PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

UNIVERSITY OF AHMED DRAIA ADRAR

FACULTY OF SCIENCES AND TECHNOLOGY

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

A Dissertation Submitted in Partial Fulfillment of the Requirement for Master's Degree in Computer Science

Aspect-based Sentiment Classification Model Employing Dialect Normalization and Deep Learning

Master Candidates	:	Chennafi Mohammed Elamine Bedlaoui Hanane
Major	:	Intelligent Systems
Supervisor	:	Dr. Abdelghani Dahou

May 2022

الجمهورية الجزائرية الديمقراطية الشعبية

People's Democratic Republic of Alge[illegible]

Ministry of Higher Education and Scientific Research
University Ahmed Draia of Adrar
The central library

وزارة التعليم العالي و البحث العلمي
جامعة أحمد دراية- أدرار
المكتبة المركزية
مصلحة البحث البيبليوغرافي

شهادة الترخيص بالإيداع

انا الأستاذ(ة) : دحو عبد العالي

المشرف مذكرة الماستر الموسومة بـ : Aspect-based Sentiment classification Model Employing Dialect Normalization and Deep Learning

من إنجاز الطالب(ة): شناني محمد الأمين

و الطالب(ة): بدلاوي حنان

كلية : علوم وتكنولوجيا

القسم : الرياضيات والاعلام الآلي

التخصص: أنظمة ذكية

تاريخ تقييم / مناقشة: جوان 2022

أشهد ان الطلبة قد قاموا بالتعديلات والتصحيحات المطلوبة من طرف لجنة التقييم / المناقشة، وان المطابقة بين النسخة الورقية والإلكترونية استوفت جميع شروطها.

وبإمكانهم إيداع النسخ الورقية (02) والاليكترونية (PDF).

امضاء المشرف:

ادرار في : 14 JUIL. 2022

مساعد رئيس القسم:

أ. كحيلي محمد
رئيس قسم الرياضيات والإعلام الآلي

وزارة التعليم العالي والبحث العلمي
قسم الرياضيات والإعلام الآلي
جامعة أدرار كلية العلوم والتكنولوجيا

ملاحظة :لاتقبل أي شهادة بدون التوقيع والمصادقة.

Dedication

This thesis is proudly dedicated to
All my beloved family
My mother, my sister Ikram, my brothers, and my father for their love and support all along my academic years.
Great debt of gratitude to my partner Chennafi Mohammed ElAmine, who has faith in me and supported me all the way.
Thank you for your ongoing love, support, and inspiration.

Bedlaoui Hanane.

To my mother and father, who have supported me throughout this journey with passion and devotion..
To my brothers and sisters, who have always been there for me during the tough times.
To my friend Hanane who made me improve my skills.
I dedicate this achievement.

Chennafi Mohammed ElAmine.

Acknowledgment

"Give glory to God, who can alter all fates".

Our first and big appreciation goes to our supervisor Dr Abdelghani Dahou, for his guidance, marvelous supervision, and encouragement. His significant engagement in mentoring, kind support, constructive critique, and advice throughout our Masters degree trip are gratefully acknowledged. He is the epitome of what it is to be a leader and a role model. Thank you very much teacher.

Also, we extend our thanks to all our teachers of the mathematics and computer science department for their kind support during our university education, as well as to our friend and classmate Elmasri Tayeb for his continuous encouragement, advice and moral support.

Last but not least, we want to express our heartfelt gratitude to our families for their unwavering support and hope. This thesis would not have been completed without such hope. Thank you all.

Chennafi Mohammed ElAmine.

Bedlaoui Hanane.

Abstract

Sentiment analysis (SA) is one of the most important fields of natural language processing (NLP) due to its wide range of applications and the benefits associated with using it. It is defined as the process of identifying the sentiment polarity (positive, negative, or neutral) of a natural language given text. Researchers have recently focused their attention on Arabic SA due to the massive amounts of user-generated content on social media and e-commerce websites in the Arab world. Most of the research in this fieldwork is on the sentence and document levels. This study tackles the aspect-level sentiment analysis for the Arabic language, which is a less studied version of SA. Because Arabic NLP is challenging, few Arabic annotated corpora available, and the massive number of Arabic dialects, limited studies have attempted to detect aspect-based sentiment analysis on Arabic texts. Specifically, this study considers two ABSA tasks: aspect term polarity and aspect category polarity, and using the text normalization of the Arabic dialect after making the classification task. We present a Seq2Seq model for dialect normalization that can serve as a pre-processing step for the ABSA classification task by reducing the number of OOV words. Thus, the model's accuracy increased. The results of the conducted experiments show that our models outperformed the existing models in literature on both tasks and datasets.

keywords : sentiment analisys , aspect based sentiment analisys , dialect normalization , aspect term polarity, aspect category polaity .

Abstract in Arabic

يعد تحليل المشاعر أحد أهم مجالات معالجة اللغة الطبيعية نظرًا لتطبيقاته الواسعة النطاق والفوائد المرتبطة باستخدامه. يتم تعريفه على أنه عملية تحديد قطبية المشاعر (إيجابية أو سلبية أو محايدة) لنص معطى بلغة طبيعية. ركز الباحثون مؤخرًا على الاهتمام بتحليل المشاعر العربية بسبب الكميات الهائلة من المحتويات التي ينشئها المستخدمون على مواقع التواصل الاجتماعي ومواقع التجارة الإلكترونية في العالم العربي. تعمل معظم الأبحاث في هذا المجال على مستويي الجملة والوثيقة. تتناول هذه الدراسة تحليل المشاعر على مستوى الجانب للغة العربية وهي النسخة الأقل دراسة في مجال تحليل المشاعر. نظرًا لأن البرمجة اللغوية العصبية للغة العربية يمثل تحديًا والعدد القليل من مجموعة البيانات العربية المتاحة والعدد الهائل من اللهجات العربية ، فقد حاولت دراسات محدودة تناول موضوع تحليل المشاعر القائم على الجوانب في النصوص العربية. على وجه التحديد ، تعالج هذه الدراسة في مهمتين من مهام تحليل المشاعر القائم على الجانب: تحديد قطبية مصطلح الجانب و تحديد قطبية فئة الجانب ، وتستفيد من تطبيع النص في اللهجة العربية بعد إجراء مهمة التصنيف. بتعبير أدق ، نقدم نموذج تطبيع اللهجات والذي يمكن أن يكون بمثابة خطوة معالجة مسبقة لمهمة تصنيف تحليل المشاعر القائم على الجانب عن طريق تقليل عدد الكلمات الخارجة عن قاموس مفردات النموذج وبالتالي زيادة دقة النماذج الخاصة بنا. تظهر النتائج المحققة للتجارب التي تم إجراؤها أن نماذجنا تفوقت على النماذج الموجودة في الدراسات السابقة في كل من المهام ومجموعات البيانات.

