UzABSA: Aspect-Based Sentiment Analysis for the Uzbek Language

Sanatbek Matlatipov*[1], Jaloliddin Rajabov[1], Elmurod Kuriyozov[2,3], Mersaid Aripov[1]
[1] National University of Uzbekistan named after Mirzo Ulugbek, Department of Applied Mathematics and Computer Analysis, 4, University Street, Tashkent city, 100174, Uzbekistan
[2] Universidade da Coruña, CITIC, Grupo LYS, Depto. de Computación y Tecnologías de la Información, Facultade de Informática, Campus de Elviña, A Coruña 15071, Spain
[3] Urgench State University, Department of Computer Sciences, 14, Khamid Alimdjan street, Urgench city, 220100, Uzbekistan
{s.matlatipov, j.rajabov}@nuu.uz, e.kuriyozov@udc.es, mirsaidaripov@mail.ru

Abstract

The objective of enhancing the availability of natural language processing technologies for low-resource languages has significant importance in facilitating technological accessibility within the populations of speakers of these languages. Our current grasping shows that there are no established linguistic resources available open source to develop aspect-based sentiment analysis (ABSA) tools tailored to the Uzbek language. This work aims to address the aforementioned gap by presenting the first high-quality annotated ABSA dataset - UzABSA. The data used in this study was obtained from a compilation of online reviews of Uzbek restaurants. Consequently, the constructed dataset has a length of 3500 reviews at the document level and 6100+ sentences at the sentence level. The popular approach to language resources of this kind explores four distinctive characteristics, namely Aspect Terms, Aspect Term Polarities, Aspect Category Terms, as well as Aspect Category Polarities. To the best of our knowledge, it is the first and the largest ABSA dataset for the Uzbek language. To evaluate the annotation process of our dataset, we used established statistical techniques such as Cohen's kappa coefficient and Krippendorff's α to assess agreement between annotators. Subsequently, a classification model, namely K-Nearest Neighbour (KNN), was used to evaluate the performance of the created dataset. Both sets of evaluation techniques demonstrate comparable levels of accuracy. The first findings across the various tasks showed promising outcomes, with accuracy rates ranging from 72% to 88%. This study not only highlights the significance of our acquired dataset but also plays a valuable tool for scholars interested in furthering sentiment analysis in the Uzbek language.

Keywords: Aspect-based Sentiment Analysis, Uzbek Language, Sentiment Dataset, low-resource languages

1. Introduction

Sentiment analysis (SA) is a critical component of natural language processing. It addresses the processing of opinions, feelings, and subjectivity by collecting, analyzing, and summarizing sentiment. It has gotten a lot of interest not just in academics, but also in business since it provides real-time feedback via online reviews on websites, where it may take advantage of people's thoughts on particular items or services. The task's underlying premise is that the whole text has an overall polarity. To conduct a more comprehensive analysis of the aforementioned viewpoint, it is necessary to develop an annotated Aspect-Based Sentiment Analysis (ABSA) corpus. Therefore, ABSA is critical in recognizing fine-grained emotions in user expressions(Zhang and Liu, 2017). Currently, Aspect-Based Sentiment Analysis has reached significant advancement in performance by using deep learning (including transformer-based) models by a thorough evaluation and aspect extraction methods(Chauhan et al., 2023). On the other hand, low-resource languages still lack access to those performance improvements. Using pre-trained language models such as BERT together with fine-tuning methods for ABSA classification tasks(Hoang et al., 2019; Chauhan et al., 2023) for both sentence-level and text-level documents has shown prominent accuracy results. However, to be able to perform such classification tasks, they require high-quality annotated ABSA data. It is essential to note that natural language processing (NLP) technologies, including sentiment analysis tools, get advantages from considering the particular features of the language being analyzed(Jang and Shin, 2010; Kincl et al., 2019).

Mostly spoken in Uzbekistan, the Uzbek language contains relatedness to the Turkic group and has a distinct agglutinative typology, like all others in the group, where words are formed by stringing morphemes without changing the spelling or phonetics of the word. Being a part of the Karluk group of the Turkic language, Uzbek has a common feature at the same time with all its members: vowel harmony and gender inflections but at the same time differ from them with some phonetic, lexical, and grammatical developments. Uzbek is different from most other Turkic languages in non-vowel harmony and exposure to the heavy influence of Persian, Arabic, and Russian, so it possesses a different vocabulary and phonemic structure. Affixes

SIGUL2024 Workshop, pages 394–403

define the grammatical relationship in Uzbek and permit the forming of new words through them to bring out an exceptional, systematic, and regular expression of the grammarians. This morphological characteristic is of huge importance to processing the language's elements in an application of natural language processing (NLP) and therefore forms a very interesting focus in the Turkic language world for research in the field of computational linguistics (Turaeva, 2015).

