Weakly-Supervised Learning for Aspect-Based Sentiment Analysis of Urdu Tweets

Zoya
Department of Computing
School of Electrical Engineering & Computer Science
National University of Sciences & Technology
H-12, Islamabad, Pakistan
zmaqsood.phdcs17seecs@seecs.edu.pk

Abstract

Aspect-based sentiment analysis (ABSA) is vital for text comprehension which benefits applications across various domains. This field involves the two main sub-tasks including aspect extraction and sentiment classification. Existing methods to tackle this problem normally address only one sub-task or utilize topic models that may result in overlapping concepts. Moreover, such algorithms often rely on extensive labeled data and external language resources, making their application costly and time-consuming in new domains and especially for resource-poor languages like Urdu. The lack of aspect mining studies in Urdu literature further exacerbates the inapplicability of existing methods for Urdu language. The primary challenge lies in the preprocessing of data to ensure its suitability for language comprehension by the model, as well as the availability of appropriate pre-trained models, domain embeddings, and tools. This paper implements an ABSA model Huang et al. (2020) for unlabeled Urdu tweets with minimal user guidance, utilizing a small set of seed words for each aspect and sentiment class. The model first learns sentiment and aspect joint topic embeddings in the word embedding space with regularization to encourage topic distinctiveness. Afterwards, it employs deep neural models for pre-training with embedding-based predictions and self-training on unlabeled data. Furthermore, we optimize the model for improved performance by substituting the CNN with the BiLSTM classifier for sentence-level sentiment and aspect classification. Our optimized model achieves significant improvements over baselines in aspect and sentiment classification for Urdu tweets with accuracy of 64.8% and 72.8% respectively, demonstrating its effectiveness in generating joint topics and addressing existing limitations in Urdu ABSA.

1 Introduction

Opinion mining and sentiment analysis have gained significant importance in analyzing user-generated content for various applications. However, the vast volume of textual content on numerous platforms like social media makes manual analysis impractical. Hence, there is a need for automatic computational frameworks to extract opinions from unstructured texts, leading to the development of sentiment analysis and opinion mining as a research field.

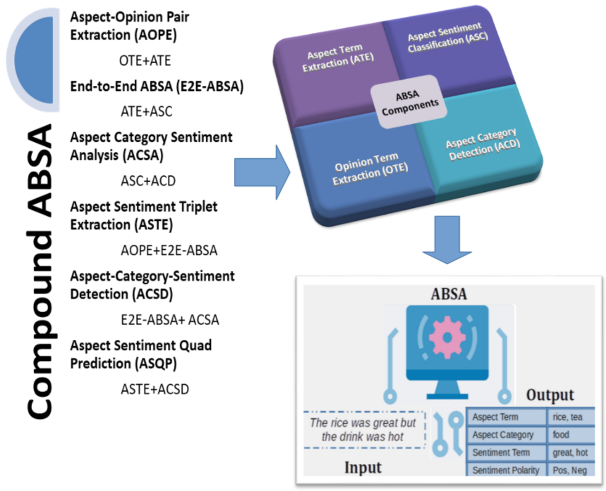

Figure 1: ABSA Individual vs. Compound Elements

Conventional sentiment analysis studies mainly perform prediction at the sentence or document level Yu and Hatzivassiloglou (2003); Pang et al. (2002), identifying the overall sentiment towards the whole sentence or document. However, ABSA (Cai et al., 2021) surpasses traditional sentiment analysis by incorporating finer granularity and enabling the identification of two crucial components: target aspect and sentiment in a given sentence or document. The target represents an entity or aspect of an entity, while the sentiment signifies an expression of opinion. For instance, consider sentences from restaurant reviews "The rice was great" and "The drink was hot". Both sentences convey sentiments but the term 'great' in the first sentence represents a positive sentiment related to the rice aspect, whereas the term 'hot' in the second sen-

Proceedings of the Student Research Workshop associated with RANLP-2023, pages 78–86, Varna, Bulgaria, Sep 4–6, 2023.
https://doi.org/10.26615/issn.2603-2821.2023_009

tence signifies a negative sentiment associated with the drink aspect. ABSA primarily revolves around four target components: aspect category, aspect term, sentiment term, and sentiment polarity Hemmatian and Sohrabi (2019). A comprehensive understanding of opinions at the aspect level requires considering multiple combinations of these components in different scenarios, generally referred to as compound ABSA. However, early ABSA research focused on individual components, such as capturing aspect terms while ignoring terms related to opinions like Aspect Term Extraction (ATE) Yang et al. (2020). Nowadays, various frameworks have been proposed to tackle such compound ABSA tasks that aim to extract and link individual elements Zhao et al. (2020); Cai et al. (2021). Such tasks are further denoted with several abbreviations, based on the specific combinations they address in compound ABSA. Figure 1 illustrates these tasks with their underlying combinations of ABSA individual components, as well as a complete ABSA of an opinion.