الكلمات المفتاحية : تحليل المشاعر ،تحليل المشاعر القائم على الجانب ، تطبيع اللهجة ،تحديد قطبية مصطلح الجانب ،تحديد مصطلح فئة الجانب .

List of Contents

List of Figures

List of Tables

List of Abbreviations

NLP	Natural Language Processing
SA	Sentiment Analysis
ABSA	Aspect-based Sentiment Analysis
DA	Dialectal Arabic
MSA	Modern Standard Arabic
ML	Machine Learning
OOV	Out-of-vocabulary
BERT	Bidirectional Encoder Representation from Transformer
ACSA	Aspect category sentiment analysis
ATSA	Aspect-term sentiment analysis
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-term Memory
SVM	Support Vector Machines
MUSE	Multilingual Universal Sentence Encode
CRF	Conditional Random Field
GRU	Gated Recurrent Network
BiLSTM	Bidirectional Long Short-term Memory
E2E-ABSA	End-to-End Aspect-based sentiment analysis
SPD	Sentiment polarity detection
AE	Aspect Extraction
ASPC	Aspect sentiment polarity classification
AOTE	Aspect opinion target extraction
Seq2seq	Sequence-to-sequence
MLM	Masked language modeling
NSP	next sentence prediction
NLI	Natural language inference
QA	Question answering
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
MADAR	Multi-Arabic Dialect Applications and Resources
PADIC	Parallel Arabic DIalectal Corpus
HAAD	Human Annotated Arabic Dataset
SemEval-2016	Semantic Evaluation Workshop 2016
ALG	Algiers
ANB	Annaba
TUN	Tunisia
PAL	Palestine

SYR	Syria
MAR	Morocco
CAMeL	Computation approaches for modeling languages

Chapter 1: Introduction

The world knows a significant development of web sectors and a growing interest in social media in the current years. Hence, users produce and generate a considerable amount of data each day. This user-generated content includes essential information about people's opinions on multiple topics. Therefore, there was a growing need to understand human opinions and feelings to make it easy the process of decision-making about products, places, and services. Sentiment analysis (SA) or opinion mining is a subfield of natural language processing (NLP) that considers identifying the sentiment of a given text automatically as it is positive, negative, or neutral [1].

One of the SA classification levels is aspect-based. Aspect-based sentiment analysis (ABSA) is a more fine-grained and complex task than SA; it is concerned with determining the aspect terms presented in a document, as well as the sentiment expressed against each term [2].

, For instance, Figure 1.1 illustrates an example restaurant review with basically two different target aspects and their related sentiment expressions. As shown, the first target aspect, LOCATION, expressed by the aspect expression "view of the river," has a positive polarity. In contrast, the second target aspect, "FOOD," represented by the aspect expression "sushi rolls," has a negative polarity.

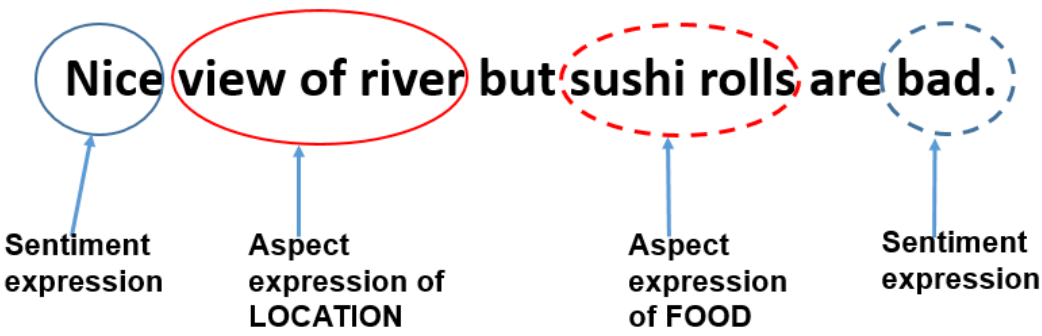

Figure 1.1: An aspect-based sentiment analysis example. The aspect expressions "view of river" and "sushi rolls" correspond to the sentiment expressions "nice" and "bad."

As mentioned in [3], there are four primary tasks of ABSA that can be identified: Aspect Term Extraction (T1), Aspect Term Polarity (T2), Aspect Category Identification (T3), and Aspect Category Polarity (T4). In this work, we concentrate on tasks (T2) and (T4).

Recently, the field of Arabic NLP has become more interesting and challenging. The Arabic language is the original language for twenty-two (22) countries across Asia and Africa, spoken by 300 million people. It is an official language from the six of the United Nations [4]. Two types of Arabic can be considered: Modern Standard Arabic (MSA) and dialects (vernaculars). MSA originated from Classical Arabic and is utilized throughout the Arab world in education, media, literature, official documents, and old and new books. Before the mid-1990s, the only documented versions of Arabic were classical Arabic and MSA. After that, the documentation of different Arabic dialects was pushed due to the widespread use of internet services and mobile applications. In addition, a new version of Arabic known as Arabizi has been designed in which Arabic characters have been replaced with the Roman alphabet [5]. The Arabic dialects are used in the daily informal communications between people who live in the same country. Linguists have classified the Arabic vernaculars into seven main regional groups, which are Egyptian, Maghrebi, Mesopotamian, Sudanese, Arabian Peninsula, Andalusian, and Levantine [5].