To our knowledge, there is no available ABSA dataset for the Uzbek language. Therefore, it is indeed helpful to transfer the language from low-resource(Nguyen and Chiang, 2017; Mukhamadiyev et al., 2023; Matlatipov et al., 2020) to high-resource. To fill that gap, we created, to our knowledge, the first high-quality ABSA dataset for the Uzbek language in a sentence-level(it can also be further merged into document-level because of its ID structure)[1] which is derived from online Uzbek restaurant reviews(Matlatipov et al., 2022), each systematically annotated to aid specific aspects of SA. The annotation covers four detailed areas agreed on the Annotation guideline: identifying specific Aspect Terms (T1) and their associated sentiments/polarities (T2), and categorizing broader Aspect Categories (T3) along with their polarities (T4). To ensure the validity and reliability of the corpus we established manual evaluations that measure consistency between human annotators. Therefore, we've used two widely accepted metrics for this purpose: Cohen's kappa(Cohen, 1960; Rau and Shih, 2021) and Krippendorff's α (Krippendorff, 2004) which underlines our commitment to data quality.

The main contributions of the paper are as follows:

1. The first annotated dataset for aspect-based sentiment analysis in the Uzbek language comprises reviews sourced from the domain of Uzbek restaurants which was preprocessed as well as cleaned from our previous work(Matlatipov et al., 2022). These reviews were collected by accessing accessible URLs on Maps. The data size for sentence-level analysis consists of 6175 instances, while for document-level analysis, it comprises 6500 reviews. It is worth noting that reviews have a maximum of 19 sentences.

2. An annotation guideline has been developed and made available at the project repository. The annotators were tasked with identifying aspect words, aspect term polarity, pre-defined aspect categories, and aspect category polarities to achieve the specified purpose. The primary emphasis of the guideline was the inquiries around the determination of which words or categories should be annotated as aspect terms, as well as which terms or categories should not be annotated with good examples to understand.

3. Evaluated the dataset using inter-annotator agreement using Cohen's Kappa, Krippendorff's α as well as classification model, namely K-Nearest Neighbour (KNN). All the accuracy results are comparable and reliable as follows:
   For the effective usage of the dataset, we used a machine learning model for aspect term extraction, aspect category extraction and sentiment polarity classification tasks. The evaluation exhibited for the T1 task an F1-accuracy of 75%, precision of 75.1%, and recall of 74.6%. T2 reported a simple ratio accuracy of 83%. T3 achieved an F1-accuracy of 87.8%, precision of 88%, and recall of 87.6%. T4 recorded a ratio accuracy of 85.3%.

2. Related Work

Aspect-based sentiment analysis (ABSA) has attracted significant interest in recent years owing to its capacity to provide more detailed sentiment analysis compared to conventional sentiment analysis methods(Liu, 2012). The mission of ABSA entails the identification of attitudes and aspects, which is a quite complex undertaking.

Datasets and benchmarks play a fundamental role in the assessment and advancement of ABSA. The workshops organized under the name SemEval (Semantic Evaluation) have played a crucial role in this aspect by presenting various tasks related to Aspect-Based Sentiment Analysis over the years. The SemEval-2014 Task 4 focused on the analysis of restaurant and laptop reviews(Pontiki et al., 2014), where participants were required to identify and classify different features within the evaluations. Subsequent endeavours, such as SemEval-2015 Task 12(Pontiki et al., 2015) and SemEval-2016 Task 5(Pontiki et al., 2016), built upon the preceding trials by using supplementary datasets, such as hotel reviews, and necessitating more detailed assessments of sentiment based on specific aspects.

In addition to SemEval, the dataset of Amazon product reviews, which was highlighted by McAuley et al. (2015), encompasses many product categories and has served as a fundamental resource for research on ABSA—the Yelp Dataset Chal-

[1] https://huggingface.co/datasets/Sanatbek/aspect-based-sentiment-analysis-uzbek

lenge[2] is considered to be a significant dataset that provides a diverse collection of restaurant reviews. This dataset is highly regarded within the ABSA (Aspect-Based Sentiment Analysis) community since it is recognized as a useful resource. The selection of a dataset, taking into account its domain specialization, the accuracy of annotations, and the intricacy of the reviews, may significantly impact the results of a sentiment analysis model. Benchmarks, particularly those derived from projects such as SemEval, serve as a foundation for evaluating various approaches, cultivating an atmosphere of competition and cooperation. This dynamic has played a crucial role in driving improvements in the field of aspect-based sentiment analysis(Nakov et al., 2019).

