Hotels Arabic Reviews Sentiment Analysis

Abstract

There are several ways for guests to express their views, as hotels' primary goal is to fulfill their needs and elevate their expectations. Thus, in order to assist them in improving their services, we have developed two models in this project that aid in understanding visitor opinions. The Machine Learning Model (ML) and the Deep Learning Model (DL) are the two models. For ML, we employed Support Vector Machines (SVM), while for DL, we employed Long-Short-Term Memory (BLSTM). Both models perform well when it comes to natural language processing (NLP), particularly when dealing with Arabic text. We achieved a high accuracy, with 71% for the SVM of the ML model and 73% for the BLSTM of the DL model.

Introduction

Understanding guest feedback is essential for hotels, particularly with the explosion of the number of reviews from Arabic-speaking guests. due to the internet and websites like booking and Google Maps that allow guests to rate and write comments about the hotels and service. They offer a large number of valuable reviews that can improve the domain of Arabic semantic analysis. Moreover, there have been very advanced and high-performance artificial intelligence models that handle semantic analysis. Unfortunately, there is a leak in Arabic semantic analysis research and projects due to the complexity of the Arabic language. Our project leverages natural language processing techniques, specifically sentiment analysis, to interpret hotel reviews. Helping hotels to provide better customer service and improve guest experience.

Background/Related Work

Sentiment analysis of Arabic reviews is one of the Natural Language Processing (NLP) problems, which can solved by machine learning and deep learning techniques like Support Vector Machines (SVM) and Long-Short-Term Memory (LSTM). This Section presents a review of the literature that defines and discusses the use of SVM and LSTM in complex linguistic features of the Arabic language in light of technological progress and challenges to be faced.

- ### Support Vector Machines (SVM)

  The Support Vector Machine (SVM) is a machine learning algorithm containing labelled training data during training to find the patterns. SVM identifies an optimal hyperplane in high-dimensional space for data classification. This has the ability to deal with very complex, high-dimensional relationships in such a way that it supports maximum distance between different data classes (Zainuddin & Selamat, 2014). Various works have successfully used SVM for sentiment analysis in Arabic, which we will mention below:

  Alyami et al.(Alyami & Olatunji, 2020) proposed an SVM model for the task of sentiment classification in Arabic texts data from Twitter, focusing on topics discussed under Saudi Arabia's Vision 2030. Classifying data into two categories: positive or negative. Some of the techniques followed to reach the best possible classification accuracy are light stemming, feature extraction (Ngrams, emoji, and tweet-topic features), parameter optimization, and set reduction. The result of the SVM model showed a high classification accuracy, up to 89.83%.

  Louati et al.(Louati et al., 2023) developed the SVM-based sentiment analysis of Arabic student course reviews at Prince Sattam bin Abdulaziz University (PSAU). The preprocessed text of feedback from university students is cleaned from any sort of noise in the form of removal of irrelevant information. The former went into extensive details, including tokenization, stop-word removal, and finally using SVM for sentiment classification. Performance of SVM compared with the CAMeLBERT model shows that the former is competitive, classifying 69.62% of the reviews as positive.

- **Long-Short-Term Memory (LSTM)**

Long-Short-Term Memory (LSTM) is a deep learning model based on Recurrent Neural Network architecture. LSTM solves the long-term dependency and gradient problems that occur in the RNN. To clarify, RNN has short-term memory while LSTM has both long and short memory to leverage the long sequential dependency between features. Also, the backpropagation process in RNN is applied to matrix multiplications, which solves the problem of vanishing gradients. At the same time, LSTM backpropagates only on the long-term memory, which is a pairwise multiplication of parameters. This structure allows it to maintain long-term dependencies more effectively than traditional RNNs (Van Houdt et al., 2020). Moreover, the LSTM includes three gates: the forget gate that removes unneeded data, the input gate that adds new data, and the output gate that updates the short term memory. These architectures made LSTMs capable of holding the information for a long duration(Van Houdt et al., 2020).

One of the specialized forms of LSTM is the Bidirectional LSTM (BLSTM). It takes data through layers in both the forward and backward directions, hence capturing context far better. This feature, serving as a form of post-processing, comes in quite handy in Arabic Sentiment Analysis, where the full context of words and statements should be taken into consideration (Elfaik & Nfaoui, 2020). In another work, the authors present a method for sentiment analysis based on Word2Vec and BLSTM to extract opinions and polarity from Arabic text. The proposed model, using the vectorization of words and the power of BLSTM, was able to generate a considerably better accuracy compared to all other Machine learning (ML) , Convolutional neural network (CNN), or LSTM architectures. The maximum achieved accuracy in this research was 94.88%, which accounts for the ability of the proposed BLSTM-based deep learning approach to hold long-range information and give a correct representation of the future semantics of word sequences in Arabic sentiment analysis(Osama, Arabi, & Amira, 2023).