From the main challenges and problems knowns by Arabic SA is the ambiguity and complexity of Arabic morphology and orthography. The Arabic language has a set of morphemes like affixes, prefixes, and suffixes, which express linguistic features such as a person, instance, and gender. Furthermore, Arabic has a set of 16 morphemes, which adds to the amount of ambiguity caused by diverse meanings of the same word, making Arabic text processing more difficult. As illustrated in Figure 1.2, the term of (فسيرونها) has many morphemes that express an English sentence, "and they will see it" [6] [4].

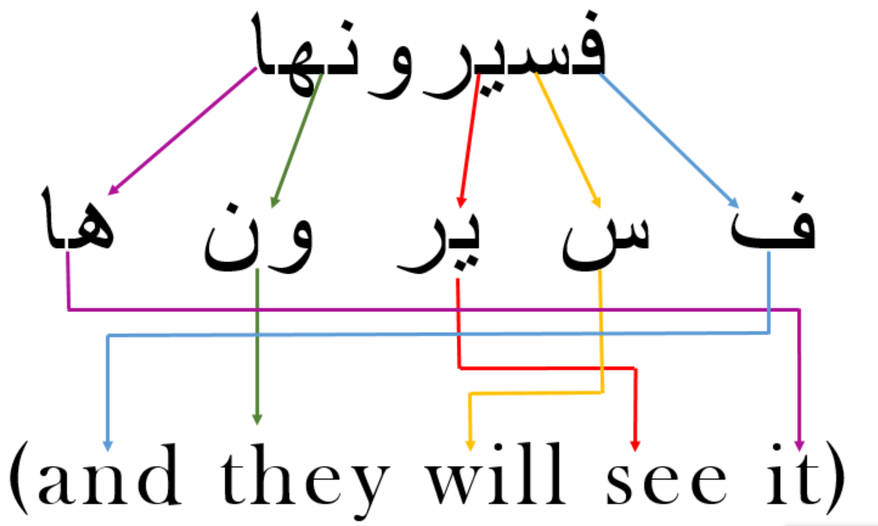

Figure 1.2: Example of the ambiguity of Arabic words morphology.

Orthographically, the lack of diacritical markings in the text creates lexical ambiguity, which is a challenging problem for computer systems [4]. For instance, the undiacritized word (درس) may have several meanings like "lesson," "study," and "taught." Moreover, the massive number of Arabic dialects and the unavailability of dialectical Arabic (DA) language resources are well-known problems that lead to the lack of training dataset, making the research in this field more complicated. Consequently, these dialectical words are considered out of vocabulary words (OOV) in many pre-trained language models.

Here comes the role of the normalization phase that removes the out of vocabulary words (OOV) or the dialectical word and replaces them with a suitable form of Modern

Standard Arabic (MSA) in order to improve the performance of the aspect-based level.

Our study adopted a sequence-to-sequence model for text normalization that aims to transform dialect into MSA. In addition to that, we adopted two aspect-based sentiment analysis models that rely on the pre-trained Bidirectional Encoder Representation from Transformers (BERT) language model for Arabic to perform aspect-based classification for both tasks aspect category polarity and aspect sentiment classification.

The key contributions of our work are:

- We proposed a new solution that improves the results of ABSA by converting the dialectical text into MSA using text normalization.
- We used a pre-trained based model (BERT) with sentence pair input to solve the Arabic ABSA classification task, including MSA and dialect, instead of traditional Machine learning (ML) algorithms.
- We adopt a sequence-to-sequence model for normalizing out-of-vocabulary (OOV) words from our dataset.
- We pre-processed and built a training dataset for the normalization model using two well-known datasets.

The rest of the research is given as follows. Chapter 2 addresses the recent related work on ABSA. In chapter 3, we explained the methodology used in our research; Section 4 discusses the results of our experiments. Finally, chapter 5 presents the study's conclusion and future work.

Chapter 2: Related Works

ABSA's research methodologies are classified into traditional machine learning techniques and deep learning techniques. ABSA's early efforts were mostly based on deep learning techniques. In this chapter, we will mention some of the most recent works tackling English and Arabic ABSA.

2.1 English aspect-based sentiment analysis:

Xue et al. [7] demonstrated by their extensive experiments on SemEval datasets provided by [8], the improved performance of their efficient Gated Tanh-ReLU Units (GTRU) with gating mechanisms model compared to other neural models. The proposed model controls the sentiment flow according to the aspect information for aspect-category sentiment analysis (ACSA) and aspect-term sentiment analysis (ATSA) tasks. In addition, The aspect and sentiment information is modeled separately by two conventional layers.

Liu et al. [9] introduced a Gated Alternate Neural Network model (GANN), a new type of neural network architecture. Also, a GTR (Gate Truncation RNN) module is used for learning good aspect-dependent sentiment clue representations. Convolution and pooling algorithms produce more exact representations by using a gating mechanism to control information and extract key local sentiment clue features and get the position invariance of features. They experimented with their models using four Chinese and three English datasets and achieved the best results.

Li et al. [10] conducted experiments on two benchmark datasets originating from SemEval [8] and re-prepared in [11] to assess the efficiency of the BERT embedding component on the End-to-End Aspect-Based Sentiment Analysis task (E2EABSA). The experimental results illustrate that their BERT-based model outperforms state-of-the-art works.

Xu et al. [12] improved the accuracy of both two tasks, aspect term sentiment analysis (ATSA) and aspect category sentiment analysis (ACSA), by proposing a Recurrent Neural Network model with Target Embedding (RTE) using the target enhance technique. Experiments are conducted on the SemEval workshop and a Twitter dataset. The proposed RTE model achieved the best results compared to state-of-the-art models.

Trueman et al. [13] evaluated their model based on convolutional stacked bidirectional long short-term memory with a multiplicative attention mechanism using SemEval-2015 and SemEval-2016 datasets on both aspect category and sentiment polarity detection tasks.

2.2 Arabic Aspect based sentiment analysis:

The number of works available in the Arabic ABSA is still restricted due to the language complexity.

