BMI 550: Assignment 2
JaMor Hairston, MSHI, MS
Emory University School of Medicine
Department of Biomedical Informatics
October 24, 2023

Abstract

This study uses Natural Language Processing (NLP) techniques to predict Freezing of Gait (FOG) instances from patient fall reports. The primary goal of employing machine learning algorithms and text analysis is to develop a robust model capable of accurately identifying and predicting occurrences of FOG based on narrative descriptions of falls. This predictive framework holds the potential for facilitating timely interventions and personalized healthcare strategies for individuals experiencing FOG symptoms. This use case found forecasting the need for early intervention and personalized medicine plausible.

Introduction

Freezing of Gait (FOG) presents a significant challenge for individuals with neurological disorders such as Parkinson's disease, often leading to impaired mobility and increased risk of falls. Accurate prediction of FOG instances from patient-reported fall descriptions is crucial in proactive healthcare interventions and in improving patient outcomes. This study uses advanced NLP techniques and machine learning models to predict the likelihood of FOG occurrences based on textual narratives of falls. By harnessing predictive analytics in this context, the aim is to enable early detection and personalized management strategies for individuals susceptible to FOG episodes.

Adjusting the objective to predict FOG from fall reports emphasizes the predictive aspect, aiming to forecast the likelihood of FOG incidents from the textual descriptions provided by patients who have experienced falls. This alteration aligns the study's focus more explicitly with the predictive modeling aspect to anticipate and identify FOG occurrences based on textual data, fostering proactive healthcare interventions and support strategies.

Methods

Data Preprocessing The train and test datasets containing textual descriptions of falls were preprocessed to remove noise, standardize text, handle missing values, and perform tokenization for further analysis. The training dataset included 300 unique records with 29 variables – including IDs, demographics, and descriptive text. The test data included 72 unique records and the same 29 variables of the training set. Textual data, 'fall_description', was cleaned, normalized, and transformed into a format suitable for feature extraction. **Feature Engineering** To conduct the instructed feat engineering step, two primary feature engineering techniques were employed: 1. TF-IDF for N-Grams: Text data was vectorized using TF-IDF (Term Frequency-Inverse Document Frequency) representation with n-gram ranges to capture sequential word combinations. 2. Clustering-Based Features: Utilizing K-means clustering, text features were grouped into clusters to derive additional categorical features. **Model Building** Multiple machine learning classifiers: Logistic Regression, Random Forest, Gradient Boosting, SVM, and

Naive Bayes were trained on the aforementioned engineered features. Hyperparameter tuning and cross-validation were conducted to optimize model performance. **Classifier Evaluation** The evaluation metrics, micro F1 score, accuracy, precision, recall, and AUC were employed to assess the models' performance on the test dataset. Per the assignment instructions, the overall micro-averaged F1 score was used to evaluate the performance of the classifiers. **Training Set Size vs. Performance Exploration:** An analysis was conducted to explore how the variation in training set sizes affects the classifiers' performance. Models were trained and evaluated on subsets of varying sizes to understand the impact of data volume on predictive accuracy. **Ablation Study** Finally, a systematic ablation study was conducted to gauge the importance of individual features. Each feature was removed individually, and the resulting model performance was assessed using the micro F1 score.

Results

In comparing the performance of the classifiers, illustrated in Table 1, the Logistic Regression classifier performed the highest in accurately classifying FOG instances from textual descriptions, based on the micro F1 score – with a micro-averaged F1 score of 0.8169.

	Accuracy	F1_Micro	F1_Macro	Precision	Recall	AUC
Log Reg	0.8169	0.8169	0.8156	0.8230	0.8169	0.8159
Random Forest	0.7324	0.7324	0.7245	0.7560	0.7324	0.7302
XGB	0.7606	0.7606	0.7598	0.7622	0.7606	0.7599
SVM	0.5070	0.5070	0.3604	0.7535	0.5070	0.5139
Naïve Bayes	0.6479	0.6479	0.6025	0.7527	0.6479	0.6433