NLP advancements on the Uzbek language: Although there is currently no existing aspect-based sentiment analysis corpus available for the Uzbek language, there have been notable efforts to create natural language processing (NLP) resources and models, which may be regarded as a comprehensive advancement in resource creation for the language. Several noteworthy contributions have been made in the field, such as the development of datasets for sentiment analysis(Kuriyozov et al., 2022; Matlatipov et al., 2022), (Rabbimov et al., 2020) investigated the effect of emoji-based features in Uzbek texts' opinion classification, and more specifically movie review comments from YouTube. They tested some of the classification models, and feature ranking was performed to evaluate the discriminating ability of the emoji-based features. There is also a paper related to semantic assessment(Salaev et al., 2022b). The list of stop words as a source, a paper by Madatov et al. (2023) proposed the collocation method of detecting stop words of the corpus as well as stop-words dataset containing 731,156. Various natural language processing (NLP) tools have been created to facilitate NLP research and applications on Uzbek texts. These tools include transliteration between existing alphabets (Salaev et al., 2022a), syllabification tool (Salaev et al., 2023), as well as neural machine translation models (Allaberdiev et al., 2024). Nevertheless, further endeavours are required to enhance the efficacy of natural language processing (NLP) models when applied to Uzbek texts. Rabbimov and Kobilov (2020) conducted a study that focuses on the challenge of multi-class text categorization specifically for texts composed in the Uzbek language. Matlatipov and Vetulani (2009) studied Uzbek morphology which is one of the early and first works for Uzbek NLP. Uzbek morphology is studied using a theoretical framework that analyzes morphotactic and morphophonemic standards. The authors created the UZMORPP system for automated Uzbek morphological parsing. System Prolog implementation is supplied. (Abdurakhmonova et al., 2022) MorphUz is a Morphological analyzer(Mengliev et al., 2021) tool that is capable of segmenting a given text consisting of words into a sequential arrangement of morphemes. The first open-source and the biggest WordNET for the Uzbek language was created by (Agostini et al., 2021). The authors aim to provide a dataset for aspect-based sentiment analysis for the Uzbek language and assess the performance of several models using evaluation metrics such as F1-Score, Cohen's kappa, and Krippendorf's alpha. The TFIDF algorithm was used by the researchers, who utilized word-level and character-level n-gram models as methods for feature extraction. In addition, a list of stop-words was generated to eliminate them throughout the process of vectorizing the data. The researchers achieved a notable accuracy rate of 88% during their evaluation of an aspect-category recognition task using a specific dataset. The constraints of this study include a constrained dataset obtained just from a singular domain outlet, thereby yielding a limited scope for analysis and application.

3. Dataset

Restaurant domain[3] annotated corpora is used(Matlatipov et al., 2022), which is collected from The Google Maps based on Uzbek cuisine's locations where local national food reviews are the primary target. The sizes of the training and test data are shown in Table 1.

Name	**Train**	**Test**
absa-uz-all	5327	848
absa-uz-inter-annotator	760	760

Table 1: The length of the dataset where the first one is what is called gold(big) data and the second one is used for inter-annotator agreement between annotators

3.1. Tasks

1. **Task1(T1) Aspect term extraction:** Given a set of sentences with pre-identified entities (e.g., restaurants), identify the aspect terms present in the sentence and return a list containing all the distinct aspect terms. An aspect term names a particular aspect of the target entity.

[2] Yelp Dataset Challenge. `https://www.yelp.com/dataset/challenge`

[3] `https://huggingface.co/datasets/Sanatbek/Uzbek-restaurant-domain-sentiment-reviews`

- (e.g. "**Xizmat** va **xodimlar** muomilasi menga yoqdi, ammo **ovqat** yamon ekan"/ "I liked the **service** and the staff, but the **food** was bad").

2. **Task2(T2) Aspect term polarity:** For a given set of aspect terms within a sentence, determine whether the polarity of each aspect term is positive, negative, neutral or conflict (i.e., both positive and negative).

   - (same example above: **Xizmat** va **xodimlar** muomilasi menga yoqdi, ammo **ovqat** yamon ekan“ === xizmat: positive, xodimlar: positive, ovqat: negative).

3. **Task3(T3) Aspect Category detection:** Given a predefined set of aspect categories (ovqat(food), xizmat(service), narxi(price), muhit(environment, atmosphere), and boshqa(misc.)), identify the aspect categories discussed in a given sentence. Aspect categories are typically coarser than the aspect terms of task 1, and they do not necessarily occur as terms in the given sentence.

4. **Task4:Aspect category polarity:** Given a set of pre-identified aspect categories (e.g., food, price), determine the polarity (positive, negative, neutral or conflict) of each aspect category.

3.2. Annotation Process

The annotation process for the dataset adheres to the rules established by SemEval 2014 (Pontiki et al., 2014) shared task. Two annotators used BRAT (Stenetorp et al., 2012), a web-based annotation tool, that was suitably customized to meet the requirements of the ABSA task using annotation guideline??. The last step is the conversion of annotation format-based datasets into other suitable formats, such as JSONL, XML, and Parquet, therefore making them accessible on the HuggingFace platform. The annotation of each aspect term, together with its corresponding sentiment, is performed for every review sentence. Aspect categories are annotated using predefined five restaurant-related domain terms, and their polarities, namely positive, negative, neutral, and conflict. Figure 1 displays examples of the dataset, with their corresponding XML format.