Abdelgwad et al. [14] experimented with the modeling capabilities of contextual embeddings from the pre-trained BERT model on the task of Arabic aspect sentiment polarity classification with the use of sentence pair input. Three different Arabic datasets have been used including HAAD [3] , Arabic News dataset [15], Arabic Hotel Reviews datasets [8]. Their model had the best accuracy of 89.51 percent on the Arabic hotel reviews dataset.

Abdelgwad et al. [2] addressed aspect opinion target extraction (T2) and aspect polarity detection (T3) tasks by using two recurrent gated units (GRU) based models. They used the benchmark Arabic hotel reviews dataset [8]. The proposed methods achieved the best results on both tasks, with an F1 score of 70.67% for Task 2 and an accuracy of 83.98% for Task 3.

Ashi et al. [16] compared two words embedding models fastText Arabic Wikipedia and AraVec-Web for the two tasks aspect detection followed by sentiment polarity classification of the detected aspects. They used a corpus of 5K airline service-related tweets in Arabic. A support vector machine (SVM) classifier has been used for both classification tasks. They found that the fastText Arabic Wikipedia word embeddings model performed best with an accuracy of 70% for aspect detection and 89% for sentiment polarity classification of the detected aspects.

Al-Dabet et al. [17] presented two deep learning models to cover both ABSA's tasks aspect-category identification and aspect-sentiment classification. The first one is based on CNN and the stacked independent LSTM model. The second model comprises multiple layers of stacked bidirectional independent LSTMs, a position-weighting mechanism, and multiple attention mechanisms. They used the Arabic SemEval-2016 dataset for the Hotels domain was used. The two models achieved the best results, with an F1 score of 58.08% for the first model and accuracy of 87.31% for the second model.

Mohammad et al. [18] developed a deep learning model (Pooled-GRU) based on Gated Recurrent Units (GRU) and features extracted to handle two ABSA tasks: (1) aspect extraction and (2) aspect polarity classification by using Multilingual Universal Sentence Encoder (MUSE). They employed the Arabic hotel reviews dataset and got the best results in both tasks, achieving an F1 score of 93.0% in the first task and 90.86% in the second. Al-Smadi et al. [19] implemented two based LSTM neural network models. the first is a character-level bidirectional LSTM with conditional random field classifier (Bi-LSTM-CRF) for aspect opinion target expressions (OTEs) extraction and the second is aspect-based LSTM for aspect sentiment polarity classification. The experiments conducted on Arabic Hotels' reviews enhanced both tasks by 39% for aspect-OTEs extraction and 6% for aspect sentiment polarity classification.

Al-Smadi et al. [20] implemented and trained two approaches of RNN and SVM along with the lexical, syntactic, word, morphological, and semantic features and evaluated using Arabic Hotel reviews dataset. Experiment results show that the SVM approach is better than RNN in the three ABSA tasks, whereas deep RNN execution time for training and testing was faster.

Table 2.1 summarizes the given overview of the recent related works to English and Arabic aspect-based sentiment analysis.

Table 2.1: Summary of the related works.

Paper	Task	Model	Dataset
Aspect-based sentiment analysis in English			
Xue et al. [7]	ACSA and ATSA	CNN and gating mechanism	SemEval2014 datasets
Liu et al. [9]	ABSA	Gated Alternate Neural Network (GANN)	SemEval2014,four chinese and Tweeter dataset
Li et al. [10]	Aspect term (E2E-ABSA)	BERT	Two review datasets from SemEval
Xu et al. [12]	ATSA and ACSA	RNN and Target embedding	SemEval and a twitter dataset
Trueman et al. [13]	ACSA and SPD	convolutional stacked bidirectional LSTM and attention mechanism	SemEval-2015 and SemEval-2016
Aspect-based sentiment analysis in Arabic			
Abdelgwad et al. [14]	ASPC	Pre-trained model BERT	HAAD, Arabic News and Arabic Hotel Reviews datasets
Abdelgwad et al. [2]	ASPC and AOTE	GRU and CNN	Arabic hotel reviews dataset
Ashi et al. [16]	AE and ASPC	Word embedding	Arabic airline-related tweets
Al-Dabet et al. [17]	OTE extraction and ASPC	CNN and LSTM	Arabic SemEval-2016 dataset
Mohammad et al. [18]	AE and ASPC	Gated Recurrent Units (GRU)	Arabic hotel reviews
Al-Smadi et al. [19]	Aspect OTE and ASPC	Long short-term memory (LSTM)	Arabic Hotels' reviews
Al-Smadi et al. [20]	ACSA, aspect OTE extraction and ASPC	RNN and SVM	Arabic Hotels' reviews

Chapter 3: Methodology

In this chapter, a description of the used approaches will be given. Also, the proposed models of both normalization and ABSA tasks will be explained in detail. The following simple diagram illustrates our work:

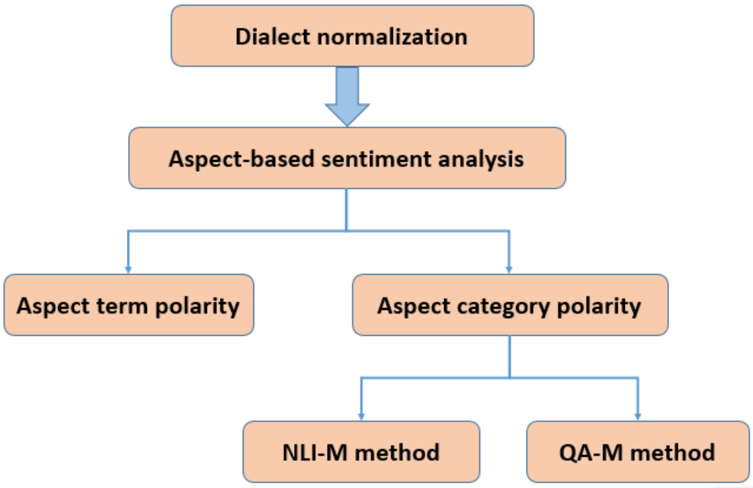

Figure 3.1: A simple diagram of the proposed work.

