Journal of Algorithms and Computation

journal homepage: http://jac.ut.ac.ir

A Multilingual Sentiment Analysis Model in Tourism

M. J. Shayegan*[1] and Y. Dastan†[1]

[1]Department of Computer Engineering, University of Science and Culture, Tehran, Iran.

ABSTRACT

In recent years, scholars have dedicated significant attention to the field of sentiment analysis. A substantial volume of feedback shared by tourists on social networking platforms, notably on Tripadvisor, manifests as reviews. The tourism sector stands to gain valuable insights from sentiment analysis applied to such reviews. Despite the extensive body of research in sentiment analysis, scant attention has been directed toward multilingual sentiment analysis, particularly within the domain of tourism. This is noteworthy given the inherently multilingual and global nature of the tourism industry. This study aims to address this gap by presenting a comprehensive multilingual sentiment analysis conducted on Tripadvisor reviews. The sentiment analysis model is crafted using various layers of a neural network. We introduce an augmented Attention-based Bidirectional CNN-RNN Deep Model (Extended ABCDM). Comparative analysis reveals that the multilingual model attains a superior F1 measure of 0.732, outperforming previous models.

Keyword: Multilingual Sentiment Analysis, Hotel and Tourism, TripAdvisor, Transfer Learning, Machine Learning, Deep Learning.

ARTICLE INFO

Article history:
Research paper
Received 01, January 2024
Accepted 28, February 2024
Available online 01, August 2024

AMS subject Classification: 68T07.

*Corresponding author: M. J. Shayegan. Email: shayegan@usc.ac.ir
†yusef.dastan@yahoo.com

1 Introduction

In recent years, the integration of digital technologies has revolutionized service delivery, particularly within the hospitality sector. Notably, the accommodation industry has embraced online platforms and mobile applications for streamlined hotel reservation processes. Moreover, the emergence of social media platforms has provided a platform for hotel promotion and the dissemination of user-generated reviews, shaping consumer decision-making in the tourism sector. The analysis of user feedback on social media channels presents a promising avenue for enhancing customer satisfaction and informing future business strategies [1]. Increasingly, tourists are turning to social media platforms such as TripAdvisor for pre-booking, highlighting the importance of data-driven insights derived from these platforms [2]. Given the linguistic diversity inherent in hotel reviews, the adoption of multilingual sentiment analysis methodologies is crucial for accurate analysis. Platforms like TripAdvisor frequently host reviews in multiple languages on the same page, necessitating robust multilingual sentiment analysis techniques for comprehensive understanding and informed decision-making within the hospitality industry.

In the context of low-resource languages, the task of sentiment analysis and polarity detection becomes notably more challenging. While deep learning models generally outperform traditional approaches in this domain, their effectiveness hinges on the availability of substantial amounts of labeled data for training purposes. However, the process of labeling raw data is often prohibitively expensive and time-consuming [3]. Consequently, the primary hurdle in developing robust sentiment analysis models lies in sourcing datasets of sufficient size to adequately address the complexities of the task.

One novel solution to the problem of finding a dataset containing texts in a rare language is to use language-independent embeddings [4]. With language-independent embeddings, a model can be trained on a language with abundant resources, and that training can then be leveraged for low-resource languages.

Despite the advancements in sentiment analysis, particularly in the realm of monolingual models, the potential of multilingual research remains largely untapped. For instance, previous studies have demonstrated enhancements in monolingual model performance through techniques such as graph-based word connectivity [5] and novel feature selection [6]. However, the exploration of multilingual sentiment analysis remains relatively underexplored in recent research endeavors.

This study seeks to address this gap by proposing a comprehensive model for multilingual sentiment analysis specifically tailored for TripAdvisor, a prominent tourism social network characterized by the diverse linguistic composition of its user-generated content. The primary contribution of this research lies in the development and evaluation of a robust multilingual sentiment analysis model specifically designed for application within the TripAdvisor platform, thereby advancing the understanding and application of sentiment analysis in the context of multilingual tourism social networks.