Aspect	Value	Pos.	Neut.	Neg	Con
Terms	7412	4153	1601	1555	103
Categories	7724	4488	1518	1547	171

Table 2: Distribution of Aspect Terms and Categories in terms of counted Values, Positive, Neutral, Negative and Conflict for the UZABSA Dataset.

The data are shown in the table 2 reveals clear trends in sentiment distribution for aspect phrases and categories within the UzABSA dataset. The recorded count for aspect keywords is 7412, whereas the count for categories is 7724. The prevalence of positive emotion is evident in both classifications, with 4153 occurrences identified in aspect terms and an even higher figure of 4488 in aspect categories. It is worth noting that there is a tight correlation between the incidence of neutral feeling and aspect phrases, with a total of 1601 instances. However, the number of negative sentiments within aspect categories somewhat exceeds the number of neutral sentiments, with 1547 occurrences compared to 1518. It is worth noting that the sentiment of conflict, although occurring less often, is nevertheless evident with 103 occurrences for aspect terms and 171 occurrences for aspect categories.

4. Methodology

We are given the corpus of reviews where the main objective is to use a model $\mathcal{M}_{T_{1|2|3|4}}$ that predicts aspect terms(T_1), aspect terms polarities(T_2), aspect categories(T_3) and aspect categories polarities(T_4) from X:

$$\mathcal{M}_{T_{1|2|3|4}} : X \rightarrow \hat{Y}_{T_{1|2|3|4}}$$

The K-Nearest Neighbors (KNN) technique was used to construct the function $\mathcal{M}_{T_{1|2|3|4}}$ for aspect-based sentiment analysis. The technique included four distinct tasks.

The Aspect Term extraction(Task T_1), involves the extraction of aspect terms. The input data X underwent preprocessing, which included tokenization, stemming, and stop word removal. The K-nearest neighbours (KNN) algorithm was used to train a model for predicting aspect terms (T_1) based on the feature space(TF-IDF word embeddings).

The Aspect Term Polarity Prediction (Task T_2) involves predicting the polarity of aspect terms. The K-nearest neighbours (KNN) algorithm effectively performed multi-class classification to reliably forecast the polarities of aspect terms (T_2).

The Task of Aspect Category Extraction (Task T_3): Predefined aspect categories, such as "food quality" and "service," were established. The K-nearest neighbours (KNN) algorithm was used to classify phrases into distinct groups after a preprocessing step. The multi-class capacity of the model played a vital role in task T_3.

The task of Aspect Category Polarity Prediction (Task T_4) involves predicting the polarity of aspect categories. The polarity of retrieved aspect categories was assessed using sentiment analysis methods. The aspect category polarities (T_4) were predicted by the KNN algorithm using the classified characteristics.

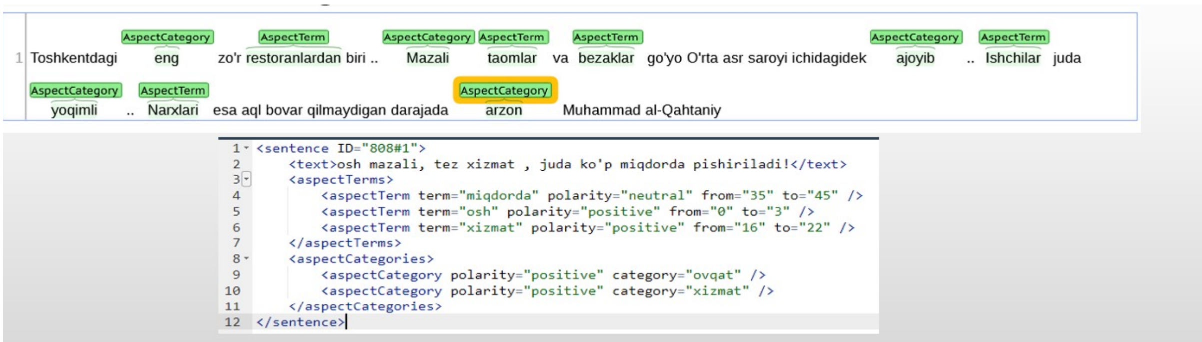

Figure 1: Sample review annotated in the BRAT tool with five aspect terms and five predefined aspect categories. The below image is an XML snippet that corresponds to the annotated sentence

The process of assessing the performance and effectiveness of a model. The function $\mathcal{M}_{T_{1|2|3|4}}$ that was created was subjected to a thorough evaluation utilizing metrics such as F1-accuracy, precision, and recall. Moreover, the validation process included comparing inter-annotator agreement data(small portion), namely Cohen's Kappa and Krippendorff's α evaluation metrics which will be discussed below.