3.1 Sequence-to-sequence learning

The transformation of one sequence into another is involved in many ML tasks. Sequence-to-sequence models, also known as seq2seq models, commonly represent such tasks in deep learning. Sequence to sequence learning is the process of training models to transform sequences from one domain to another domain sequences. This model is based on the encoder-decoder architecture to produce a sequence output for a sequence input. Recurrent neural networks (RNN) are used in both the encoder and the decoder to handle sequence inputs. The encoder's hidden state is directly used to initialize the decoder's hidden state in order to pass information from the encoder to the decoder. This can be used for free-form question answering [21], text normalization [22], text summarization [23], and machine translation [24]. In general, it can be used whenever you need to generate text. With mechanisms like attention [25], coverage [26], and copying [27], several extensions of the Seq2Seq models have been proposed.[28]

3.1.1 Encoder-Decoder Architecture

The Encoder-Decoder model is divided into two components: first, an encoder that takes an input $\mathcal{X}$ (in this example, a phrase) and generates an intermediate representation $\mathcal{Z}$ (or code) that emphasizes its key features; and second, a decoder that processes that collection of features and generates the needed output $\mathcal{Y}$. (in this case, a normalized phrase). $\mathcal{Z}$ is a matrix with the dimensions $\mathcal{Z} \in \mathcal{M}_{f*l}$, where f is the number of features to encode for each input value [22]. This model's fundamental diagram is shown in Figure 3.2.

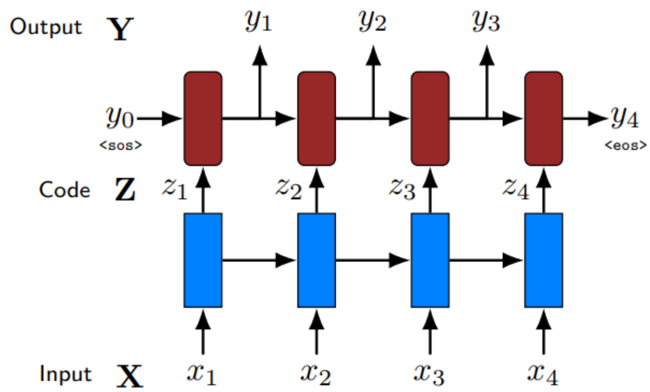

Figure 3.2: An encoder-decoder architecture.

This figure illustrates an encoder-decoder architecture in its most basic form. The encoder is represented by the color blue. While the Decoder is red. The system generates an output sequence, $Y = y_1, y_2 ... y_n$ from an input sequence, X = $x_1, x_2 .. x_n$. and $Z = z_1, z_2 ... z_n$ is the intermediate code. There are two special symbols: <sos> and <eos>, which stand for the start of sequence and end of the sequence, respectively.

It is a simplified model of the encoder-decoder architecture in which the input, intermediate representation, and output are all the same size. For example, this can be the case with image processing tasks. Nonetheless, they might have different sizes in many NLP applications. In our case, all intermediate code, the input, and the output are always the same length.

3.1.2 Word-level sequence-to-sequence model for text normalization

The text normalization problem is described as a sequence-to-sequence learning problem, in which the goal is to learn how to transfer one sequence to another. Our normalization proposed model is based on a sequence-to-sequence learning framework introduced by [28] which is a hybrid model to address the OOV problem. It consists of two encoder-decoder models; the first is a word-based Seq2Seq model that reads the informal text sequences and transforms them into a continuous-space representation passed on to the decoder to generate the target normalized sequence. The second one, capture local spelling errors and morphological variations of OOV words and correct them.

Accordingly, we adopt the word-based Seq2Seq model to handle our research task, which can serve as a pre-processing step for the ABSA task. We represent the dialectical sentence as input sequence of words $x = [x_1, ..., x_T]$ with length T and MSA as output sequence of words $y = [y_1, ..., y_L]$ with length L. The input sequence x is read and transformed by the encoder module into a context-specific hidden state sequence $h = [h_1, ..., h_T]$. Since our model is bi-directional, two encoders are used; one for forwarding text reading mode and another for backward text reading. The concatenation of the two encoder modules

$$h_t = [g_f(x_t, h_t - 1); g_b(x_t, h_t + 1)] \tag{3.1}$$

yields the final hidden state at time t. The forward and backward encoder units are denoted by g_f and g_b, respectively. Based on the previous word $y_j 1$ and decoder state $s_j 1$, a hidden state sequence $s_j = g_s(s_j 1, y_j 1, c_j)$ is produced by the decoder. Based on the attention mechanism [25], the context vector c_j is calculated as a weighted sum of encoder hidden states. Then, the Softmax classifier predicts each target word. The Sequence-to-sequence model is trained by maximizing the log-likelihood of the training data. Scheduled sampling [29] is often used when calculating the conditional probability in order to avoid the occurrence of compounding errors. Figure 3.3 presents an example of a source (dialect) and target (MSA) pair of sentences for which the seq2seq model helps in appropriately normalizing the content.

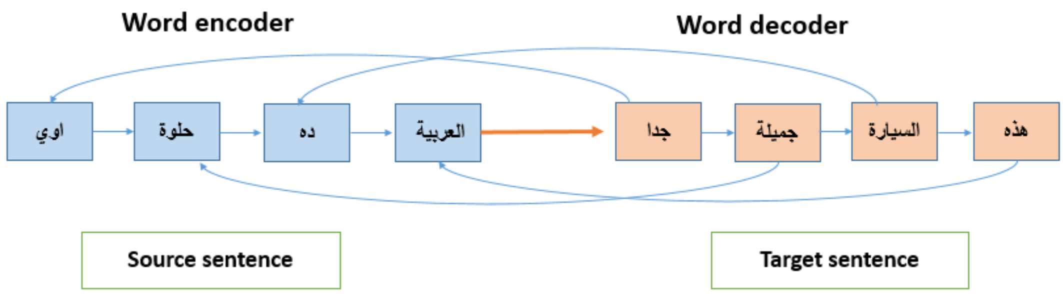

Figure 3.3: Seq2seq model for dialect normalization.