Most of the research in ABSA has been conducted in resource-rich languages, primarily English. Recent studies generally adopt supervised frameworks Hu et al. (2021); Yang et al. (2020) that require a large number of labeled sentences for training. Some studies leverage word embeddings in unsupervised He et al. (2017a) or weakly supervised settings Wu et al. (2020); Wang et al. (2021) for aspect extraction without annotated documents. These methods often rely on external language resources or syntactic structures such as POS tags or lexicons. Nevertheless, opinions are expressed in multiple languages in real-world scenarios. Acquiring large annotated data and accurate language resources for new domains or low-resource languages is challenging and expensive. Cross-lingual and cross-domain transfer requires language-specific knowledge of the target language and domain. Besides, previous methods relying on translation systems are limited by translation quality Zhang et al. (2021).

Alike other low-resource languages, Urdu also encounters difficulties in terms of data availability, language tools, and research resources. Additionally, Urdu is a comparatively complex language holding characteristics of multiple languages that pose troubles in minor processing tasks Khattak et al. (2021). Besides, the use of informal language on social media platforms further complicates the preparation of suitable datasets for machine learning models. As a result, ABSA tasks have been largely overlooked in the Urdu language and pre-trained models developed for other languages may not generalize effectively. Due to limited literature, many approaches and difficulties as discussed for English need to be explored for the applicability of ABSA in the Urdu language. Meanwhile, the popularity and widespread usage of Urdu on online platforms demand sincere efforts to address these challenges for the advancement of Urdu technologies.

This paper presents a significant contribution to ABSA in Urdu by addressing its unique facets and challenges for Urdu texts. We adopt a pioneering step by employing a weakly-supervised model Huang et al. (2020) developed for English ABSA to unlabeled Urdu tweets on the budget topic. Our methodology involves rigorous pre-processing to ensure data quality and model compatibility to avoid memory leakage, training disruptions caused by invalid tokens, and out-of-vocabulary issues. Moreover, tokenization tools are deeply analyzed to determine the optimal approach for token formation in Urdu, taking into account their pivotal role in comprehending language context and optimizing resource utilization. We evaluate the model with vanilla settings and conduct experiments with architecture modifications specifically replacing CNN with BiLSTM and utilizing Fasttext embeddings. Clustering techniques and graph network analysis are applied to enhance seed words selection rather than manually.

2 Literature Review

Several approaches have been proposed for aspect extraction and sentiment classification in ABSA. Some methods focus on independently addressing these two sub-tasks, while others adopt a joint approach to simultaneously solve them.

2.1 Isolated Extraction of Aspect and Sentiment

The literature review covers various methods for individual element extraction of ABSA. Supervised methods use token-level classification and multi-label classification for ATE and Aspect Category Detection (ACD) Yang et al. (2020); Yin et al. (2020); Hu et al. (2021) respectively. Different techniques such as sequence labeling, dependency paths, and word embeddings have been employed for ATE, while ACD studies use word co-occurrence patterns, hybrid features, and neural models. Opinion Term Extraction (OTE) includes aspect opinion co-extraction (AOCE) and target-oriented opinion word extraction (TOWE) using sequence labeling, dependency-tree, attention-based models, and those considering syntactic structures and positional embeddings with various encoder structuresWang et al. (2016); Li and Lam (2017); Wu et al. (2020). Deep learning-based Aspect Sentiment Classification (ASC) models Zhou et al. (2019); Liu et al. (2020) have shown performance improvements, especially using pre-trained language models and neural network-based dependency parsing. Contrarily, Weakly-supervised methods use few keywords for aspect guidance and data augmentation for self-training Chen and Qian (2020); Wang et al. (2021), while unsupervised methods He et al. (2017b); Luo et al. (2019) use contrastive learning and autoencoder models for aspect extraction and ACD.

The aforementioned methods primarily target the extraction of aspect or opinion words in isolation, which may not fully capture the comprehensive aspect-level opinion. To achieve a deeper understanding, extracting multiple sentiment elements and recognizing their relationships by incorporating their semantic meanings is crucial.

