HauBert: A transformer Model for Aspect-based Sentiment Analysis of Hausa-Language movie reviews

Aminu Musa[1], Fatima Mohd Adam[1], Umar Ibrahim[1], Abubakar Yakubu Zandam[2]
[1]Federal Universiy Dutse, Nigeria musa.aminu@fud.edu.ng
[2]Federal University Technology Babura University, Nigeria

ASEC 2024 Conference

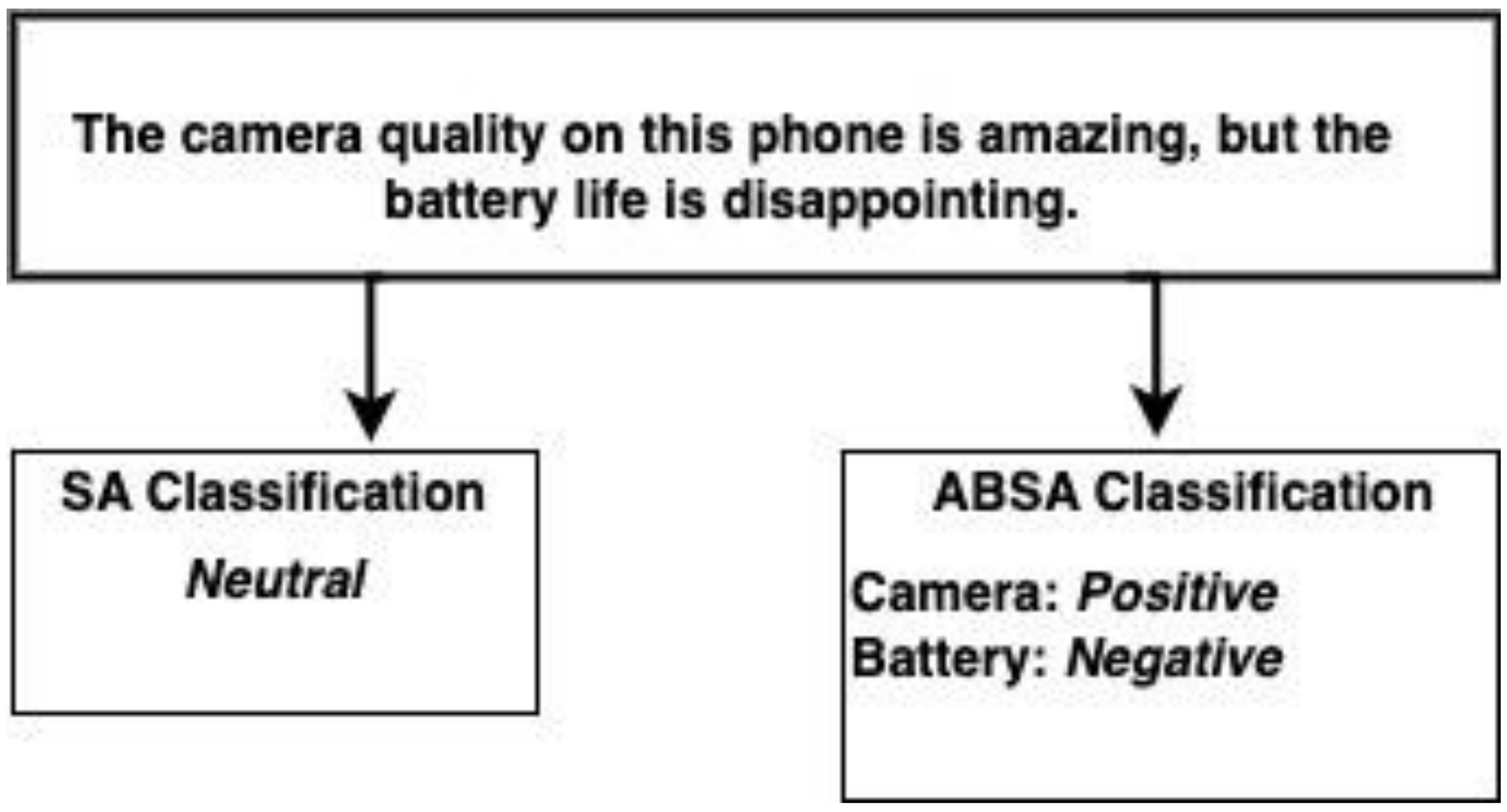
Problem

- SA and ABSA has advanced significantly for widely spoken languages such as English, The major drawback is that the vast majority of works on SA/ABSA were carried on high-resource languages such as English Language. This creates a gap when it comes to low-resource languages, especially African languages like Hausa.