3.2 BERT (Bidirectional Encoder Representations from Transformers)

BERT (Bidirectional Encoder Representations from Transformers) is a natural language processing (NLP) deep learning technique in which deep neural networks use unstructured language representation and bidirectional Transformer-based models. BERT is a pre-trained language model that evaluates both the left and right sides of a word's context simultaneously. [30].

It improves performance in various NLP applications, such as sentiment analysis, machine translation, and question answering systems. Compared to other models like ELMo, which train left and right independently, BERT can extract more context information from a sequence. BERT is pre-trained on two different but related NLP tasks using the bidirectional capabilities: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). MLM's objective is to hide a low-probability word in a sentence. The model replaces a word with the token [MASK] when masked. The model then attempts to predict the masked word using context from both the left and right sides of the masked word with the help of transformers. BERT employs the next sentence prediction (NSP) task in addition to the MLM to train a competent language model that detects and understands sentence links.

BERT is a stack of encoders based on the transformer architecture. A transformer architecture is an encoder-decoder network that uses self-attention on the encoder side and attention on the decoder side. In their paper, Devlin et al. [30] presented the following BERT two architectures:

- **BERT Base:** has 12 layers in the Encoder stack, 12 attention heads, 768 hidden units, and 110M parameters
- **BERT Large:** has 24 layers, 16 attention heads, 1024 hidden units, and 340M parameters.

A preprocessing phase should be conducted before feeding a raw sentence to BERT. The text input to the BERT model is first tokenized using the WordPiece tokenization approach. [24]. This produces a sequence of tokens, each representing a single word. There are two additionally specialized tokens added to the set of tokens: a classifier token $[CLS]$ at the beginning and a separation token $[SEP]$ at the end of a sentence. After that, three separate embedding layers are applied to this set of tokens: Token Embedding, Segment Embedding, and Position Embedding Layer, with the same dimensions, which are afterward added and transmitted to the encoder layer.

3.2.1 Language models

Language models are essential in solving NLP tasks by learning word occurrence and prediction patterns from unannotated text data. They learn the context using techniques like word embedding. The language model is trained on a massive number of training data that representations of words allow related words to have similar representations based on the context. Our experiments utilized AraBERT and ArabicBERT as our base pre-trained language models. AraBERT [31] is a pre-trained Arabic language model built on the BERT base architecture. With extra preprocessing, AraBERT has the same BERT Base configuration. On 8,2 billion words, the ArabicBERT model [32] was pre-trained. The corpus and vocabulary set are not just for MSA; they also include dialectical Arabic. We used araBERTv2 in aspect term polarity and ArabicBERT for aspect category polarity.

3.2.2 Sentence Pair Classifier Task

The goal of the Sentence Pair Classifier is to determine the semantic relationships between two sentences. Due to the BERT model's capabilities and efficiency in dealing with sentence pair classification problems, the ABSA task was converted into a sentence-pair classification task utilizing the pre-trained BERT model. Figure 3.4 represents an architecture of pair sentence classification using BERT, where E is the input embedding for both sentences, Ti is the contextual representation of token i (Tok_i) for both sentences, $[SEP]$ is a special symbol that separates the two sentences, whereas $[CLS]$ is another special token for classification output that sums up all of the hidden state's output from all tokens in the input sentence.

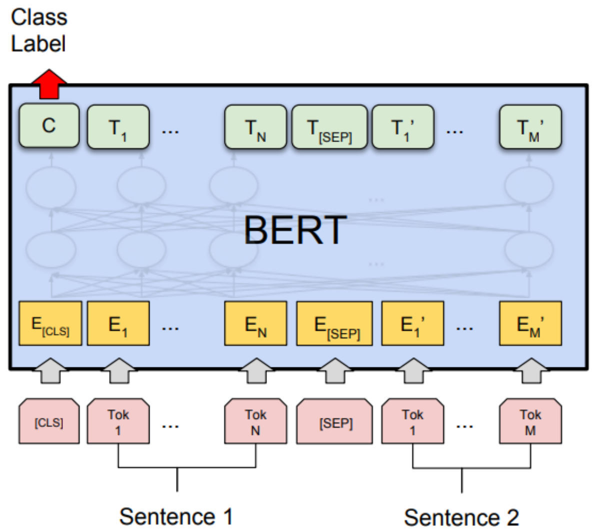

Figure 3.4: Pair sentence classification using BERT.

In our study, we proposed to solve ABSA as a sentence-pair classification task using BERT by constructing auxiliary sentences. Two pre-trained language models BERT are used for both tasks Aspect Category Polarity and Aspect Sentiment Classification.

3.2.3 BERT base model for aspect category polarity via Constructing Auxiliary Sentence

To handle the aspect category polarity task, a BERT model proposed by Sun et al.[33] was adopted with some modifications to deal with our task. An auxiliary sentence is constructed from the aspect to transform aspect category polarity, which aims to determine fine-grained sentiment polarity towards a given category associated with a term, into a sentence-pair classification task, such as question answering (QA) and natural language inference (NLI). The pre-trained BERT model was fine-tuned and evaluated on the HAAD task dataset for Aspect Category Polarity. The final hidden state or (output of the transformer) of the first token is used as input to obtain a fixed-dimensional pooled representation of the input sequence. A classification layer whose parameter matrix is $\mathcal{W} \in \mathcal{R}^{k*H}$ was added where $\mathcal{K}$ represents the number of categories and the vector $\mathcal{C} \in \mathcal{R}^{H}$.The softmax function is used to determine the probability of each category $\mathcal{P}$:

$$\mathcal{P} = softmax(\mathcal{C}\mathcal{W}^{T}) \tag{3.2}$$

3.2.3.1 Construction of the auxiliary sentence

In our work, we used two methods to convert the aspect category polarity task into a sentence pair classification task:

- **Sentences for QA-M:** it refers to the question-answering task. The auxiliary sentence generated from the category is a question, and the question format is

required to be the same with all samples.
As an example," رواية رائعة، انا حسيت وكأني عايش", the category here is "المشاعر" where the generated sentence is "ما رأيك في المشاعر؟"

- **Sentences for NLI-M:** For the natural language inference (NLI) task, the conditions used to generate sentences are less stringent, and the form is considerably more straightforward in that the auxiliary sentence contains only the category of the sentence. For the previous example, the auxiliary sentence formed is: المشاعر.