2.2 Joint Extraction of Aspect and Sentiment

Recent research has focused on joint ABSA tasks involving multiple sentiment elements and can be further divided into sub-tasks such as ATE, ACD (presented in Figure 1) aim to extract individual elements. Regardless of the method used, OTE is commonly viewed as an auxiliary task, as it provides valuable insights into the existence of aspect terms and sentiment orientation Liang et al. (2021). Several modeling paradigms like pipeline, joint, and unified have been explored for these tasks. Mao et al. (2021) employed a joint method for the End-to-End ABSA task, which trains the subtasks jointly and uses two label sets to predict the aspect boundaries and sentiment labels, and the final prediction is derived by combining the outputs of the two subtasks. Another approach dismisses the boundary between the subtasks and employs a unified tagging scheme, where both sentiment elements are denoted in the tag of each token Li et al. (2019). In the context of Aspect-Opinion Pair Extraction(AOPE), early studies have employed a pipeline approach to extract aspect and opinion terms in pairs Chen et al. (2020). An alternative method is to extract aspects first and then identify corresponding opinion terms Gao et al. (2021). Unified approaches to address AOPE aim to mitigate potential error propagation from the pipeline method by considering joint term and relation extraction Zhao et al. (2020).

Aspect Category Sentiment Analysis (ACSA) is commonly solved by pipeline approach Hu et al. (2019), where aspect categories are detected first and then sentiment polarities are predicted for those categories. Huang et al. (2020) learns joint topic embeddings for sentiment and aspect pairs and aims to enhance topic distinctiveness as compared to existing models primarily based on conventional topic models LDA often resulting in topic overlapping. The Aspect Sentiment Triplet Extraction (ASTE) task Peng et al. (2020) proposes a two-stage pipeline method for extracting the triplets. In the first stage, two sequence tagging models extract aspects, sentiments, and opinion terms separately. In the second stage, a classifier is employed to identify valid aspect-opinion pairs from the predicted aspects and opinions, thereby constructing the triplet prediction. To better capture the relationships between multiple sentiment elements, several unified methods have been proposed for ASTE, such as multi-task learning frameworks Zhang et al. (2020). The recently proposed Aspect Sentiment Quad Prediction (ASQP) task Cai et al. (2021) focuses on predicting all four sentiment elements in a quadruplet form for a given text item.

Researchers have explored the use of pre-trained language models (PLMs) for different ABSA tasks Mao et al. (2021); Chen et al. (2021). They have found that a simple linear classification layer combined with PLMs can outperform complex neural ABSA models. However, relying solely on PLMs as context-aware embeddings may not be sufficient, as ABSA tasks require capturing dependency relations. Additional designs are needed to fully utilize contextualized embeddings from PLMs and improve the robustness of PLM-based ABSA models Xing et al. (2020).

Early works such as Mitchell et al. (2013) favor the pipeline method, while Li et al. (2019) illustrates that using a tailor-made neural model alongside the unified tagging scheme yields better performance. Hence, research works based on either method can achieved good performance, making it challenging to compare and determine the superiority of one method over the others. Further exploration is needed in this regard. Moreover, many ABSA tasks have been solved by supervised approaches and others are weakly-supervised approaches. However, weakly supervised models trained on a specific dataset or domain may struggle to generalize well to new or diverse data. They may not effectively capture the nuances and variations in sentiment expressions across different contexts and complex languages. These models may produce suboptimal results when faced with such linguistic challenges. Manually selected seed words or heuristics to guide the learning can limit the model's ability to adapt to new domains or capture emerging sentiment patterns that are not covered by the seed words. It can be challenging to establish precise correspondences between aspect terms and sentiment expressions, especially when multiple aspect terms or sentiments coexist within a sentence. This ambiguity can result in noisy or incorrect predictions.

2.3 ABSA in Urdu Language

Existing studies primarily focus on a document or sentence-level labeling and employ machine-learning techniques for binary classification Amjad et al. (2021); Rani and Anwar (2020). Lexicon and rule-based methods are commonly used for opinion mining in Urdu due to limited labeled data and resources required for advanced techniques Khattak et al. (2021). As supervised and weakly-supervised methods requiring large labeled datasets or relying on other linguistic resources are less applicable to Urdu. Therefore, fine-grained ABSA remains unexplored even at its basic level.

3 Methodology

Our methodology consists of three main stages: pre-processing, seed word selection, and model implementation. Figure 2 illustrates the entire process.