A novel attention-based bidirectional CNN-RNN deep learning model for sentiment analysis is introduced in [7]. This research employs a multilingual approach, as detailed in the aforementioned study, wherein hotel reviews from TripAdvisor serve as the primary data

source. The initial phase involves the collection of user-generated reviews, followed by dataset preparation and subsequent pre-processing and embedding operations. Notably, words from diverse languages are transformed into vectors using preprocessing algorithms like Language-Agnostic SEntence Representations (LASER). Subsequently, the model integrates Long Short-Term Memory (LSTM), attention layer, and CNN components as per the methodology outlined in [7].
The structure of the paper is as follows: Section two provides a comprehensive review of related works, while Section three delineates the methodology employed. In section four, the research findings are elucidated, and section five entails the evaluation of the research outcomes.

2 Literature review

The application of sentiment analysis in hotel recommendation systems represents a burgeoning area of research, garnering significant attention within the realm of hotel management and the tourism industry [2]. Luo et al. [8] noted substantial shifts in sentiment analysis trends within English-language online hotel reviews in recent years. Neural networks, notably Convolutional Neural Networks (CNN) as demonstrated in [9] and [10], and LSTM networks employed in studies such as [11], [12], [13], [14], and [15], have emerged as prominent tools for sentiment analysis. Recent advancements have seen the integration of multilingual models, replacing traditional single-model approaches. Hybrid models, featuring distinct components for embedding and classification, have gained traction, facilitating text processing and classification without necessitating translation. These innovative models enable direct model transfer without intermediary translation steps. Noteworthy tools like fastText and LASER are instrumental in embedding multilingual texts [4]. LASER, developed by Facebook (now Meta), is an artificial neural network model trained to convert sentences into language-agnostic vectors, eliminating the need for prototypes and extensive datasets. It achieves this by producing vectors that represent sentences irrespective of language, wherein semantically similar sentences are mapped to proximate vectors, as illustrated in Figure 1 [16]. This capability underscores LASER's efficacy in multilingual sentiment analysis, accomplished at an accelerated pace without prior translation requirements.
The subsequent discussion provides an overview of prior research endeavors in sentiment analysis within the tourism domain, with a focus on bilingualism, tourism-related contexts, and the pursuit of heightened accuracy. Park et al. conducted a study employing Linguistic Inquiry and Word Count (LIWC), revealing distinct patterns in customer feedback within the hospitality sector [1]. Their findings underscored that loyal patrons tend to convey more positive sentiments in their comments, indicative of a stronger emotional attachment to the hotel. Conversely, individuals expressing negative sentiments in their feedback are less likely to revisit the establishment in the future.