For each sentence, the related category is polarized as positive, negative, or neutral, with the other categories being labeled as none. Because the number of categories in HAAD dataset is 15, we suggested to randomly selecting three none-labeled categories instead of all categories in order to have a balanced dataset.
Example: S1 = "كتاب يحكي تاريخ الاندلس. رائع بحق"

Table 3.1: An example of HAAD dataset with NLI-M method for S1.

Sentence	Auxiliary sentence	English translation	Sentiment
S1	الاسلوب	Style	Positive
S1	الخاتمة	conclusion	None
S1	السياق	Context	None
S1	الهوامش	Margins	None

Table 3.2: An example of HAAD dataset with QA-M method for S1.

Sentence	Auxiliary sentence	English translation	Sentiment
S1	ما رأيك في الأسلوب ؟	What do you think about the style?	Positive
S1	ما رأيك في المشاعر ؟	What do you think about the sentiments?	None
S1	ما رأيك في الحبكة ؟	What do you think about the plot?	None
S1	ما رأيك في السياق ؟	What do you think about the context?	None

3.2.4 BERT Post-Training model for Aspect Sentiment Classification

In this study, the Aspect sentiment classification was considered as a sentence-pair classification task, so we suggested a BERT-based model proposed by the authors Xu et al. [34] to handle this task, where their model made use of a post-training approach to improving both the domain and task knowledge. The goal of aspect sentiment classification (ASC) is to identify the polarity of sentiment expressed on an aspect retrieved from a review sentence as (positive, negative, or neutral). The aspect and a review sentence mentioning that aspect are the two input sentences to ASC. Then, the output is considered as a class of polarity. The input sequence is formulated as follow: $x = ([CLS]; a_1; , , , ; a_m; [SEP]; s_1; , , , ; s_n; [SEP])$, where $a_1; , , , ; a_m$ denotes the aspect with m tokens and $s_1; , , , ; s_n$ represents the review sentence with n tokens that contains that aspect. The polarity distribution is predicted to be as follows:

$$L_4 = softmax(W_4.h[CLS] + b_4) \tag{3.3}$$

where $h[CLS]$ is the aspect-aware representation of the whole input and $W_4, b_4 \in R^3$ (3 is the number of polarities). On $[CLS]$, the function is applied along the label dimension: $L_4 \in [0, 1]^3$. The introduction of the post-training step to boost the performance of Aspect Sentiment Classification is required due to the insufficient supervised training data significantly which limits the performance gain across the task.

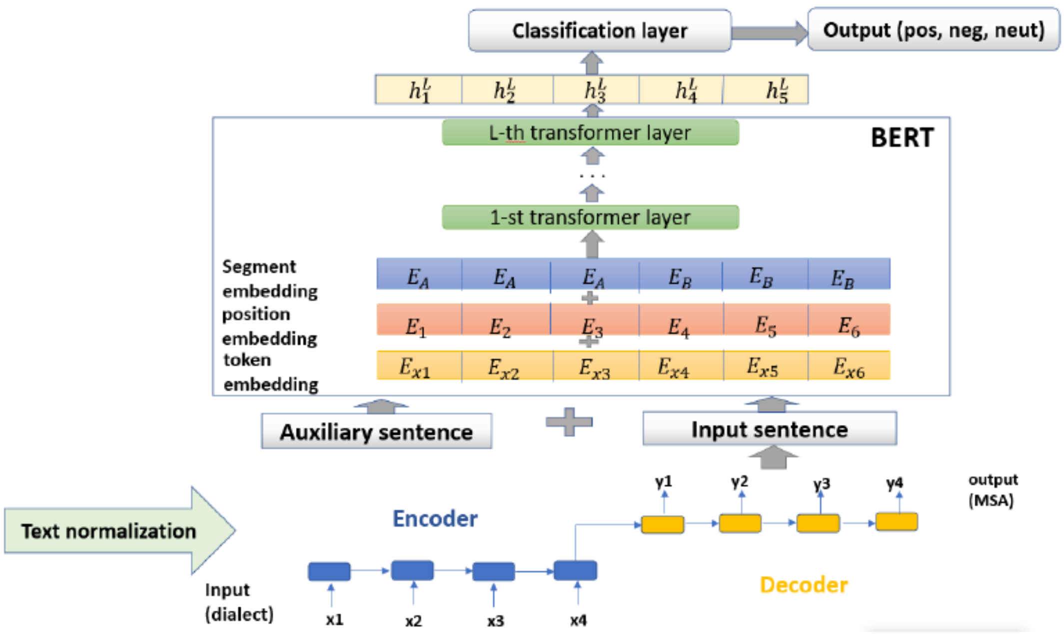

Figure 3.5: The architecture of the proposed model.

Figure 3.5 shows the proposed model architecture. The model consists of two components: a normalization part that contains a sequence to sequence model which converts the dialect into MSA. The input and output of this process should have the same length. The second part is the Aspect-based sentiment analysis which consists of a pre-trained language model BERT that accepts two sentences as input. The first one is the output of the normalization framework (MSA), and the second one is the auxiliary sentence related to the first one. The auxiliary sentence may be the aspect for the aspect sentiment classification task or the category (NLI-M method) or a question about the category (QA-M method) for the aspect category polarity task.

Chapter 4: Results and Discussions

This chapter details the three experiments carried out as part of our research. In the first experiment Ex1, we normalized only the Egyptian samples from HAAD and SemEval-2016 datasets. The normalization framework was trained only on the Cairo dialect from the MADAR dataset and experimented on aspect term polarity and aspect category polarity models. Ex2 provided a training normalization model using the whole MADAR and PADIC dataset and experimented with Egyptian dialect on both ABSA task models in the second experiment. Finally, the last experiment, Ex3, shows the results of using all HAAD and SemEval-2016 datasets with the same trained normalization model of the second experiment on both ABSA's models. Also, a comparison between our model and previous works is made. The chapter includes datasets descriptions, performance measurements, and ABSA model configuration. The following table summarizes our three experiments:

Table 4.1: Experiments summary.