Table 1: Classifier Results Comparison

Using this classifier, the training set size v. performance analysis (Figure 1) revealed an increasing trend in model performance with larger training set sizes. However, performance gains began to plateau beyond a certain volume, suggesting diminishing returns beyond a specific dataset size.
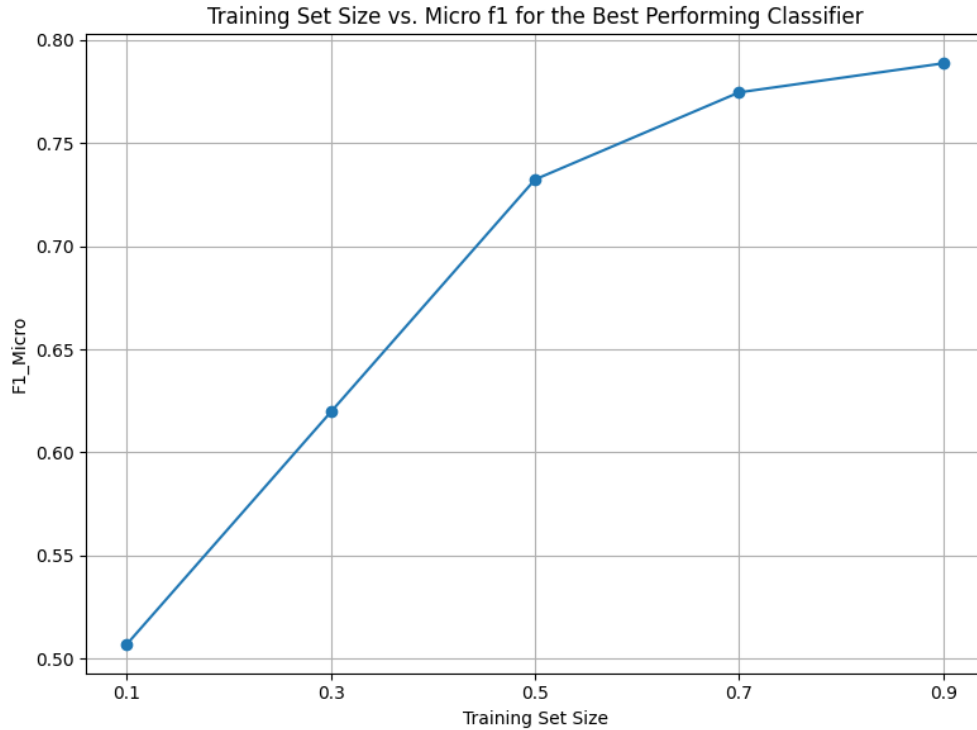
Figure 1: Training Set vs. Performance Graph

The ablation study demonstrated varying impacts of individual features on the model's micro F1 score. Certain clusters derived from text data exhibited a more significant influence on the classifier's performance when removed.

Discussion and Conclusion

When considering the study's primary focus on predictive modeling aimed to forecast Freezing of Gait (FOG) instances from textual fall reports, it has produced valuable insights. The predictive models, particularly the Logistic Regression classifier, showcased notable capabilities in foreseeing the likelihood of FOG occurrences based on narrative descriptions of falls. The emphasis on predictive analytics underscores the potential for interventions and personalized care strategies for individuals susceptible to FOG episodes. Examining the importance of features highlighted the significance of specific textual clusters derived from fall reports in driving the predictive accuracy of FOG identification. Understanding the relevance of these features could clarify the essential information embedded within textual narratives that aid in effectively predicting FOG incidents. This feature-focused analysis provides critical guidance for feature selection and further model refinement for enhanced predictive capabilities.

Predictive modeling for Freezing of Gait (FOG) identification from fall reports demonstrates promising potential in healthcare interventions and personalized care. The successful application of Natural Language Processing (NLP) techniques signifies a notable stride toward accurately predicting FOG occurrences based on textual descriptions of falls.

Moving forward, continued advancements in feature engineering, model optimization, and leveraging larger datasets can augment the predictive capabilities of the FOG identification system. These improvements would further empower healthcare professionals with early insights into potential FOG episodes, fostering proactive interventions and tailored care plans for individuals at risk of FOG-related mobility challenges.

References

-   Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python (NLTK). "O'Reilly Media, Inc."
-   OpenAI. (2023). ChatGPT (Nov 1st version) [Large language model]. https://chat.openai.com/chat - Used for code debugging.

Code

The code produced for the rule-based system and analysis can be found on GitHub at the following web address https://github.com/jmhairston/BMI550Assignment2.