Ray et al. [2] introduced a hotel recommender system leveraging the BERT model

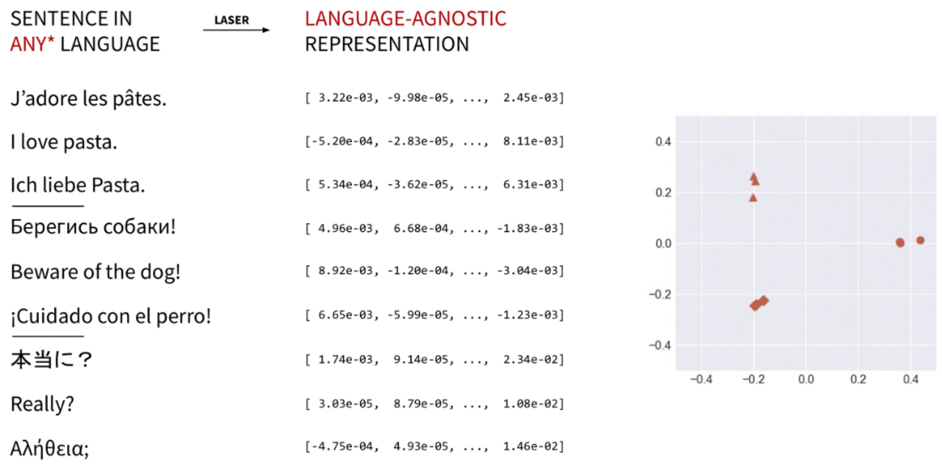

Figure 1: A view of LASER vectorization of different sentences

for sentiment analysis, an open-source framework for natural language processing (NLP) adept at deciphering contextual nuances in textual data. Following sentiment analysis, the system employed the Random Forest classification algorithm to generate appropriate responses to user queries.
Ghasemi et al. [3] devised a multilingual sentiment analysis model utilizing the Digikala dataset for Persian texts and Amazon for English texts. The model integrated BiBOWA and VecMap for word preprocessing and vectorization while employing CNN-LSTM for text classification.
Kanclerz et al. [4] developed a multilingual model by adapting preprocessing techniques and incorporating algorithms like fastText and LASER. Evaluation of this model employed CNN and Bidirectional LSTM (BiLSTM) algorithms across diverse datasets encompassing hotel, medical, school, and product domains.
Lou et al. [8] employed deep learning and fine-grained sentiment analysis on eLong.com data, revealing that factors such as cleanliness, noise levels, sanitation, and modern amenities hold greater sway over hotel customer satisfaction than room pricing.
Basiri et al. [7] introduced the bidirectional deep CNN-RNN (ABCDM) model, integrating GloVe word embeddings and bidirectional GRU layers, augmented with attention mechanisms. The model's output underwent polar analysis via a CNN neural network, outperforming six existing sentiment analysis algorithms [17-22].
Bijari et al. [5] used a structural graph to extract characteristics that can preserve the relations and sequences between words. Their method has been useful on different datasets such as Stanford University and IMDB datasets.
Chang et al. [6] enhanced the Categorical Proportional Difference (CPD) method by introducing the Modified CPD (MCPD), thereby optimizing its performance. They incorporated the Balance Category Feature (BCF) strategy, which entails selecting attributes equally from both positive and negative examples. Post MCPD and BCF evaluation, they employed the Support Vector Machine (SVM) algorithm for classification. This approach,

which selects features from both positive and negative samples, improves sentiment classification accuracy.

Chambua and Niu [23], based on a comprehensive review study, deduced that enhancing prediction accuracy is most effectively achieved through feature extraction, word weighting, and integrating multiple improvement methods. Aguero-Torales et al. [24] explored multilingual sentiment analysis articles and determined that CNN and CNN+BiLSTM classifier models yield the most accurate classification results. Additionally, they observed significant innovation in embedding techniques across most multilingual studies. Catelli et al. [25] addressed the challenge of limited Italian texts by employing Transfer Learning techniques, thereby achieving satisfactory sentiment analysis results through modifications to the BERT model.

Jin et al. [26] reported good performance utilizing Multinomial Naive Bayesian (MNB) classifiers, Support Vector Machine (SVM), and logistic regression (LR). Wu et al. [27] derived text polarity using Naive Bayesian and LSTM methods, followed by predictions using the ARIMAX model based on three criteria: bullish index, average index, and variance index.

Bian et al. [28] employed Conditional Random Field (CRF) and BiLSTM for sentiment analysis, subsequently extracting aspects through convolution. They then utilized a sentiment analysis model based on a dictionary to analyze sentiments related to the reviews. Venugopalan et al. [29] reported optimal results by integrating Latent Dirichlet Allocation (LDA) with the BERT model. Despite considerable efforts to enhance model accuracy in sentiment analysis, limited attention has been given to multilingual sentiment analysis, leaving ample room for improvement. Further research is warranted, particularly within tourism social networks like TripAdvisor, which is the focus of this study.

3 Methodology

Figure 2 illustrates the overall methodology employed in this research. The PolEmo dataset [30] serves as the foundation of this study, encompassing reviews across diverse domains including products, schools, pharmaceuticals, and hotels. For this investigation, solely the hotel segment of the dataset is utilized. These hotel datasets are sourced from TripAdvisor and are in the Polish language. The reviews within the dataset are categorized into six distinct classes:

- Strong Positive: Indicating wholly positive sentiments.
- Weak Positive: Showing mostly positive feelings with a few small negative points.
- Neutral: Neither positive nor negative sentiments.
- Weak Negative: Containing mostly negative feelings with a few positive aspects.

- Weak Negative: Containing mostly negative feelings with a few positive aspects.
- Weak Negative: Containing mostly negative feelings with a few positive aspects.
- Strong Negative: Signifying entirely negative sentiments.
- Ambivalent: Denoting a balanced mix of positive and negative sentiments within the text.