3.1 Dataset

We used the Twitter API to extract approximately 100 Urdu tweets daily from each trending topic in Pakistan. Our dataset consists of around 5 million tweets covering diverse subjects from April 2020 to August 2020. We filtered out non-Urdu tweets, truncated tweets, and retweets. The dataset is based on generic query terms and encompasses all topics discussed during that period.

3.2 Pre-processing

Urdu language processing involves several pre-processing steps to structure the input text and reduce noise. These steps enhance the overall data quality and improve the accuracy for further linguistic and computational analysis. We have categorized them into three levels: tweet, token, and character.

3.2.1 Tweet Pre-processing-Level I

This involves cleaning the tweets by normalizing Unicode, punctuation marks, diacritics, hashtags, URLs, emojis, white spaces, hyperlinks, email addresses, phone numbers, mentions, and special characters like currency symbols. Regular expressions and Urdu-specific libraries are used for this purpose.

Unicode Normalization: Urdu text normalization is crucial due to its inclusion in the Arabic Unicode block, which can lead to multiple Unicode representations for Urdu alphabets. This step addresses variations in Unicode values caused by Arabic Unicode or orthographic changes and ensures accuracy in language models.

Stopwords Removal: insignificant words with minimal impact on the model, are removed to reduce vocabulary size. Urdu Language Processing (ULP) categorizes them as generic (applicable to all domains e.g. prepositions) and specific (domain-specific). A total of 1264 stopwords from various sources are collected and removed from the dataset.

Duplicate and Null Removal: Duplicate tweets resulting from related hashtags are eliminated to ensure dataset uniqueness. Additionally, tweets with no useful information like tweets that fell below a specified length threshold or null text are excluded, resulting in a reduced dataset size of approximately 1.2 million tweets.

Consecutive Words & Sentences Removal: We conducted an analysis of consecutive words and sentences within tweets and eliminated any repetitive instances to a certain limit, thus reducing the unnecessary length in the tweets e.g. سلیکٹڈ بجٹ سلیکٹڈ بجٹ... (selected budget selected budget...) or چور چور چور چور... (thief thief thief thief...). This step ensure that the tweets were concise and free from redundant content, resulting in more streamlined and focused text for further analysis

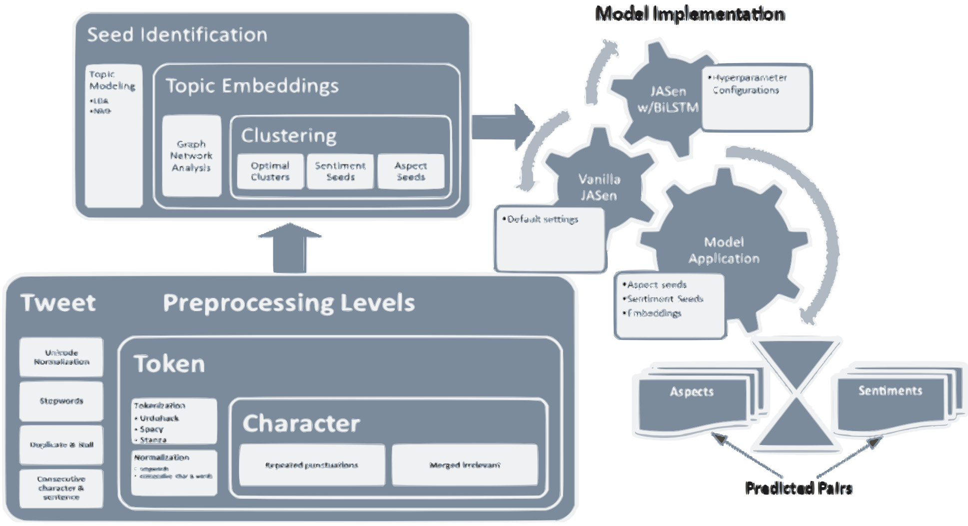

Figure 2: Methodology

3.2.2 Token Pre-processing-Level II

At this level of processing, tokens were formed from the input tweet and their correctness was observed for accuracy and consistency ensuring that they corresponded to the intended linguistic units and were free from errors or irregularities.

