

An Ensemble Model for Automatic Grading of Evidence

Yuting Guo¹, Yao Ge¹, Ruqi Liao², Abeed Sarker¹

¹Department of Biomedical Informatics, Emory University

²Department of Industrial and System Engineering, Georgia Institute of Technology



EMORY
UNIVERSITY
SCHOOL OF
MEDICINE

Introduction and Methodology

Approaches:

- Modified RoBERTa model adapted to multi-document classification.
- SVMs with feature engineering.
- Ensemble model: converts the discrete grading problem into a continuous score generation problem.

RoBERTa

Data reorganized as shown in Figure 1.

Inference:

- class probability vectors of multiple publications averaged, and
- class with the highest probability chosen as the final prediction.

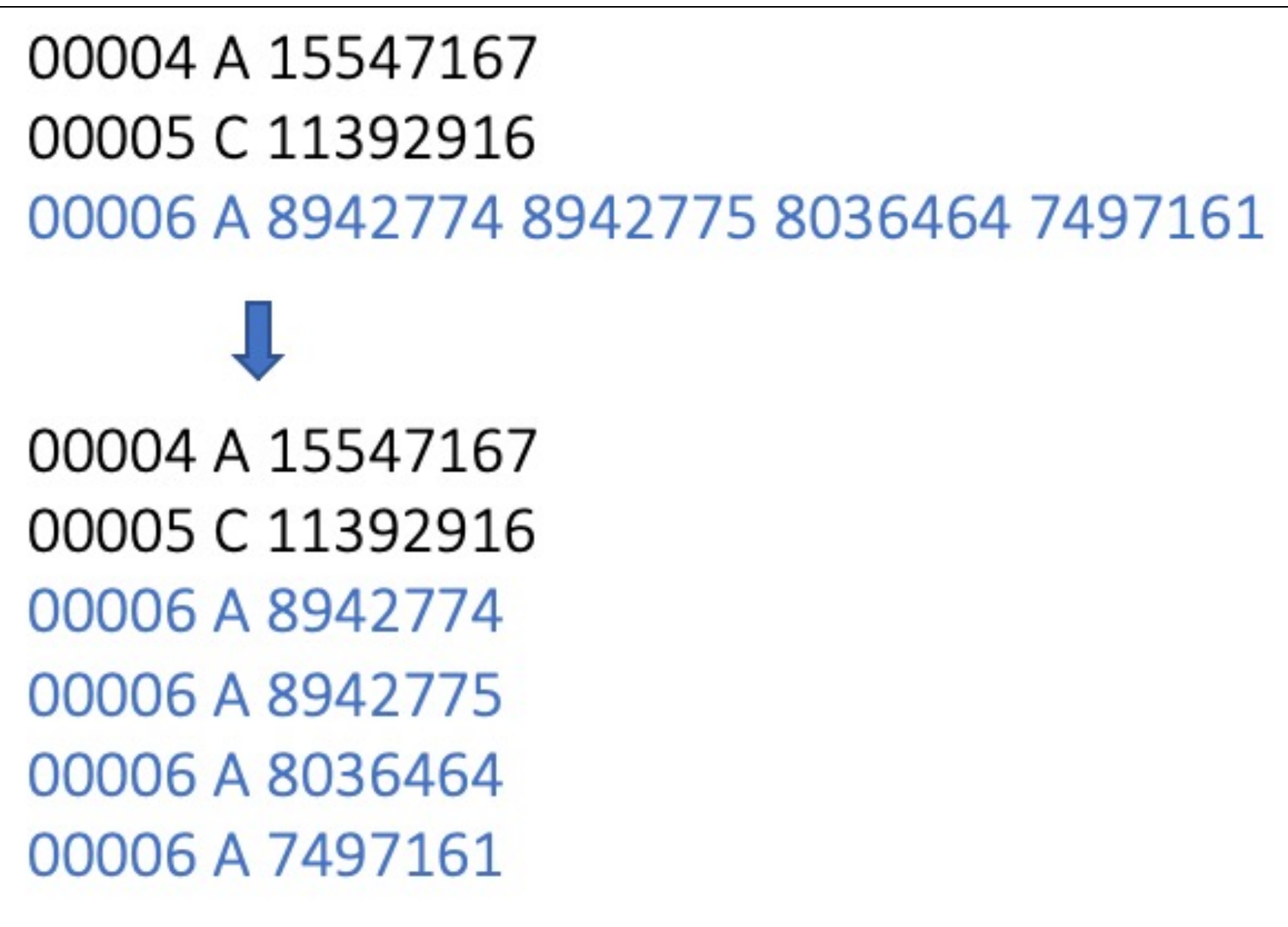


Figure 1:
Data example.

Feature Engineering

- N-gram Features from titles and abstract bodies (N=1, 2, 3, 4).
- Consistency of experimental results mentioned in abstracts.
- Publication types (e.g., randomized clinical trial, systematic review).
- Medical Subject Headings (MeSH) terms.

