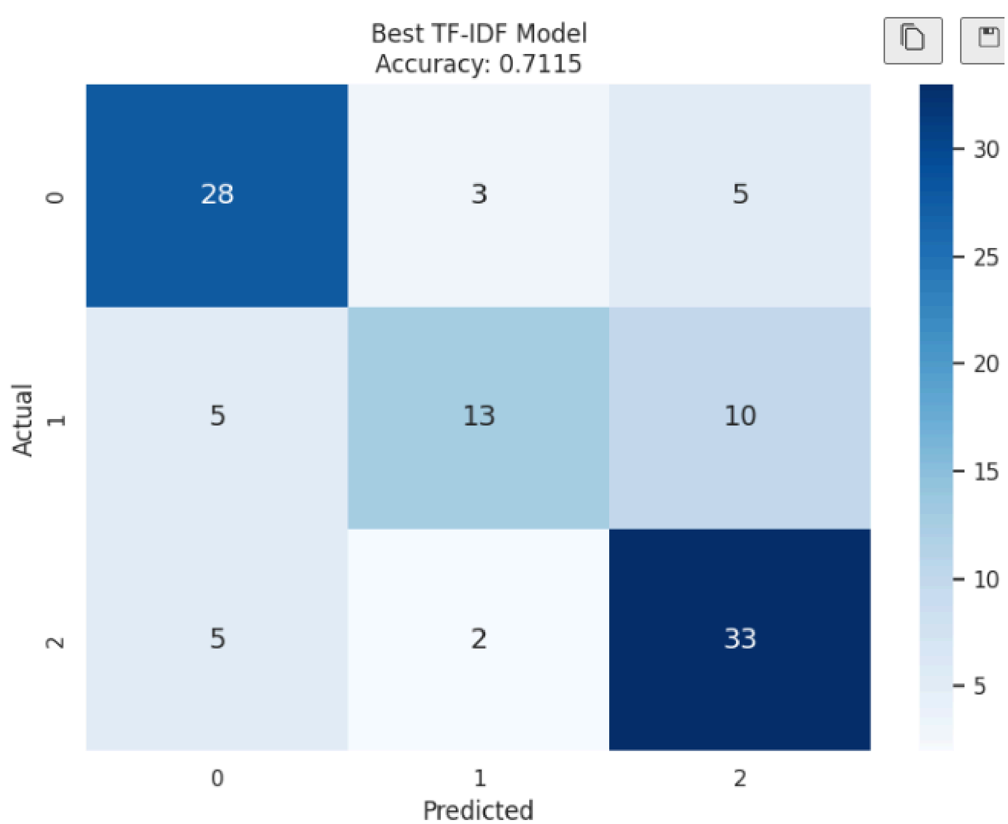
The confusion matrices for the best-performing models reveal interesting classification patterns:

Best BoW Model (Accuracy: 0.7212):

- Strong diagonal elements indicate good class separation
- Class 0: 25 correct predictions out of 36 samples (69.4% correct)
- Class 1: 18 correct predictions out of 28 samples (64.3% correct)



- Class 2: 32 correct predictions out of 40 samples (80% correct)



Best TF-IDF Model (Accuracy: 0.7115):

- Similar pattern to BoW but with slight differences
- Class 0: 28 correct predictions out of 36 samples (77.8% correct)
- Class 1: 13 correct predictions out of 28 samples (46.4% correct)
- Class 2: 33 correct predictions out of 40 samples (82.5% correct)

3.5 Detailed Performance Metrics

The classification reports provide deeper insights into model performance beyond accuracy:

TF-IDF Model (Accuracy: 0.654):

- Precision ranges from 0.62 to 0.71
- Recall varies significantly across classes (0.29 to 0.81)
- F1-scores range from 0.39 to 0.75
- Class 1 shows particularly poor recall (0.29)

BoW Model (Accuracy: 0.701):

- More consistent performance across metrics
- Precision ranges from 0.66 to 0.74
- Recall is more balanced (0.64 to 0.78)
- F1-scores range from 0.67 to 0.74
- Better macro and weighted averages (0.70) than TF-IDF (0.65)

3.6 Optimal Model Configurations

The best performing models were achieved with the following configurations:

BoW Model:

- C=10.0
- penalty=l1
- solver=liblinear
- class_weight=None
- Accuracy: 0.7212

4. Conclusions and Recommendations

Based on the comprehensive analysis presented in this report, several conclusions can be drawn:

1. **Feature Representation**: Bag of Words consistently outperforms TF-IDF for this particular text classification task, achieving higher accuracy and more balanced performance across classes.

2. **Optimal Configuration**: The best performance is achieved with moderate regularization strength (C=10.0 for BoW, C=1.0 for TF-IDF) using L1 regularization and the liblinear solver.

3. **Class Weighting**: Class balancing appears to have minimal impact on overall performance, suggesting that the native class distribution may be appropriate for this classification task.

4. **Practical Implications**: For applications where overall accuracy and balanced class performance are priorities, BoW with moderate regularization represents the optimal approach among those tested.