The efficacy of the KNN-based technique was proved in its application to aspect-based sentiment analysis in Uzbek restaurant reviews.

5. Evaluations

To evaluate Gold(G) with Test(T) dataset, we have used F1-score, Cohen's kappa coefficient and Krippendorff's α to evaluate the accuracy of aspect terms and aspect category detection tasks. The biggest annotated corpus is evaluated as 6000 training data and 848 test data using F1-score, whereas, the inter-agreement evaluation dataset contains 313 reviews with 760 sentences and annotation made only for sentence-level which have been calculated using Cohen's kappa coefficient and Krippendorff's α as following:

5.1. Metrics used for inter-annotator agreement

The ABSA task evaluation has been evaluated between two annotators, who were native speakers of the Uzbek language. To check the quality of annotations by different annotators we calculate inter-rater/inter-coder agreements of the same document on 760 sentences where one of them is taken from what is considered a gold dataset. The reason of limited time and source, annotators could only partially annotate the same reviews, whereas the rest of the corpus is annotated only once. Firstly, we calculated Cohen's Kappa (κ)(Cohen, 1960) to quantify the inter-annotator agreement among annotators. **Cohen's Kappa:** measures the validity coefficient of UzABSA dataset where agreement between two annotators are classified N objects into C mutually exclusive categories [4]. Cohen's Kappa k coefficient takes into account the possibility of chance agreement.

$$k = 1 - \frac{1 - P(O)}{1 - P(E)} \tag{1}$$

where $P(E) = \frac{1}{N^2}\sum_{k=1}^{n}(\sum_{i=1}^{n} a_{ik} * \sum_{j=1}^{n} a_{kj})$ and $P(O) = \frac{1}{N}\sum_{k=1}^{n}(a_{kk})$. Here, $P(O)$ is the actual agreement among raters, p_e is the hypothetical probability of chance agreement, $n \in |\{G \cup T\}|$ is a number of classes created by Gold and Test dataset and $a \in A$ the number of times raters i, j predicted category k. Below, the A confusion matrix (Figure 2) is illustrated for T2, T3 and T4 tasks, whereas, the T1 task has more than 650 classes, so we decided to skip the illustration.

Krippendorff's (α) (Krippendorff, 2004) is a measure of inter-coder agreement(Krippendorff, 2004), which is used for assessing the reliability of UzABSA annotations. The reason we chose α-agreement as it handles incomplete (missing) data, any number of values available for coding a variable, binary, nominal, ordinal, interval, ratio, polar, and circular metrics, as well as small sample sizes of the reliability data are all applicable. It also adapts to incomplete data and missing values.

$$\alpha = 1 - \frac{D_o}{D_e} \tag{2}$$

where:

- D_o is the observed disagreement.

$$D_o = \sum_{i=1}^{N}\sum_{j=i+1}^{N} \delta(x_i, x_j)$$

The dissimilarity function $\delta(x_i, x_j)$ is for categorical data to quantify the dissimilarity between annotations for data points i and j.

[4] https://en.wikipedia.org/wiki/Cohen%27s_kappa

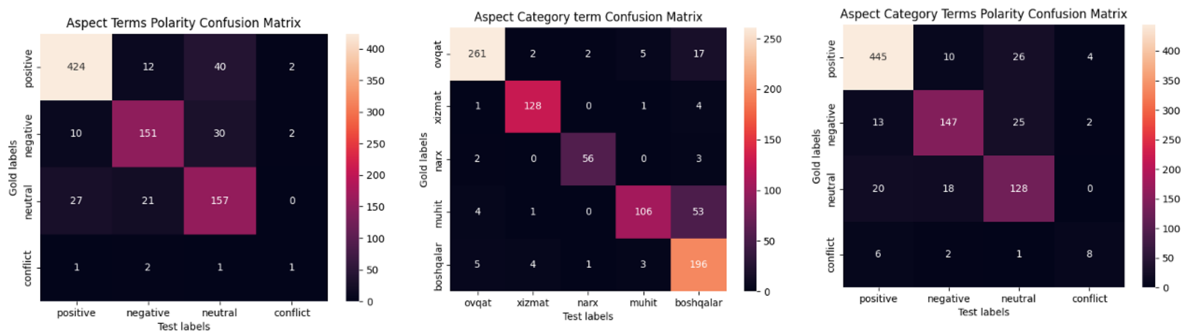

Figure 2: Confusion Matrices for T2(left), T3(Middle), T4(Right)

- D_e (Expected Disagreement):

$$D_e = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{l=1}^{L} \delta_l \cdot \delta_l$$

where δ_l is the expected probability of disagreement for label l.