Tokenization: We compared multiple linguistic tools such as Qi et al. (2020), ALi (2020), and Vasiliev (2020) with the traditional space-based method to determine the most effective tokenization strategy for the Urdu language that yielded the lowest error rate and produced accurate tokenization results. To assess the performance, we provided a test set containing different forms of incorrect tokens caused by the absence of spaces or the use of informal language on the Twitter platform. Among the various options, we found that Urduhack ALi (2020) performed relatively better demonstrating a better level of accuracy. As a result, we selected Urduhack as the preferred library for tokenizing our dataset.

Normalization: Misspelled words that were joined together in tweets were normalized to their standard form at various levels of processing. *Stopwords Normalization:* Consecutive stopwords merged due to the absence of spaces at various levels like تھیاک (wasone) two stopwords forming incorrect word corrected to تھی اک (was one). Additionally, prior to normalization, these joined words were checked against a list of valid frequent words in Urdu to ensure that they do not form any other correct word.*Consecutive Characters & Words Normalization:* We noticed consecutive irrelevant characters within a token as well as the repetition of words without spaces often resulted in incorrect and lengthy tokens e.g.پاکسسسستان,فریفریفری (Pakisssstan, freefreefree) and rectified them.

3.2.3 Character Pre-processing-Level III

This level of pre-processing involved dividing the token further into individual units such as characters and compared with valid Urdu alphabets to facilitate further analysis.*Repeated punctuations removal:* We discovered the missed tokens in the cleaning process consisting of consecutive punctuation marks due to the absence of space and rectified them e.g.'****'.*Merged irrelevant character removal:* We identified and corrected instances where punctuation, digits, alphabets, or misprinted strings were merged inside a token, leading to incorrect formations like نواز ہوــنے کاc002u (Nawaz'sbecomingu200c).

3.3 Optimal Seed Words Selection

We employed topic modeling to identify the general topics within our dataset of tweets using conventional topic models like LDA & NMF Jelodar et al. (2019). Specifically, we focused on the topic of 'budget', which yielded approximately 10K related tweets. To analyze this topic further, we generated tweet embeddings using vanilla multilingual SBERT Reimers and Gurevych (2019). To ensure the accuracy of our methodology, we recognized the significance of seeding aspects and sentiment words. As a result, we utilized various techniques and clustering algorithms to obtain reliable seed words for each sub-topic under the 'budget' category. These techniques involved determining the optimal number of clusters as aspect categories through distance-driven clustering or other cluster validation metrics. Our experiments indicated that on average 3-9 clusters provided optimal results (as depicted in Figure 3), and we ultimately selected 9 clusters as our final choice. However, we encountered overlapping words across different clusters, which presented challenges in finalizing the aspect categories. To address this, we performed graph analysis where edges represented cosine similarities between embeddings. We set edge weights to 0 for values below a threshold of 0.7. This approach allowed us to group the most similar tweets within each cluster while obtaining distinct terms across the clusters. The predicted aspect categories and their corresponding aspect terms are presented in Table 1.

3.4 Model Application & Analysis

We experimented with the weakly-supervised JASen model as our baseline, developed for the English language, for Urdu tweets related to budget topic. We analyze its performance under various experimental configurations involving modifications to the dataset, embeddings, model architecture, and hyperparameters. Our dataset consists of approximately 10194 tweets of which we labeled 194 tweets as our test dataset and the remaining 10000 tweets as our unlabeled training dataset. Initially, we applied this model with default settings on our tweets dataset and compared it with the originally reported results on restaurant and laptop reviews to establish a performance benchmark. However, we employed FastText embeddings for representing the tweets, considering their effectiveness in capturing subword information. Though, we did not achieve up to mark results on these experiments for Urdu dataset but the reason could be the chosen seeded words as all sentiment words did not fit well for every aspect category. Subsequently, we made architectural modifications by replacing the CNN component with a BiLSTM layer to capture temporal dependencies in the tweet data. We tested the model with different hyperparameter settings and identified the best-performing configuration like alpha=0.05, hidden size=300, embedding size=100. The modified JASen model with BiLSTM and FastText embeddings outperformed the vanilla settings and showed improved accuracy and sentiment analysis results on Urdu tweets presented in Tables 2, 3, 4 and 5. However, there were no significant improvements observed on the benchmark datasets. These findings emphasize the importance of model modifications and the choice of embeddings for enhancing ABSA in Urdu. The limitations encountered during experimentation, such as memory leakage and out-of-vocabulary words, model disruption due to bad token as highlighted in section 3.2, were addressed through enhanced pre-processing of Urdu text.