Approach

Our goal is to analyze hotel reviews to help hotel owners understand and interpret guest views in text data; rather than just analyzing text manually, we use two methods. The first one is a statistical-based method, which is SVM, and the second one is a deep learning-based method, which is Bidirectional LSTM. Each method is discussed in detail below:

Support Vector Machines (SVM)

Support Vector Machine (SVM) is a supervised machine learning model that is used for classification or regression tasks. It's used for NLP tasks involving Arabic text data, which we used to address our problem by categorization type. It is a popular choice in several Arabic language processing applications due to its versatility, robustness, and ability to handle high-dimensional feature spaces. Optimizing SVM's performance in Arabic NLP tasks requires extensive preparation and evaluation with different parameters. SVM is suited to handling such irregularities in Arabic text since it can ignore noisy features and learn from structured data representations. It can often generalize well to previously unseen Arabic text samples and performs well even with small training data. Figure 1 illustrate the process.

Long-Short-Term Memory (LSTM)

We used the Bidirectional LSTM architecture, a type of Recurrent Neural Network (RNN), for our Deep Learning model. That architecture is powerful for handling sequential data in NLP tasks, including those involving Arabic text. Its ability to capture long-term dependencies and contextual information makes it well-suited for understanding the intricacies of Arabic language syntax, morphology, and semantics. LSTM can effectively identify named entities, analyze sentiment, and machine translation. Figure 2 shows the architecture of BLSTM.

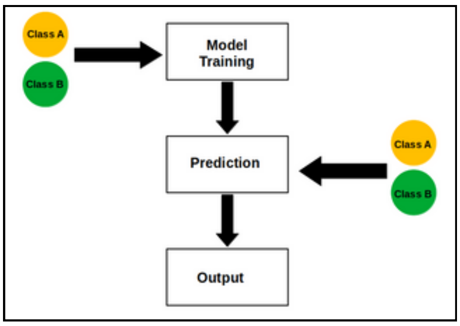

Figure.1: SVM

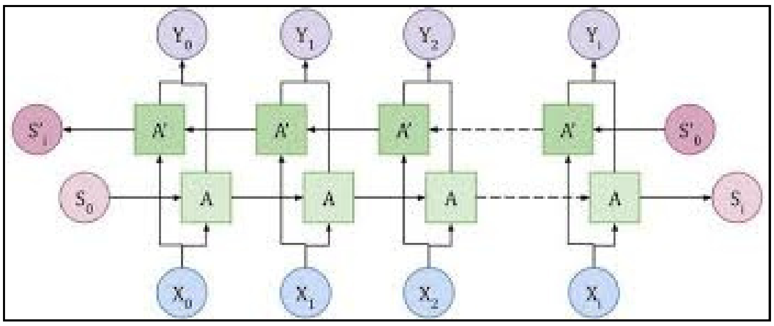

Figure.2: LSTM

Dataset Description

The Hotel Arabic-Reviews Dataset (HARD) is presented in a paper by Elnagar, A. et al. (2018) for Arabic hotel reviews extracted from the Booking.com website. The dataset is specialized for sentiment analysis and machine learning applications in the area of hospitality. However, the HARD dataset is very large, containing 490,587 samples, enabling it to build powerful tools for hotel sentiment analysis for hospitality if computational resources are available. The researchers provide both balanced and unbalanced subsets in Modern Standard Arabic (MSA) and Dialectal Arabic (DA). Furthermore, a subset of the dataset is available on Hugging Face at https://huggingface.co/datasets/hard with 105,698 samples with four labels. We selected 2500 samples for each class and show the statistics in the following figure.

	text	label
0	جيد جداً . مواقف السيارات صغيرة جدا	3
1	...ضعيف. للأسف لم يعجبني شئ ... الممرات بين الغرف	2
2	ضعيف جداً . كل شي	1
3	. .أتعس وأفشل فندق سكنت فيه بعمري"	2
4	...من أسوأ الفنادق التي عرفتها". اللوبي فقط الت"	2

```
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   text    10000 non-null  object
 1   label   10000 non-null  int64
dtypes: int64(1), object(1)
memory usage: 156.4+ KB
```

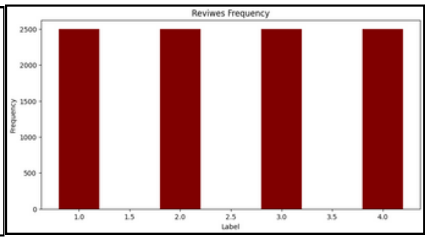

Figure.3: Dataset

Details of Experimental Procedures

This section will explain our implementation of machine learning and deep learning methods, including SVM and Bidirectional LSTM. We use Google Colab Pro to access cloud GPUs and high memory for software.