for T_1 and T_3: The harmonic mean of precision(P) and recall(R) are used to evaluate $\mathcal{M}_{T_{1|3}}$ using F1-score:

$$F1_{T_{1|3}} = \frac{2 \cdot \mathsf{P}_{T_{1|3}} \cdot \mathsf{R}_{T_{1|3}}}{\mathsf{P}_{T_{1|3}} + \mathsf{R}_{T_{1|3}}} \quad (3)$$

for T_2 and T_4: Only the harmonic mean of precision(P) is used to evaluate $\mathcal{M}_{T_{2|4}}$:

$$P = \frac{\sum(|G \cap T|)}{|T|} \quad (4)$$

5.2. Results of the Evaluation

The evaluation results for small inter-annotator agreement data are shown in table 3.

1. T_1**:** F1-scores have been calculated by two annotators' agreements where the comparison with 881 aspect terms for gold and 876 aspect terms for the test dataset. The result retrieved 75% F1-accuracy with 75.1% Precision as well as 74.6% Recall.
Cohen's Kappa score retrieved 72% accuracy whereas Krippendorffś alpha for nominal matrix retrieved 55%.

2. T_2**:** Simple ratio accuracies have been calculated by two annotators' agreements where comparison output 727 correctly annotated out of 876 aspect term polarities. The ratio accuracy performed 83%.
Cohen's Kappa score retrieved 72.4% accuracy whereas Krippendorffś alpha for nominal matrix retrieved 88%.

3. T_3**:** F1-score have been calculated by two annotators' agreements where the comparison with 855 aspect categories for gold and 851 aspect categories for the test dataset. The result retrieved 87.8% F1-accuracy with 88% Precision as well as 87.6% Recall.
Cohen's Kappa score retrieved 83.4% accuracy whereas Krippendorffś $alpha$ for nominal matrix retrieved 83.3%.

4. T_4**:** Simple ratio accuracies have been calculated by two annotators' agreements where comparison output 726 correctly annotated out of 851 aspect category polarities. The ratio accuracy performed was 85.3%.
Cohen's Kappa score retrieved 75% accuracy whereas Krippendorffś alpha for nominal matrix retrieved 78%.

The evaluation results for small absa-uz-all data are shown in table 4.

The assessment ratings for the whole UzABSA dataset are shown in Table 4. In the context of task T_1, the F1-score was determined to be 44.8%, with accuracy calculated at 48% and recall measured at 42%. In the context of task T_2, the accuracy score achieved the greatest value, namely 55%. Task T_3 attained an F1-score of 64%, with an accuracy of 70% and a recall of 59%. The accuracy score achieved the greatest value of 67% in task T_4.

The findings shown in Table 3 demonstrate the effectiveness of UzABSA in measuring inter-annotator agreement, hence shedding information on the dataset's consistency across various tasks. The assessment shown in Table 4 provides an expanded study of the whole dataset, highlighting the difficulties encountered in attaining precise aspect-based sentiment analysis in the Uzbek language. The disparities in accuracy, recall, and F1-score seen across different tasks highlight the intricate nature of aspect-based sentiment analysis and emphasize the need for more refinement and study in this domain. The following sections provide a more in-depth analysis and discussion of these results with a conclusion.

Table 3: UzABSA evaluation scores for small inter-annotator agreement data. Number of Best scores per task are highlighted.

tasks	Aspect count for train	Aspect count for test	F1-score	Cohen's kappa	Krippendorff's α
T_1	881	876	**0.75**	0.72	0.55
T_2	876	727	0.83	0.724	**0.88**
T_3	855	851	**0.878**	0.834	0.833
T_4	851	726	**0.85**	0.75	0.78

Table 4: UzABSA evaluation scores for all data. The numbers with the best scores per task are highlighted.

tasks	Aspect count for train	Aspect count for test	F1-score	Precision	Recall
T_1	7412	2822	**0.448**	0.48	0.42
T_2	6703	1302		**0.55**	
T_3	6655	1069	**0.64**	0.7	0.59
T_4	6807	1200		**0.67**	

6. Conclusion and Discussion

This study showcases substantial advancements in the domain of aspect-based sentiment analysis within the context of the Uzbek language. Initially, we carefully selected and annotated an innovative dataset that was particularly designed for this particular objective. The dataset used in this study was obtained from evaluations specifically about Uzbek restaurants. Before analysis, the dataset underwent thorough pre-processing and cleaning procedures, which were informed by previous research efforts conducted by Matlatipov et al. (2022). The dataset used in our study consisted of 6500 reviews, which were analyzed at the sentence level. Specifically, we focused on 6175 occurrences, with each review including no more than 19 sentences.