Aspect Category	Aspect Terms
Govt. Spending	Expenditure (خرچ), Salary (تنخواہ), Pension (پنشن), Poor (غریب), Loans (قرضوں)
Subsidy & Grant	Subsidy (سبسڈی), Project (پروجیکٹ), Plans (منصوبے), Specific (مختص), Funds (فنڈز)
Social Welfare	Employees (ملازمین), Agriculture (زراعت), Reforms (اصلاحات), Education (تعلیم), Health (صحت)
Defense Spending	War (جنگ), Armed Forces (افواج), Defense (دفاع), Enemy (دشمن), Terrorism (دہشت)
Economic Policy	Financial (مالی), Policy (پالیسی), Crisis (بحران), Economic (اقتصادی), Livelihood (معیشت)
Corruption & Taxation	Corruption (کرپشن), Mafia (مافیا), Wealth (مال), Thieves (چوروں), Taxes (ٹیکس)
Media Coverage	Press (پریس), Report (رپورٹ), Media (میڈیا), Journalist (صحافی), News (نیوز)
Regional Policy	Regions (اضلاع), Areas (علاقوں), Sindh (سندھ), Provinces (صوبے), Limited (محدود)
Politics	Pakistan (پاکستان), Cabinet (کابینہ), State (ریاست), Assembly (اسمبلی), Politics (سیاست)

Table 1: Aspect seed words on tweets of budget topic

Method	Accuracy	Precision	Recall	macro-F1	Dataset
JASen	83.83	64.73	72.95	66.28	Restaurant reviews
JASen	71.01	69.55	71.31	69.69	Laptop reviews
JASen	54.6	46.63	44.19	41.33	Budget tweets
JASen w/BiLSTM	70.3	70.9	71.4	68.7	Laptop reviews
JASen w/BiLSTM	64.8	43.5	45.5	48.8	Budget tweets

Table 2: Aspect identification results (%): English Reviews vs. Urdu Tweets

Method	Accuracy	Precision	Recall	macro-F1	Dataset
JASen	81.96	82.85	78.11	79.44	Restaurant reviews
JASen	74.59	74.69	74.65	74.59	Laptop reviews
JASen	54.2	46.41	50.00	42.14	Budget tweets
JASen w/BiLSTM	75.2	75.7	75.4	75.2	Laptop reviews
JASen w/BiLSTM	72.8	54.6	51.3	52.4	Budget tweets

Table 3: Sentiment identification results (%): English Reviews vs. Urdu Tweets

4 Conclusion

We propose an approach to introduce joint aspect-based sentiment analysis in the Urdu language using advanced techniques with minimal guidance. The weakly-supervised JASen model, originally developed for English, is applied to Urdu tweets to capture fine-grained information. Fasttext embeddings and the vanilla settings of the model are leveraged to establish a benchmark, followed by experiments with modifications in the model architecture and default settings. Extensive pre-processing is performed to prepare the dataset due to the complexity and informality of the language used in tweets, making it compatible with model training and resource utilization. The results demonstrate improvements in the model's performance on the Urdu dataset, attributed to the enhancements in pre-processing and model architecture modifications.