To assess the model's applicability across different languages, the original Polish reviews are professionally translated into Dutch, English, French, German, Italian, Portuguese, Russian, and Spanish [4]. The dataset comprises several text files, including separate files designated for Training, Validation, and Testing purposes in Polish. Subsequently, translations of the test file are generated into various languages including English, Spanish, and Portuguese, among others [4].
The LASER tool is employed for multilingual text embedding, converting sentences into 1024-dimensional vectors [16]. Regardless of the language, LASER ensures that sentences with identical meanings are mapped to proximate vectors, facilitating cross-lingual analysis.
Each text in the dataset typically contains around 24 sentences, with a typical middle value of 22 sentences. To streamline the learning and evaluation process, it's advisable to opt for powers of 2 when selecting input sizes for a neural network. Thus, numbers like 32 and 64 are suitable choices for the model [4].
To address the issue of insufficient sentences in some reviews, padding was implemented. This technique involves adding zero embedding vectors to the matrix of review sentence embeddings until it aligns with the network input. Consequently, each sentence in the review is considered within a framework of either 32 or 64 dimensions. The research methodology proposed here builds upon the ABCDM model [7], incorporating three familiar neural networks: Convolution, BiLSTM, and BiGRU. The ABCDM model introduces a novel attention layer, which, while not standard in the TensorFlow Library, has been utilized in prior studies [17] [7]. Within the attention layer, it's possible to identify significant phrases that may have minimal occurrences.
The model's input consists of a sequence of sentences converted into vectors, structured as 32 x 1024 dimensions per input record, with 32 sentences and 1024 associated vectors. This input is fed into two neural networks, BiLSTM and BiGRU, and subsequently passes through two attention layers before being processed by a lambda layer. Each lambda layer yields two outputs, totaling four outputs. These outputs are then directed to four Convolution layers, where important features are extracted using pooling techniques (note: all pooling layers are not depicted in Figure 2). These eight pooling layers are eventually concatenated together using a concatenate layer. Finally, the data is normalized using a batch-normalization layer, a technique that facilitates faster and more stable training of deeper neural networks [31]. The data is then classified into six distinct categories by a simple density layer.

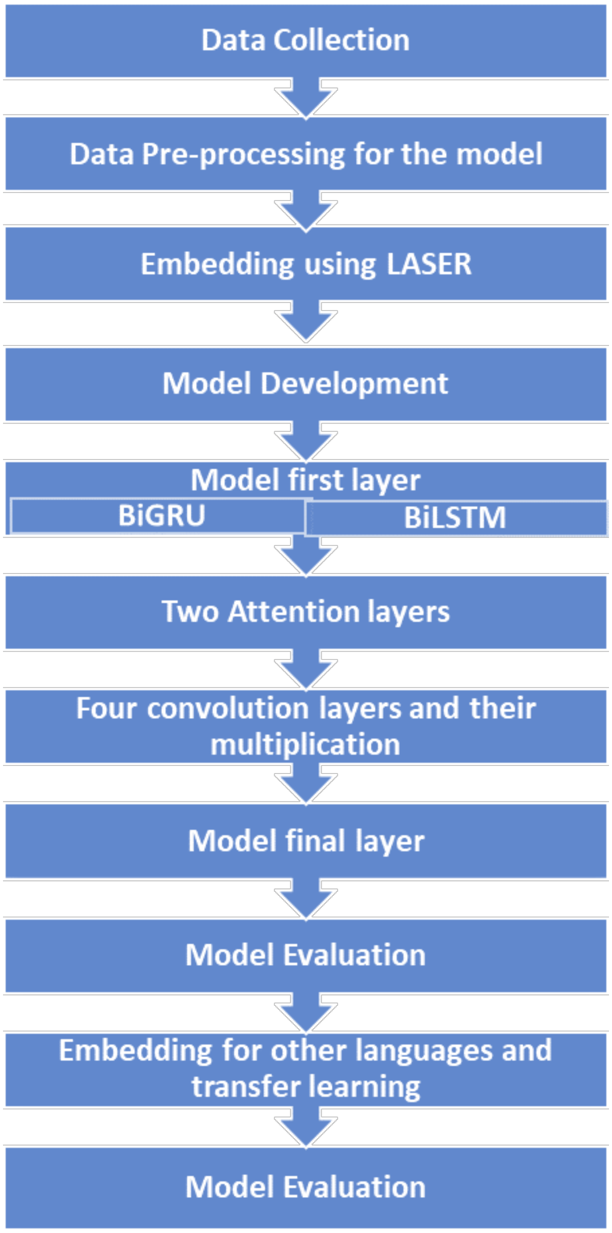

Figure 2: Overall Research Approach

This study utilizes a multilingual dataset, with the training and validation sections available only in Polish, while the test section is provided in English, Dutch, German, Spanish, Russian, French, Portuguese, Italian, and Polish [4]. A model is trained in Polish, with other languages processed via transfer learning. Transfer learning involves training a model on data from one context and applying the acquired knowledge to another context for prediction or classification. Multilingual sentiment analysis stands to benefit significantly from transfer learning [25].