To guarantee the quality and uniformity of our annotations, we have created a detailed annotation guideline. The guideline, which may be accessed via a designated URL, offers comprehensive directions to annotators about the identification of aspect terms, aspect term polarity, pre-defined aspect categories, and aspect category polarities. The guideline emphasised the intricate work of choosing the words or categories that should be annotated as aspect terms. This was further supported by providing illustrative examples to enhance clarity and understanding.

In addition, our research included meticulous assessment procedures to substantiate the efficacy and dependability of the annotated dataset. Inter-annotator agreement data, such as Cohen's Kappa and Krippendorff's α, were used to evaluate the level of consistency among the annotators. Furthermore, we have used the K-Nearest Neighbour (KNN) method, a machine learning model, to perform aspect word extraction, aspect category extraction, and sentiment polarity classification tasks. The assessment findings on small inter-annotator agreement data showcased our dataset's resilience and our methodology's efficacy. In the context of aspect term extraction (T1), our results indicate an F1-accuracy of 75%, accompanied by a precision of 75.1% and a recall of 74.6%. In the task of aspect category extraction (T2), we achieved a straightforward ratio accuracy of 83%. In the task of sentiment polarity classification (T3), our model demonstrated a noteworthy F1 accuracy of 87.8%. Additionally, it achieved a precision of 88% and a recall of 87.6%. Finally, in the task of aspect category polarity classification (T4), we obtained an accuracy ratio of 85.3%.

The challenges encountered in the whole dataset are shown in Table 4. Task T_1 shows a significant decrease in the F1-score, suggesting difficulties in extracting aspect terms. This might be attributed to the presence of different and sophisticated linguistic expressions. Task T_2 has the maximum level of accuracy, indicating precise polarity assignments for the aspect terms that have been found. Task T_3 exemplifies a well-balanced compromise between accuracy and recall, hence showcasing the dataset's effectiveness in detecting aspect categories. Task T4 has a high level of accuracy, suggesting that the dataset has the potential to determine the polarity of aspect categories accurately. Nevertheless, the lack of recall values indicates possible opportunities for expanding the dataset and improving the model. The findings of this study highlight the intricate and subtle nature of aspect-based sentiment analysis in the Uzbek language. This research brings attention to the difficulties encountered in accurately identifying specific aspect words, categorizing them, and determining their related polarity. The resolution of these issues has the potential to facilitate the development of sentiment analysis models that are more precise and dependable in future research.

The aforementioned contributions jointly provide a useful resource within the field of aspect-based sentiment analysis in the Uzbek language. The dataset we have carefully selected, together with the comprehensive annotation guideline and rigorous assessment measures, provides a solid foundation for future progress in sentiment analysis research, specifically in the context of Uzbek restaurant reviews.

7. Data Availability

All the code used in this work is openly available at https://github.com/SanatbekMatlatipov/uzabsa. Also, the UzABSA dataset has been uploaded to the HuggingFace Models Hub at https://huggingface.co/datasets/Sanatbek/aspect-based-sentiment-analysis-uzbek.

8. Acknowledgements

This publication has been produced within the framework of the Grant “REP-25112021/113 - UzUDT: Universal Dependencies Treebank and parser for natural language processing on the Uzbek Language”, funded under the MUNIS Project, supported by the World Bank and the Government of the Republic of Uzbekistan. The statements do not necessarily reflect the official position of the World Bank and the Government of the Republic of Uzbekistan.

9. Conflicts of Interest

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

10. Bibliographical References

Nilufar Abdurakhmonova, Ismailov Alisher, and Rano Sayfulleyeva. 2022. Morphuz: Morphological analyzer for the uzbek language. In *2022 7th International Conference on Computer Science and Engineering (UBMK)*, pages 61–66.

Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova, and Mukhammadsaid Mamasaidov. 2021. UZWORDNET: A lexical-semantic database for the Uzbek language. In *Proceedings of the 11th Global Wordnet Conference*, pages 8–19, University of South Africa (UNISA). Global Wordnet Association.

Bobur Allaberdiev, Gayrat Matlatipov, Elmurod Kuriyozov, and Zafar Rakhmonov. 2024. Parallel texts dataset for uzbek-kazakh machine translation. *Data in Brief*, pages 110–194.

Ganpat Singh Chauhan, Ravi Nahta, Yogesh Kumar Meena, and Dinesh Gopalani. 2023. Aspect based sentiment analysis using deep learning approaches: A survey. *Computer Science Review*, 49:100576.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 20(1):37–46.

Mickel Hoang, Oskar Alija Bihorac, and Jacobo Rouces. 2019. Aspect-based sentiment analysis using BERT. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pages 187–196, Turku, Finland. Linköping University Electronic Press.

Hayeon Jang and Hyopil Shin. 2010. Language-specific sentiment analysis in morphologically rich languages. In *Coling 2010: Posters*, pages 498–506.