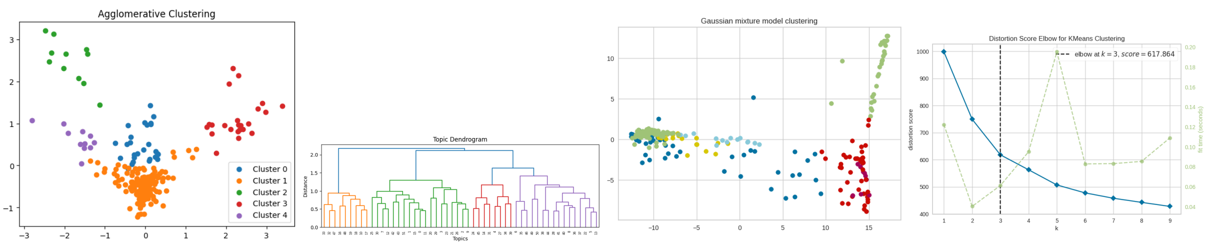

Figure 3: Inspecting optimal number of clusters

	Govt. Spending	Subsidy & Grant	Social Welfare	Defense	Economy	Corruption & Tax.	Media	Regional Policy	Politics
Positive	روزگاری اعتماد اشرافیہ انتخاب ایمانداروں Expenditure Elite Confidence Trustworthy Elections	منصوبہ جامع بندی صنعتوں مختص اسکیموں Comprehensive Projects Industry Specific Schemes	معیار ضروریات کسان ترجیحات اضافے Standards Essential for Farmers Additional Preferences	قربان مضبوط پڑوسی قوت سلامتی Strong Sacrifice Neighbors Health Power	معاشی تجارتی بہتری استحکام موزوں اقدامات Economic Trade Better Strengthening Appropriate Initiatives	نیاٹیکس بزنس فرینڈلی سرپرائز ورثے New tax Business Friendly Surprises Legacies	تبصرے تجزیے ماہرانہ اینکرز معلومات Comments Expert Analysis Anchors' Knowledge	سہولیات کالج اسکولز پختون مطالبہ College Facilities Schools of Khyber Demand	پارلیمانی ووٹنگ بجٹ اتحادیوں اعتماد Parliamentary Voting Coalition Budget Confidence
Negative	خودکشی گزارہ کٹھپتلی ظالمانہ تنگ Suicidal Brutal Oppression Narrow	اندھابانٹے ہولڈنگ بیکار مذموم مسخرے Indebted Holdings Useless Condemned Mockeries	بیرزگاری پیناڈول کارتوس نقصانات غربت Poverty Pinadole Cartouches Losses Poverty	ماتم خطرہ حملہ دھمکی موت Mourning Threat Attack Threat Death	دیوالیہ بدحالی بیروزگاری سود سختیاں Disaster Desolation Unemployment Profit Hardships	نقصان آٹا ذخیرہ سکینڈل ناقص Flawed Flour Hoarding Scandal Incomplete	جوائنٹ سربراھی شکار مذمت تقریر Joint Leadership Condemned Hunting Speech	غیرمناسب ناانصافی کٹوتی قبائلی دشمنی Inappropriate Unfair Katoori Tribal Enmity	گونج خطرات دستور حکومتی اپوزیشن Roar Threats Government Orders Opposition

Table 4: Aspect terms retrieved by joint topics

Tweet	**Ground Truth**	**Prediction**
انشاء اللہ بجٹ عوام دوست خان امید (Inshallah budget people's friend Khan hope)	(politics, pos)	(govt. spending, pos)
عوام دشمن بجٹ غریبوں محنت کشوں سرکاری ملازمین (People's enemy budget poor laborers government employees)	(govt. spending, neg)	(govt.spending, neg)
کٹھ پتلی حکومت کٹھ پتلی بجٹ نامنظور (Puppet government Puppet budget unacceptable)	(politics, neg)	(politics, neg)
عوام معاشی قتل عام پاکستان بدترین انسان دشمن وفاقی بجٹ (People economic massacre common Pakistan worse human enemy federal budget)	(economy, neg)	(politics, neg)

Table 5: Comparison of Model Predictions with Ground Truth

In future work, we intend to conduct additional experiments involving variations in hyper-parameter settings, embeddings, seeded knowledge, and architectural changes to further explore advanced approaches for ABSA in the Urdu language. We also aim to identify any limitations these approaches may have when applied to Urdu. Another promising direction is the development of annotated datasets to leverage state-of-the-art deep learning methods and enhance the performance of ABSA in Urdu.

Limitations

Our experimentation focused only on the "budget" topic in the ABSA domain, leaving the other topics largely unexplored with a large dataset. Additionally, we did not explore the use of other embeddings such as sentence BERT embeddings, transformer architectures with multi-head attention, and pre-trained language models with varying hyper-parameter settings. Furthermore, the current methodology does not enable the prediction of multiple aspects in a sentence along with their associated sentiments.

References

Ikram ALi. 2020. Urduhack: A python library for urdu language processing. https://docs.urduhack.com/en/stable/#urduhack.

Maaz Amjad, Noman Ashraf, Alisa Zhila, Grigori Sidorov, Arkaitz Zubiaga, and Alexander Gelbukh. 2021. Threatening language detection and target identification in urdu tweets. *IEEE Access*, 9:128302–128313.

Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 340–350.

Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. 2020. Synchronous double-channel recurrent network for aspect-opinion pair extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 6515–6524.

Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceed-*

ings of the AAAI conference on artificial intelligence, volume 35, pages 12666–12674.

Zhuang Chen and Tieyun Qian. 2020. Enhancing aspect term extraction with soft prototypes. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2107–2117, Online. Association for Computational Linguistics.

Lei Gao, Yulong Wang, Tongcun Liu, Jingyu Wang, Lei Zhang, and Jianxin Liao. 2021. Question-driven span labeling model for aspect–opinion pair extraction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 12875–12883.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017a. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 388–397.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2017b. An unsupervised neural attention model for aspect extraction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 388–397, Vancouver, Canada. Association for Computational Linguistics.