- SVM

We use TfidfVectorizer, which converts text to numerical values by computing the frequency multiplied by the inverse document frequency. The model was trained using the support vector classification (SVC) class from the scikit-learn library in Python. The SVC class implements the SVM method for classification problems. Figure 4 shows the parameters before and after tuning. We implemented a Random Search for hyperparameter tuning to determine the best values for the kernel, regularization parameter (C), kernel Coefficient (Gamma), and degree. Based on the Random Search results, the best values are 'C': 2.56, degree: 3, gamma: 'scale,' and kernel': 'rbf.'. The runtime for the model was two and a half hours, including Random Search.

SVM Parameters before	`SVC(C=1, degree=12, gamma=0.01, kernel='sigmoid')`
SVM Parameters After	`Best Parameters: {'C': 2.560186404424365, 'degree': 3, 'gamma': 'scale', 'kernel': 'rbf'}`

Figure 4: SVM Parameters

Details of Experimental Procedures

- Bidirectional LSTM

We tokenize the data; then, we use sequence padding to make the input the same size so that our neural networks can accept the data. Also, we convert numerical labels to one-hot encoded categorical labels. For the model configuration, we calculate the total number of unique words and set the vector space size with 100, making each word in the vocabulary presented as a 100-dimension vector. These two configurations are essential when building the model architecture. The model architecture has one embedding layer, one Bidirectional LSTM, two dense layers, and two dropout layers. Additionally, we use the Adam optimizer, categorical cross entropy, for the loss function. The runtime for the model is ten minutes. Figure 5 shows the architecture of our model.

```
# Define the model
model = Sequential([
    Embedding(input_dim=vocab_size, output_dim=embedding_dim, input_length=max_len_train),# Embedding layer
    Bidirectional(LSTM(16, return_sequences=False)),# Bidirectional LSTM layer
    Dropout(0.4),# Dropout layer for regularization
    Dense(16, activation='relu', kernel_regularizer=l2(0.0001)),# Fully connected layer with L2 regularization
    Dropout(0.7),# Dropout layer for regularization
    Dense(4, activation='softmax')# Output layer with softmax activation for classification into 4 classes
])
# Compile the model
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
```

Figure 5: Details of Experimental Procedures

Details of Experimental Procedures for Valuation metric

- Recall:

  Measures the proportion of actual positives that are identified correctly.
  Recall equation for each class:
  Recall=TP/(FN+TP)

- Precision:

  Measures the proportion of positive predictions that are actually correct.
  Precision equation for each class:
  Precision= TP/(FP+TP)

- F1-score:

  Measures the trade-off between Precision and Recall.
  F1-score equation for each class:
  F1-score=2× (Precision× Recall)/ (Precision + Recall)

- Macro average:

  Calculates the average recall, precision, and f1-score for all classes.
  Macro Average equation for a metric have four classes:
  Macro Average = (M1+M2+M3+M4)/4
  where Mi indicates the metric value

Details of Experimental Procedures for Valuation metric

- Weighted average:

  Computes the average recall, precision, and f1-score for all classes, weighted by the number of examples in each class.
  Weighted Average equation for a metric have four classes:
  Weighted Average =(M1×S1)+(M2×S2)+(M3×S3)+(M4×S4)/ (S1+S2+S3+S4)
  where M indicates the metric value, Si number of true example for class i

Results

SVM

```
The confusion matrix
 [[164 250  93   5]
 [ 25 337 138   0]
 [  1  14 473   4]
 [  2  15 362 117]]

 The classification report
               precision    recall  f1-score   support

           1       0.85      0.32      0.47       512
           2       0.55      0.67      0.60       500
           3       0.44      0.96      0.61       492
           4       0.93      0.24      0.38       496

    accuracy                           0.55      2000
   macro avg       0.69      0.55      0.51      2000
weighted avg       0.69      0.55      0.51      2000
```

Figure 6: SVM result before hyperparameter tunning

```
The confusion matrix
 [[388  87  21  16]
 [132 317  39  12]
 [ 16  45 340  91]
 [ 25  16  80 375]]

 The classification report
               precision    recall  f1-score   support

           1       0.69      0.76      0.72       512
           2       0.68      0.63      0.66       500
           3       0.71      0.69      0.70       492
           4       0.76      0.76      0.76       496

    accuracy                           0.71      2000
   macro avg       0.71      0.71      0.71      2000
weighted avg       0.71      0.71      0.71      2000
```