Contribution

- Collection, and organizing the first open source Hausa Dataset for ABSA
- We proposed a transformer- based model for ABSA with improve performance compared with traditional CNN model.

Method



Training process

Algorithm 1 Fine-Tuning mBERT for ABSA in Hausa Language

Require: Preprocessed dataset $\mathcal{D}$ with inputs in Hausa; pre-trained mBERT model $\mathcal{M}$; optimizer $\mathcal{O}$; number of epochs E; batch size B; and loss function $\mathcal{L}$.

Initialize mBERT model $\mathcal{M}$ with pre-trained weights.
Add a task-specific classification head to $\mathcal{M}$ for ABSA.
Split dataset $\mathcal{D}$ into training and validation sets.
Divide the training set into batches of size B.
for epoch $e = 1$ to E **do**
  Shuffle the training dataset.
  for each batch $(\mathbf{X}, \mathbf{y})$ in the training set **do**
    Unpack the batch:
      Input IDs $\mathbf{X}_{\text{ids}}$, attention mask $\mathbf{X}_{\text{mask}}$, and labels $\mathbf{y}$.
    Transfer the batch to the computation device (e.g., GPU or CPU).
    Zero the gradients of $\mathcal{O}$.
    Compute predictions $\hat{\mathbf{y}}$ using mBERT:
      $\hat{\mathbf{y}} = \mathcal{M}(\mathbf{X}_{\text{ids}}, \mathbf{X}_{\text{mask}})$.
    Compute loss ℓ:
      $\ell = \mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$.
    Perform backpropagation to compute gradients:
      $\nabla_\theta \ell$.
    Update model parameters using the optimizer:
      $\theta \leftarrow \theta - \mathcal{O}(\nabla_\theta \ell)$.
  end for
  Evaluate the model on the validation set to compute metrics such as accuracy and F1-score.
end for
Return the fine-tuned mBERT model $\mathcal{M}$.

- Our approach to Aspect-based Sentiment Analysis (ABSA) in Hausa movie reviews involves a transformer-based model for aspect and polarity classification. The algorithm presented describes in detail dataset preparation, preprocessing, model design, training, and evaluation. Each step is carefully tailored to maximize the model's effectiveness in analyzing sentiment in an underrepresented language.

- 1652 Hausa movie review were collected and annotated for both aspect and polarity.
- Data preprocessing such Data cleaning, removing emojis and stopwords, Each text review was tokenized using a Autotokenzier from transformers and transformed into a vector using Hausa version of word2vec referred as hauwe.

- Data was split into training and validation on 70/30 ratio

Results

TABLE I: Performance of traditional machine learning of aspect word Extraction

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.70	0.71	0.70	0.67
SVM	0.7153	0.6900	0.7211	0.7021
Random Forest	0.7000	0.72	0.70	0.70
Logistic Regression	0.7624	0.7411	0.7602	0.7511

TABLE II: Performance of traditional machine learning on aspect polarity

Model	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.64	0.60	0.64	0.52
SVM	0.64	0.60	0.64	0.52
Random Forest	0.64	0.60	0.64	0.52
Logistic Regression	0.64	0.60	0.64	0.52

TABLE III: Performanc proposed model on aspect and polarity classification

Model	Accuracy
Result on Aspects Extraction:	
Naive Bayes	0.54
SVM	0.64
Random Forest	0.74
Logistic Regression	0.64
Proposed DCNN model	0.93
Result on Polarity classification:	
Naive Bayes	0.58
SVM	0.64
Random Forest	0.66
Logistic Regression	0.64
Proposed DCNN model	0.96

- We assess the performance of the mbert model trained and tested on the Hausa ABSA dataset.
- Table 1 and Table 2 presents the performance of traditional machine learning models tested on both Aspect extraction and Polarity classification on Hausa Movie reviews.
- Table 3 presents the results of our proposed model on both aspect on polarity classification. The accuracy of the model trained on Hausa movie review dataset reached the 96. 08%, which is better than traditional approaches. These results indicate the effectiveness of Haubert on handling ABSA in Low Resource Languages..

Conclusion

In this paper, we investigated the potential of pre-trained mutllingual transfer model finetuned on Low resource language on Aspect Based Sentimenta Analysis. Our research introduces a new dataset for ABSA, and new Model Huabert for ABSA in Hausa Language.