Tomáš Kincl, Michal Novák, and Jiří Přibil. 2019. Improving sentiment analysis performance on morphologically rich languages: Language and domain independent approach. *Computer Speech and Language*, 56:36–51.

Klaus Krippendorff. 2004. *Content Analysis: An Introduction to Its Methodology (second edition)*. Sage Publications.

Elmurod Kuriyozov, Sanatbek Matlatipov, Miguel A. Alonso, and Carlos Gómez-Rodríguez. 2022. Construction and evaluation of sentiment datasets for low-resource languages: The case of uzbek. In *Human Language Technology. Challenges for Computer Science and Linguistics*, pages 232–243, Cham. Springer International Publishing.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.*, 5(1):1–167.

Khabibulla Madatov, Shukurla Bekchanov, and Jernej Vičič. 2023. Automatic detection of stop words for texts in uzbek language. *Informatica*, 47(2).

Gayrat Matlatipov and Zygmunt Vetulani. 2009. *Representation of Uzbek Morphology in Prolog*, pages 83–110. Springer Berlin Heidelberg, Berlin, Heidelberg.

Sanatbek Matlatipov, Hulkar Rahimboeva, Jaloliddin Rajabov, and Elmurod Kuriyozov. 2022. Uzbek sentiment analysis based on local restaurant reviews. *CEUR Workshop Proceedings*, 3315:126 – 136. Cited by: 1.

Sanatbek Matlatipov, Ualsher Tukeyev, and Mersaid Aripov. 2020. Towards the uzbek language endings as a language resource. In *Advances in Computational Collective Intelligence*, pages 729–740, Cham. Springer International Publishing.

Julian McAuley, Rahul Pandey, and Jure Leskovec. 2015. Inferring networks of substitutable and complementary products. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.

Davlatyor Mengliev, Vladimir Barakhnin, and Nilufar Abdurakhmonova. 2021. Development of intellectual web system for morph analyzing of uzbek words. *Applied Sciences*, 11(19).

Abdinabi Mukhamadiyev, Mukhriddin Mukhiddinov, Ilyos Khujayarov, Mannon Ochilov, and Jinsoo Cho. 2023. Development of language models for continuous uzbek speech recognition system. *Sensors (Basel)*, 23(3):1145.

Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2019. Semeval-2016 task 4: Sentiment analysis in twitter.

Toan Q. Nguyen and David Chiang. 2017. Transfer learning across low-resource, related languages for neural machine translation. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 296–301, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

I M Rabbimov and S S Kobilov. 2020. Multi-class text classification of uzbek news articles using machine learning. *Journal of Physics: Conference Series*, 1546(1):012097.

Ilyos Rabbimov, Iosif Mporas, Vasiliki Simaki, and Sami Kobilov. 2020. Investigating the effect of emoji in opinion classification of uzbek movie review comments. In *Speech and Computer*, pages 435–445, Cham. Springer International Publishing.

Gerald Rau and Yu-Shan Shih. 2021. Evaluation of cohen's kappa and other measures of inter-rater agreement for genre analysis and other nominal data. *Journal of English for Academic Purposes*, 53:101026.

Ulugbek Salaev, Elmurod Kuriyozov, and Carlos Gómez-Rodríguez. 2022a. A machine transliteration tool between uzbek alphabets. *CEUR Workshop Proceedings*, 3315:42 – 50.

Ulugbek Salaev, Elmurod Kuriyozov, and Carlos Gómez-Rodríguez. 2022b. Simreluz: Similarity and relatedness scores as a semantic evaluation dataset for uzbek language. *1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages, SIGUL 2022 - held in conjunction with the International Conference on Language Resources and Evaluation, LREC 2022 - Proceedings*, page 199 – 206. Cited by: 2.

Ulugbek I. Salaev, Elmurod R. Kuriyozov, and Gayrat R. Matlatipov. 2023. Design and implementation of a tool for extracting uzbek syllables. *Proceedings of the 2023 IEEE 16th International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering, APEIE 2023*, page 1750 – 1755. Cited by: 0.

Maksud Sharipov and Ogabek Sobirov. 2022. Development of a rule-based lemmatization algorithm through finite state machine for uzbek language. *CEUR Workshop Proceedings*, 3315:154 – 159.

Maksud Sharipov and Ollabergan Yuldashov. 2022. Uzbekstemmer: Development of a rule-based stemming algorithm for uzbek language. *CEUR Workshop Proceedings*, 3315:137 – 144.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France. Association for Computational Linguistics.

Rano Turaeva. 2015. Linguistic ambiguities of uzbek and classification of uzbek dialects. *Anthropos*, 110(2):463–476.

Lei Zhang and Bing Liu. 2017. *Sentiment Analysis and Opinion Mining*, pages 1152–1161. Springer US, Boston, MA.