Through the integration of LASER and transfer learning methods, this study has developed a model capable of classifying tourist reviews in a language-agnostic manner and extracting tourist sentiment. It's worth noting that the ability to train a model in one language and apply it across different languages allows for training with high-resource languages like English and utilization with low-resource languages such as Italian or any other language.

4 Findings and Results

In this study, the F1 measure is used for evaluation as in [4] (Equation 1):

$$F1\text{-measure} = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \tag{1}$$

We utilize the F1 measure and precision criteria to evaluate the algorithms due to the unbalanced nature of the dataset. We compare the performance of sentiment analysis using four models: the CNN model, the BiLSTM model, the AC-BiLSTM model as introduced in [17], and an enhanced version of the ABCDM model [7]. The CNN and BiLSTM models are adopted from [4]. Below are the parameters of the model:
verbose = 2
In a neural network, verbose is a choice regarding how you would like to view the output during the training process.
epoch = 64
epochs is a hyperparameter that determines how many times the learning algorithm will run through the entire training set.
batch_size = 32
The batch size refers to the number of samples processed before the model is updated. An epoch is defined as the number of complete passes through a training dataset.
Through iterative experimentation, these criteria were refined. We utilize three models—BiLSTM, CNN, and ACBiLSTM—as benchmarks for comparison. The ACBiLSTM model is a fusion of BiLSTM with an attention layer, enhancing its performance. It comprises a CNN layer followed by a BiLSTM layer, an extraction layer, an attention layer akin to the one in the ABCDM model, a flattening layer, a BiLSTM layer for recognition, and a simple density layer. This model, derived from [17], is considered robust for monolingual sentiment analysis.
Figure 2 depicts the ABCDM model adapted from [7], employing neural networks and diverse layers including attention, BiLSTM, BiGRU, CNN, Batch-Normalization, and dense. ABCDM leverages GloVe embedding in both forward and backward directions. In this study, LASER-based embedding replaces the original embedding part of ABCDM, utilizing transfer learning for embedding purposes.
After refining the embedding aspect, modifications are made to the ABCDM model. It encompasses two BiLSTM layers followed by an attention layer, connecting to four CNN layers through a lambda layer. Subsequently, the pooling layer extracts crucial features from these layers, and a concatenate layer optimizes their integration. Normalization is

then applied before classification into six predefined classes using a simple dense layer. This refined ABCDM structure, incorporating a specific attention layer from [7], differs from the attention layer in the Cross Library.
Performance outcomes for the ABCDM model are detailed in Table 1, while Table 2 illustrates the results of the models on Polish language test data and the average across other languages.

Table 1: The ABCDM model performance

	Result based on 32 Sentences			**Result based on 64 Sentences**		
Criterion / Dataset	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
Test (Polish)	0.7884	0.7287	0.7573	0.8209	0.6940	0.7521
English	0.6990	0.6372	0.6666	0.7749	0.6625	0.7143
Dutch	0.7270	0.6719	0.6938	0.7956	0.6877	0.7377
French	0.7509	0.6751	0.7109	0.7889	0.6719	0.7259
German	0.7552	0.6909	0.7216	0.8090	0.6814	0.7397
Italian	0.7431	0.6751	0.7074	0.7992	0.6656	0.7263
Portuguese	0.7517	0.6877	0.7182	0.7834	0.6845	0.7306
Russian	0.7679	0.6782	0.7202	0.8075	0.6751	0.7353
Spanish	0.7456	0.6751	0.7086	0.7818	0.6782	0.7263
Average	0.7476	0.6800	0.7116	0.7957	0.6779	0.7320

