Informe tarea 03 - Análisis de Sentimientos

Juan Esteban Rodríguez^{1,†}, Camilo Barreto^{2,†}, Luis Alejandro Medina^{3,†} and Julian Camilo Garcia^{4,†}

Abstract

En esta tarea desarrollamos y evaluamos clasificadores de sentimientos utilizando Naive Bayes (NB) y Regresión Logística (LR) en diferentes categorías del Multi-Domain Sentiment Dataset: "Books", "DVD", "Electronics" y "Kitchen". Los clasificadores fueron entrenados con representaciones de texto basadas en TF-IDF y características extraídas de lexicones. Se compararon los resultados de NB y LR en términos de precisión, recall, F1-score y accuracy.

1. Naive Bayes(NB), Logistic Regression(LR)

n este punto del proyecto, el objetivo principal es implementar y evaluar un modelo de clasificación utilizando regresión logística con ajuste de parámetros.

1.1. Metodología

10

12

13

17

18

19

20

21

22

25

27

32

33

34

35

- El proceso se llevó a cabo de la siguiente manera:
 - División del conjunto de datos: Se dividieron los datos en conjuntos de entrenamiento, validación y prueba. Esta estrategia garantiza que el modelo generalice bien a datos no vistos.
 - Vectorización de características: Para representar los datos textuales, se utilizó la transformación TF-IDF (Term Frequency-Inverse Document Frequency), convirtiendo cada documento en un vector de características numéricas.
 - Modelo de regresión logística con SGD: Se empleó el algoritmo Stochastic Gradient Descent para entrenar un modelo de regresión logística. La función de pérdida utilizada fue la log-loss, que es adecuada para problemas de clasificación. Además, se ajustaron hiperparámetros como la tasa de aprendizaje (eta0) y la regularización (12 penalty).
 - Criterio de parada: Para evitar sobreajuste, se implementó un criterio de parada temprana basado en el número de épocas sin mejoras significativas en la precisión del modelo.

1.2. Resultados

	naive bayes tf	naive bayes tfidf	SGD classifier tf	SGD clasifier tfidf
presicion macro	0.884278898	0.853136733	0.806600263	0.841987894
recall macro	0.796340301	0.84525469	0.797392215	0.83320876
f1 score macro	0.781653351	0.845527429	0.799144498	0.83381099
	naive bayes tf	naive baves tfidf	SGD classifier tf	SGD clasifier tfidf
presicion micro	0.797663303	0.853425385	0.800672685	0.84103381
recall micro	0.797663303	0.853425385	0.800672685	0.84103381
f1 score micro	0.797663303	0.853425385	0.800672685	0.84103381

Figure 1. Enter Caption

1.3. Conclusiones

- Mejores Resultados: La representación TF-IDF es claramente superior a la representación TF para ambos modelos. Esto se refleja en una mejora consistente de las métricas de rendimiento.
- Modelos: SGDClassifier ofrece un rendimiento ligeramente superior al de Naive Bayes, especialmente con TF-IDF, lo que sugiere que es más adecuado para esta tarea de clasificación de texto.
- Recomendación: Para este problema en particular, el uso de SGDClassifier con TF-IDF es la mejor combinación, ya que ofrece el mejor equilibrio entre precisión, recall y F1-score, evitando el sobreajuste gracias al uso del criterio de parada temprana implementado en el entrenamiento por épocas.

2. Sentiment Analysis Un clasificador por cada categoría

En esta parte de la tarea, el objetivo es implementar modelos de clasificación de sentimientos utilizando dos algoritmos principales: Naive Bayes y Regresión Logística. Los datos provienen del Multi-Domain Sentiment Dataset, que incluye reseñas en cuatro categorías: Books, DVD, Electronics y Kitchen. Se busca construir clasificadores para predecir si una reseña es positiva o negativa en cada categoría, utilizando dos representaciones de características: frecuencia de términos (TF) y frecuencia inversa de documentos (TF-IDF).

2.1. Metodología

- División de los datos: Se dividieron los datos en conjuntos de entrenamiento, validación y prueba. El conjunto de entrenamiento y validación se utilizó para ajustar los hiperparámetros de los modelos y entrenarlos, mientras que el conjunto de prueba se usó para evaluar su rendimiento final.
- · Modelos utilizados:
- Naive Bayes (NB): Se utilizó el modelo de Naive Bayes Multinomial, que es comúnmente empleado para clasificación de texto. Este modelo se entrenó con las representaciones TF y TF-IDF de los textos. Regresión Logística (LR): Se implementó un modelo de Regresión Logística para clasificación binaria, utilizando también las representaciones TF y TF-IDF.
- Métricas de evaluación: Para evaluar el rendimiento de los modelos, se utilizaron las métricas precisión, recall y F1-score en sus versiones macro y micro. Estas métricas permiten medir la capacidad de los modelos para clasificar correctamente las reseñas en positivas o negativas.