Figure 7: SVM result after hyperparameter tunning

Figure 6 shows the result of the SVM model before hyperparameter tuning; for precision, class 4 is a higher class with a 0.93 score, indicating that 93% of examples predicted as class 4 are actually class 4. For recall, class 3 achieved 0.96, showing that 96% of all examples are correct; for f1-score, class 2 achieved 0.60, indicating that 60% have a balance of precision and recall for class 2. Overall performance accuracy for f1-score is 0.55, and macro average and weighted average is 0.69 for precision, 0.55 for recall, and 0.51 for f1-score

Figure 7 shows the result of the SVM model after hyperparameter tuning; for precision, class 4 is a higher class with a 0.76 score, indicating that 76% of examples predicted as class 4 are actually class 4. For recall, class 1 and class 4 achieved 0.76, showing that 76% of all examples are correct; for f1-score, class 4 achieved 0.76, indicating that 76% have a balance of precision and recall for class 4. Overall performance of accuracy, macro average and weighted average is 0.71 for all measurement.

Bidirectional LSTM

	precision	recall	f1-score	support
Class 1	0.72	0.78	0.75	512
Class 2	0.70	0.67	0.69	500
Class 3	0.75	0.69	0.72	492
Class 4	0.75	0.78	0.76	496
accuracy			0.73	2000
macro avg	0.73	0.73	0.73	2000
weighted avg	0.73	0.73	0.73	2000

Figure 8: Bidirectional LSTM Result

Figure 8 shows the result of the Bidirectional LSTM model ; for precision, class 3 and class 4 is a higher class with a 0.75 score, indicating that 75% of examples predicted as class 3 and 4 are actually from class 3 and 4. For recall, class 1 and class 4 achieved 0.78, showing that 78% of all examples are correct; for f1-score, class 4 achieved 0.76, indicating that 76% have a balance of precision and recall for class 4. Overall performance of accuracy, macro average and weighted average is 0.71 for all measurement.

Comparison

By using TfidfVectorizer for the SVM model, the model overall performance for accuracy before hyperparameter tuning is 0.55, and after the hyperparameter tuning, it is 0.71. We use the Embedding layer in the Bidirectional LSTM model, which transfers each word into a dense vector. The model's overall performance for accuracy is 0.73. Additionally, SVM before hyperparameter tuning has higher precision, While SVM after hyperparameter tuning and Bidirectional LSTM provide stable performance across precision and recall, which is reflected in their higher F1-scores. To sum up Bidirectional LSTM is the best performing with balanced metrics across all classes, it does not contain extreamly low values means it is treat all classes fairly, als it has also the highest accuracy.

Conclusion

In conclusion, advanced techniques like Support Vector Machine (SVM) and deep learning were applied to understand the human feelings that exist in the Arabic hotel reviews. Based on this analysis, they are categorized into four classes. As powerful as these methods are, they are also affected by the complex structure of the Arabic language. It is such complexity that can prevent an accurate classification of emotions. All these are drawbacks, which require training analysis tools of large data with high-powered computing resources (GPUs) to adapt well to Arabic text. Despite these challenges, the project improves hotels' understanding of customer feedback and enhances services. Future work could explore domain-specific word embeddings, hybrid models, transfer learning, and aspect-based sentiment analysis to refine and optimize sentiment analysis in Arabic hotel reviews.

References

- Zainuddin, N., & Selamat, A. (2014). Sentiment analysis using support Vector Machine. 2014 International Conference on Computer, Communications, and Control Technology (I4CT). https://doi.org/10.1109/i4ct.2014.6914200
- Alyami, S. N., & Olatunji, S. O. (2020). Application of support vector machine for Arabic sentiment classification using Twitter-based dataset. Journal of Information & Knowledge Management, 19(01), 2040018. https://doi.org/10.1142/s0219649220400183
- Louati, A., Louati, H., Kariri, E., Alaskar, F., & Alotaibi, A. (2023). Sentiment analysis of Arabic course reviews of a Saudi University using support vector machine. Applied Sciences, 13(23), 12539. https://doi.org/10.3390/app132312539
- Van Houdt, G., Mosquera, C., & Nápoles, G. (2020). A review on the long short-term memory model. Artificial Intelligence Review, 53(8), 5929–5955. https://doi.org/10.1007/s10462-020-09838-1
- Elfaik, H., & Nfaoui, E. H. (2020). Deep Bidirectional LSTM network learning-based sentiment analysis for Arabic text. Journal of Intelligent Systems, 30(1), 395–412. https://doi.org/10.1515/jisys-2020-0021
- Elnagar, Ashraf & Khalifa, Yasmin & Einea, Anas. (2018). Hotel Arabic-Reviews Dataset Construction for Sentiment Analysis Applications. 10.1007/978-3-319-67056-0_3.
- hard · Datasets at Hugging Face. (n.d.-b). https://huggingface.co/datasets/hard