Table 2: Models Performance Comparison

	Result based on 32 Sentences			**Result based on 64 Sentences**		
Model / Dataset	**Precision**	**Recall**	**F1**	**Precision**	**Recall**	**F1**
CNN Test	0.8508	0.6556	0.7468	0.8379	0.6688	0.7438
CNN Average	0.8267	0.6326	0.7174	0.8214	0.6323	0.7144
BiLSTM Test	0.8264	0.6309	0.7155	0.8038	0.5300	0.6387
BiLSTM Average	0.7865	0.6029	0.6825	0.7278	0.5016	0.6131
ACBiLSTM Test	0.8137	0.6751	0.7379	0.8269	0.6782	0.7452
ACBiLSTM Average	0.8169	0.6393	0.7172	0.8183	0.6439	0.7206
ABCDM Test	0.7884	0.7287	0.7573	0.8209	0.6940	0.7521
ABCDM Average	0.7476	0.6800	0.7116	0.7957	0.6779	0.7320

The models' performance varies depending on the dataset. The extended ABCDM model demonstrates excellent results in sentiment analysis based on a test dataset comprising 32 sentences in Polish. Meanwhile, the CNN model performs better on average across various languages. ACBiLSTM outperforms other models in most languages but exhibits lower performance in the multilingual average compared to CNN. In Figure 3, the first 32 sentences are utilized to compare the models based on the F1 measure in a monolingual test.
In another scenario, the ABCDM model outperforms all other algorithms when analyzing sentiments based on the initial 64 sentences. For this model, the F1 measure for Polish language sentiment analysis reaches 0.7521, with an average of 0.732. This average surpasses that of any other model. Figure 4 illustrates the outcomes of different models

based on the first 64 sentences of the text.

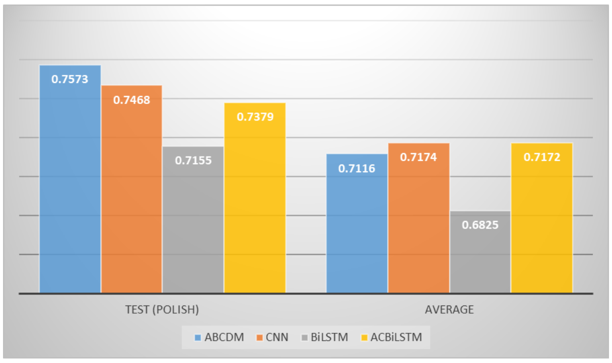

Figure 3: A comparison of the 32-sentence models based on the F1 measure

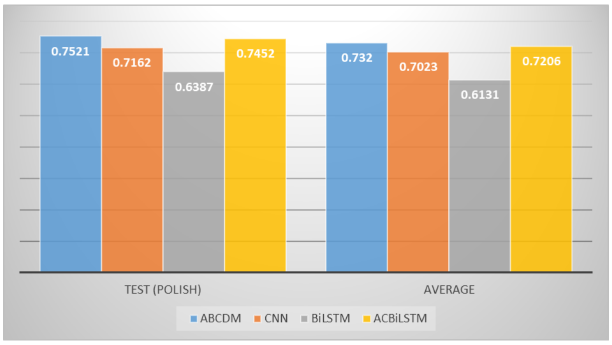

Figure 4: A comparison of the 32-sentence models based on the F1 measure

In this section, we compare the results of our model with those presented in [4]. The previous study introduced two multilingual models utilizing LASER embedding and CNN and BiLSTM neural networks. The CNN model in [4] achieved a rate of 0.7161, while the BiLSTM model achieved a rate of 0.7256. In our study, the extended ABCDM model achieved a rate of 0.732.
Figure 5 visually depicts the comparison between our research and others.

Additionally, the model undergoes evaluation using another dataset as outlined in [2]. This dataset comprises user reviews extracted from Tripadvisor in English [2]. The reviews encompass textual content, users' ratings of the hotel (referred to as "value"), and various hotel features such as cleanliness, services, location, beds, rooms, reception, and business services. To categorize the reviews, their average scores are utilized [2]. The average score ranges from 1 to 5, with scores falling between 1 and 2.5 deemed as negative, 2.5 to 3.5 as neutral, and scores above 3.5 considered positive. The testing dataset employed in this study is random and unbalanced. Specifically, the training dataset comprises 2000 English reviews of American hotels, the validation dataset consists of 500 English reviews

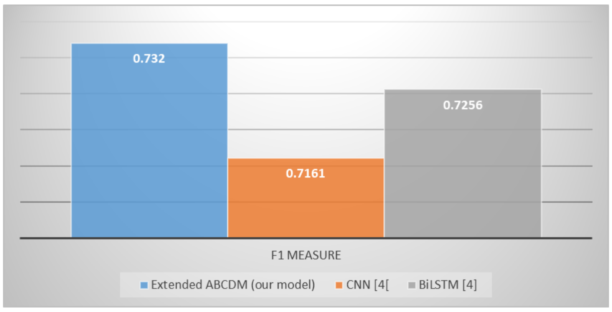

Figure 5: F1 measure comparison between the proposed model and related works

of American hotels, and the testing dataset includes 103 Persian reviews of Iranian hotels. The results obtained from the extended ABCDM model are presented in Table 3.