2.2. Resultados

- Comparación General entre Naive Bayes y Regresión Logística: La Regresión Logística tiende a superar a Naive Bayes en todas las métricas evaluadas (precisión, recall, F1-score y exactitud). Esto se debe a que LR puede capturar relaciones más complejas entre las características, mientras que NB asume independencia entre las palabras, lo cual es una limitación en el análisis de lenguaje natural.
- El rendimiento de LR fue particularmente superior en categorías como Kitchen y Books, donde el vocabulario es más diverso y complejo. Por ejemplo, en Kitchen, LR alcanzó una precisión de 0.7340 en comparación con 0.6923 de NB. Dificultad de las Categorías:
- Books fue la categoría más difícil de clasificar para ambos modelos, con un F1-score más bajo en comparación con las demás categorías. Esto sugiere que las reseñas en esta categoría tienen un lenguaje más subjetivo o con mayor variabilidad en las palabras utilizadas, lo que dificulta la predicción del sentimiento. Por otro lado, Kitchen resultó ser la categoría más fácil de clasificar, con los mejores valores de F1-score y exactitud en ambos modelos, especialmente en LR, que alcanzó un F1-score de 0.7376.

39

40

41

42

43

47

48

49

50

51

53

54

57

60

63

64

65

70

71

72

73

74

75

76

77

[†]Estos autores contribuveron igualmente a este trabajo

Diferencias en Recall:

85

86

87

88

89

93

102

103

110

112

115

116

117

119

120

121

135

137

138

139

140

141

142

 Naive Bayes mostró un mejor recall en algunas categorías, como Electronics y Kitchen, lo que indica que el modelo fue capaz de identificar más ejemplos positivos correctamente, aunque con una tasa de falsos positivos más alta. Este comportamiento sugiere que NB es más conservador en términos de predicciones negativas, lo que puede ser útil en aplicaciones donde es crucial capturar la mayor cantidad de instancias positivas.

3. Sentiment Analysis: Clasificador Único para Múltiples Categorías

Mientras que la primera parte del proyecto involucró la construcción de clasificadores individuales para cada categoría, la segunda parte, que abordamos en este análisis, se enfoca en desarrollar un único clasificador que pueda generalizar a través de todas las categorías.
Esto implica combinar los datos de todas las categorías y evaluar si un modelo consolidado puede igualar o superar el rendimiento de los clasificadores específicos de categoría.

3.1. Metodología

- Se recopilaron y combinaron todas las reseñas etiquetadas (positivas y negativas) de las cuatro categorías mencionadas. Esto resultó en un único conjunto de datos
- Se aplicó un preprocesamiento consistente al texto para normalizar y limpiar los datos
- El conjunto de datos combinado se dividió en conjuntos de entrenamiento y prueba, utilizando una proporción del 80 porciento para entrenamiento y 20 porciento para prueba.
- Se generaron dos tipos de representaciones TF y TF-IDF
- Se extrajeron características basadas únicamente en lexicones de sentimientos:

AFINN Lexicon: Proporciona puntajes de sentimiento para palabras individuales. SentiWordNet Lexicon: Asigna puntajes positivos y negativos a palabras basadas en su significado. WordStat Lexicon: Clasifica palabras en positivas, negativas y neutrales.

3.2. Resultados

3.2.1. Rendimiento de los Modelos con Características de Lexicones

Naive Bayes (Características de Lexicones):

Naive Bayes (Características de Lexicones):

Precisión: 0.5364
Recall: 0.9181
F1 Score: 0.6772
Exactitud: 0.5656

Regresión Logística (Características de Lexicones):

Precisión: 0.7194
Recall: 0.6877
F1 Score: 0.7032
Exactitud: 0.7119

La regresión logística superó al modelo de Naive Bayes en todas las métricas. Aunque Naive Bayes mostró un alto *recall*, su precisión y exactitud fueron significativamente menores, lo que indica un alto número de falsos positivos.

3.3. Analisis finales

- Las palabras con mayor peso positivo son adjetivos que típicamente se asocian con opiniones favorables, como great, excellent, best, easy, love.
- Las palabras con mayor peso negativo incluyen not, waste, worst, bad, poor, que son indicadores claros de reseñas negativas.
- La palabra not tiene un peso negativo significativo, lo que resalta su importancia en la inversión del significado de oraciones.