Fatemeh Hemmatian and Mohammad Karim Sohrabi. 2019. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial intelligence review*, 52(3):1495–1545.

Mengting Hu, Shiwan Zhao, Honglei Guo, Chao Xue, Hang Gao, Tiegang Gao, Renhong Cheng, and Zhong Su. 2021. Multi-label few-shot learning for aspect category detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6330–6340, Online. Association for Computational Linguistics.

Mengting Hu, Shiwan Zhao, Li Zhang, Keke Cai, Zhong Su, Renhong Cheng, and Xiaowei Shen. 2019. CAN: Constrained attention networks for multi-aspect sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4601–4610, Hong Kong, China. Association for Computational Linguistics.

Jiaxin Huang, Yu Meng, Fang Guo, Heng Ji, and Jiawei Han. 2020. Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6989–6999, Online. Association for Computational Linguistics.

Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. 2019. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78:15169–15211.

Asad Khattak, Muhammad Zubair Asghar, Anam Saeed, Ibrahim A Hameed, Syed Asif Hassan, and Shakeel Ahmad. 2021. A survey on sentiment analysis in urdu: A resource-poor language. *Egyptian Informatics Journal*, 22(1):53–74.

Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6714–6721.

Xin Li and Wai Lam. 2017. Deep multi-task learning for aspect term extraction with memory interaction. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2886–2892.

Yunlong Liang, Fandong Meng, Jinchao Zhang, Yufeng Chen, Jinan Xu, and Jie Zhou. 2021. An iterative multi-knowledge transfer network for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1768–1780, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Haoyue Liu, Ishani Chatterjee, MengChu Zhou, Xiaoyu Sean Lu, and Abdullah Abusorrah. 2020. Aspect-based sentiment analysis: A survey of deep learning methods. *IEEE Transactions on Computational Social Systems*, 7(6):1358–1375.

Ling Luo, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, and Dong Yu. 2019. Unsupervised neural aspect extraction with sememes. In *IJCAI*, pages 5123–5129.

Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 13543–13551.

Margaret Mitchell, Jacqui Aguilar, Theresa Wilson, and Benjamin Van Durme. 2013. Open domain targeted sentiment. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1643–1654.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8600–8607.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

Sadaf Rani and Waqas Anwar. 2020. Resource creation and evaluation of aspect based sentiment analysis in urdu. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop*, pages 79–84.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Yuli Vasiliev. 2020. *Natural Language Processing with Python and SpaCy: A Practical Introduction*. No Starch Press.

Qianlong Wang, Zhiyuan Wen, Qin Zhao, Min Yang, and Ruifeng Xu. 2021. Progressive self-training with discriminator for aspect term extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 257–268, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. Recursive neural conditional random fields for aspect-based sentiment analysis. page 616626.

Meixi Wu, Wenya Wang, and Sinno Jialin Pan. 2020. Deep weighted maxsat for aspect-based opinion extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5618–5628.

Xiaoyu Xing, Zhijing Jin, Di Jin, Bingning Wang, Qi Zhang, and Xuanjing Huang. 2020. Tasty burgers, soggy fries: Probing aspect robustness in aspect-based sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3594–3605, Online. Association for Computational Linguistics.

Yunyi Yang, Kun Li, Xiaojun Quan, Weizhou Shen, and Qinliang Su. 2020. Constituency lattice encoding for aspect term extraction. In *Proceedings of the 28th international conference on computational linguistics*, pages 844–855.

Yichun Yin, Chenguang Wang, and Ming Zhang. 2020. PoD: Positional dependency-based word embedding for aspect term extraction. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1714–1719, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136.

Chen Zhang, Qiuchi Li, Dawei Song, and Benyou Wang. 2020. A multi-task learning framework for opinion triplet extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 819–828, Online. Association for Computational Linguistics.

Wenxuan Zhang, Ruidan He, Haiyun Peng, Lidong Bing, and Wai Lam. 2021. Cross-lingual aspect-based sentiment analysis with aspect term code-switching. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9220–9230.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3239–3248.

Jie Zhou, Jimmy Xiangji Huang, Qin Chen, Qinmin Vivian Hu, Tingting Wang, and Liang He. 2019. Deep learning for aspect-level sentiment classification: survey, vision, and challenges. *IEEE access*, 7:78454–78483.