Ensemble

Converted the task of grading as a score generation problem:

- Classes A, B, and C considered to be the numbers 0, 1, and 2.
- Scores from multiple classifiers added.
- Grid search to determine optimal boundaries for grades.
- Scores smaller the lower threshold were classified as A; those larger than the higher threshold were classified as C.

Results

Model	Dev	95% CI	Test	95% CI
SVM	0.63	0.48-0.76	0.48	0.34-0.60
RoBERTa	0.58	0.44-0.70	0.48	0.34-0.62
Ensemble	0.7	0.58-0.82	0.54	0.38-0.68

Table 1: Accuracies and 95% confidence intervals(CIs) on the development and test set.

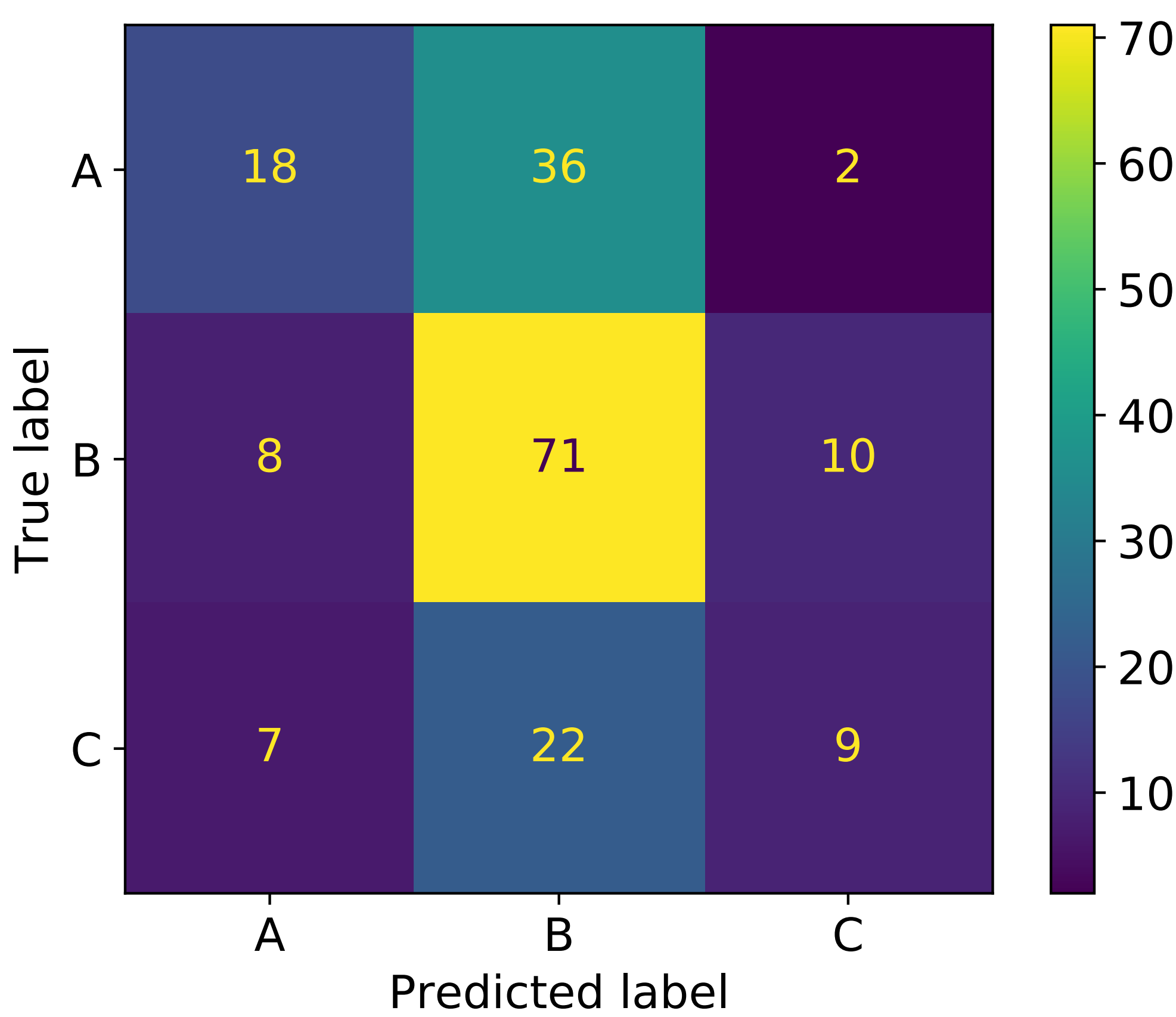


Figure 2: The confusion matrix for the result of the ensemble model on the test set.

Findings

- The ensemble model performed best both on the development set and the test set.
- All the proposed models might be overfitting (low performance on test set).