Table 3: Performance of English and Persian datasets.

Dataset / Criterion	Precision	Recall	F1
Test (English)	0.8059	0.8236	0.8146
Persian	0.8738	0.8738	0.8738
Average	0.8399	0.8487	0.8442

In this section, we introduced algorithms based on CNN, BiLSTM, ABCDM, and ACBiLSTM, commonly utilized in recent research. Each of these four models yielded distinct but comparable outcomes. Following the tests, the extended ABCDM model, CNN model, and ACBiLSTM model demonstrated strong performance. In particular, the extended ABCDM model showcased superior results when analyzing the majority of sentences.

5 Conclusion

Despite numerous studies in sentiment analysis, there's considerable scope for improvement in the multilingual tourism sector. This study extended and tailored a monolingual sentiment analysis model for multilingual use. With limited text sources, neural network models become more intricate, necessitating techniques like transfer learning. In our study, LASER was utilized for embedding, and datasets were classified using four different models: two BiLSTM and CNN models, an ACBiLSTM model from [17], and an extended ABCDM model inspired by [7]. According to our research findings, the extended ABCDM model outperformed all other models examined. The slightly lower performance of the extended ABCDM with 32 sentences can be attributed to LASER-based sentence embedding, where the attention layer may struggle to identify crucial words. However, when provided with 64 sentences, giving more information to the ABCDM model improved its performance, yielding a better F1 measure than all other models. The ABCDM model excels in monolingual sentiment analysis when using word-based embedding. However, compared to word-based embedding, it's less efficient with sentence-based embedding due to the difficulty in capturing long word relations. ABCDM is best suited when important words are scattered throughout the text. Our proposed model, an extension of the ABCDM model, achieved superior results when given more information, namely the first 64 sentences of each review. Future studies, as suggested in [2], could explore the impact of various hotel aspects such as service, cleanliness, location, affordability, and room count on traveler reactions. Additionally, multilingual sentiment analysis could be leveraged to assess customer retention rates, as discussed in [1].

References

[1] E. Park, J. Kang, D. Choi, and J. Han, "Understanding customers' hotel revisiting behaviour: a sentiment analysis of online feedback reviews," *Current Issues in Tourism*, vol. 23, no. 5, pp. 605-611, 2020.

[2] B. Ray, A. Garain, and R. Sarkar, "An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews," *Applied Soft Computing*, vol. 98, pp. 106935, 2021.

[3] R. Ghasemi, S. A. Ashrafi Asli, and S. Momtazi, "Deep Persian sentiment analysis: Cross-lingual training for low-resource languages," *Journal of Information Science*, vol. 48, no. 4, pp. 449-462, 2022.

[4] K. Kanclerz, P. Miłkowski, and J. Kocoń, "Cross-lingual deep neural transfer learning in sentiment analysis," *Procedia Computer Science*, vol. 176, pp. 128-137, 2020.

[5] K. Bijari, H. Zare, E. Kebriaei, and H. Veisi, "Leveraging deep graph-based text representation for sentiment polarity applications," *Expert Systems with Applications*, vol. 144, 2020, doi: 10.1016/j.eswa.2019.113090.

[6] J.-R. Chang, H.-Y. Liang, L.-S. Chen, and C.-W. Chang, "Novel feature selection approaches for improving the performance of sentiment classification," *Journal of Ambient Intelligence and Humanized Computing*, 2020, doi: 10.1007/s12652-020-02468-z.

[7] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharya, "ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis," *Future Generation Computer Systems*, vol. 115, pp. 279-294, 2021, doi: 10.1016/j.future.2020.08.005.

[8] J. Luo, S. Huang, and R. Wang, "A fine-grained sentiment analysis of online guest reviews of economy hotels in China," *Journal of Hospitality Marketing & Management*, vol. 30, no. 1, pp. 71-95, 2021.

[9] P. Cen, K. Zhang, and D. Zheng, "Sentiment analysis using deep learning approach," *J. Artif. Intell*, vol. 2, no. 1, pp. 17-27, 2020.

[10] L. Li, T.-T. Goh, and D. Jin, "How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis," *Neural Computing and Applications*, vol. 32, pp. 4387-4415, 2020.

[11] A. Pimpalkar, "MBiLSTMGloVe: Embedding GloVe knowledge into the corpus using multi-layer BiLSTM deep learning model for social media sentiment analysis," *Expert Systems with Applications*, vol. 203, p. 117581, 2022.

[12] N. D. Derra and D. Baier, "Working in detail: How LSTM hyperparameter selection influences sentiment analysis results," *Arch. Data Sci.* Ser. A, vol. 6, no. 1, p. 10, 2020.

[13] X. Fu, J. Yang, J. Li, M. Fang, and H. Wang, "Lexicon-enhanced LSTM with attention for general sentiment analysis," *IEEE Access*, vol. 6, pp. 71884-71891, 2018.

[14] M. Paolanti et al., "Tourism destination management using sentiment analysis and geo-location information: a deep learning approach," *Information Technology & Tourism*, vol. 23, no. 2, pp. 241-264, 2021/06/01 2021, doi: 10.1007/s40558-021-00196-4.

[15] Z. Hameed and B. Garcia-Zapirain, "Sentiment classification using a single-layered BiLSTM model," *IEEE Access*, vol. 8, pp. 73992-74001, 2020.

[16] M. Artetxe and H. Schwenk, "Massively multilingual sentence embeddings for zero-shot cross-lingual transfer and beyond," *Transactions of the Association for Computational Linguistics*, vol. 7, pp. 597-610, 2019.

[17] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325-338, 2019.

[18] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480-1489.

[19] X. Wang, W. Jiang, and Z. Luo, "Combination of convolutional and recurrent neural network for sentiment analysis of short texts," in *Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers*, 2016, pp. 2428-2437.

[20] S. Wen and J. Li, "Recurrent convolutional neural network with attention for twitter and yelp sentiment classification: ARC model for sentiment classification," in *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, 2018, pp. 1-7.

[21] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Systems with Applications*, vol. 117, pp. 139-147, 2019.

[22] A. Chatterjee, U. Gupta, M. K. Chinnakotla, R. Srikanth, M. Galley, and P. Agrawal, "Understanding emotions in text using deep learning and big data," *Computers in Human Behavior*, vol. 93, pp. 309-317, 2019.

[23] J. Chambua and Z. Niu, "Review text based rating prediction approaches: preference knowledge learning, representation and utilization," *Artificial Intelligence Review*, vol. 54, pp. 1171-1200, 2021.

[24] M. M. Agüero-Torales, J. I. A. Salas, and A. G. López-Herrera, "Deep learning and multilingual sentiment analysis on social media data: An overview," *Applied Soft Computing*, vol. 107, pp. 107373, 2021.

[25] R. Catelli et al., "Cross lingual transfer learning for sentiment analysis of Italian TripAdvisor reviews," *Expert Systems with Applications*, pp. 118246, 2022.

[26] L. Jin, L. Zhang, and L. Zhao, "Feature selection based on absolute deviation factor for text classification," *Information Processing & Management*, vol. 60, no. 3, p. 103251, 2023.

[27] D. C. Wu, S. Zhong, R. T. Qiu, and J. Wu, "Are customer reviews just reviews? Hotel forecasting using sentiment analysis," *Tourism Economics*, vol. 28, no. 3, pp. 795-816, 2022.

[28] Y. Bian, R. Ye, J. Zhang, and X. Yan, "Customer preference identification from hotel online reviews: A neural network based fine-grained sentiment analysis," *Computers & Industrial Engineering*, vol. 172, p. 108648, 2022.

[29] M. Venugopalan and D. Gupta, "An enhanced guided LDA model augmented with BERT based semantic strength for aspect term extraction in sentiment analysis," *Knowledge-based systems*, vol. 246, p. 108668, 2022.

[30] J. Kocoń, M. Zaśko-Zielińska, and P. Miłkowski, "Multi-level analysis and recognition of the text sentiment on the example of consumer opinions," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 2019, pp. 559-567.

[31] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, "How does batch normalization help optimization?," *Advances in neural information processing systems*, vol. 31, 2018.