Experiment	Normalization training dataset	Task model	task model training dataset
Ex1	Cairo dialect from MADAR	T2 and T4	HAAD and SemEval with normalized Egyptian samples
Ex2	PADIC and MADAR	T2 and T4	HAAD and SemEval with normalized Egyptian samples
Ex3	PADIC and MADAR	T2 and T4	All normalized samples from HAAD and SemEval

4.1 Hyper-parameters Setting

In our study, we used Pytorch [1] with Hugging Face's Transformers library [35] in all our experiments for the model's implementation and training. The BERT-Base model has been particularly used. The pre-trained language models "ArabicBERT" [32] and "araBERT"[31] were adopted. For the normalization framework, We use Adam optimizer [36] and set the learning rate to be 0.1 and dropout of 0.1 for all experiments, 100 epochs for the first experiment, and 50 epochs for both second and third experiments. On the other hand, for the BERT models, We use the Adam optimizer and set the learning rate to be 3e-5, with a batch size of 32 and 8 for T2 and T4, respectively. Models of T2 and T4 were trained for 5 and 4 epochs, respectively.

[1]https://pytorch.org

4.2 Performance Measures

In order to validate the strength of the proposed model, multiple models are implemented and the results are compared. The accuracy measure was used to assess the efficacy of the proposed framework, which was defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

Where:

- **(TP) True positive:** the values of positive class that are correctly categorized.
- **(TN) True negative:** Samples correctly categorized as negative.
- **(FP) False positive:** incorrectly categorized the negative as positive.
- **(FN) False negative:** misclassified the positive class as negative.

So, accuracy is defined as the number of correct samples of the total number of samples. Better performance is represented by higher accuracy.
The formulas of other performance evaluation metrics are as follow:

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

$$F1 - score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{4.4}$$

4.3 Datasets Description

We conducted our experiments using four Arabic datasets. MADAR corpus [37] is the Multi-Arabic Dialect Applications and Resources, PADIC [38] is the Parallel Arabic DIalectal Corpus, HAAD is the Human Annotated Arabic Dataset [3] and SemEval-2016 [8] is the Semantic Evaluation Workshop 2016. The first two datasets are used for the training normalization model, while the latter two are for T2 and T4. Each dataset will be described in detail in the following subsections.

4.3.1 PADIC Datasets

PADIC is a parallel corpus that includes six Maghreb and Middle Eastern dialects. Modern Standard Arabic (MSA) has been used to align each dialect. PADIC has 6400 sentences for each of the six dialects (Annaba, Algiers, Tunisian, Morocco, Syrian, and Palestinian) and MSA. Since the inputs and outputs in our model are always the same lengths, we will not use all of the datasets but only the one for which the input sentence is the same length as the output sentence. Table 4.2 shows the number of samples corresponding to the condition for each dialect.

```
<Sentence>
    <MSA>دعينا للعشاء و كان العشاء يكفي جميع المدعوين</MSA>
    <ALG>عرضونا للعشا و كان لعشا يكفي ليزانبيتي قاع</ALG>
    <ANB>عرضونا للعشا و كان العشا يكفي ليزانبيتي كل</ANB>
    <TUN>إستدعاونا للعشاء والعشاء كان يكفي الناس الكل</TUN>
    <PAL>انعزمنا عالعشاء و كان العشاء بكفي لكل المعزومين</PAL>
    <SYR>انعزمنا عالعشا و كان العشا بيكفي كل المعازيم</SYR>
    <MAR>تعردنا لعشا و كان لعشا كافي لكلشي المعرودين</MAR>
</Sentence>
```

Figure 4.1: An example of PADIC dataset. English sentence translation is "we were invited to dinner, and dinner was enough for all invited"

figure 4.2 shows an example from a PADIC dataset.

Table 4.2: Selected samples from PADIC dataset.

Dialect	ALG	ANB	TUN	PAL	SYR	MAR
Number of samples	2566	2501	2547	3949	2638	2777

4.3.2 MADAR Dataset

It is a large parallel corpus constructed by translating selected sentences in English and French from the Basic Traveling Expression Corpus (BTEC) into the dialects of 25 Arabic cities and MSA. It contains two corpora: Corpus-26 and Corpus-6:

- **CORPUS-26:** It contains 2000 sentences, and it is translated into all 25 cities' dialects in addition to MSA.
- **CORPUS-6:** It contains 12000 sentences, and it is translated into the dialects of five cities: Cairo, Beirut, Doha, Tunis, and Rabat, in addition to MSA.

As in the PADIC, we will not take all the sentences, but we will only select sentences in which the number of input sequence equals the number of output sequence. The table 4.3 shows the number of samples corresponding to the condition in each corpus.

```
96  corpus-6-train  MSA هل يمكنك أن تخبريني كيف أستطيع أن أصل إلى الأوبرا ؟
97  corpus-6-train  MSA ألم تكن مخمور بدرجة كبيرة بعض الشيء في الليلة الماضية يا عزيزي ؟
98  corpus-6-train  MSA من فضلك اتبع التعليمات الصادرة من موظفينا .
100 corpus-6-dev    MSA هل يوجد هنا محل متعدد الأقسام بالقرب من هنا .
102 corpus-6-train  MSA املأ خزان الوقود من فضلك .
103 corpus-6-train  MSA هل يوجد غرفة بسعر أرخص ؟
104 corpus-6-train  MSA أفضل سيارة بها نظام إلي لنقل الحركة .
105 corpus-6-train  MSA سأقيم في مسكن الطلبة الخاص بجامعة هارفارد .
106 corpus-6-train  MSA ما هو سعر الصرف اليوم ؟
```

Figure 4.2: An example of MADAR dataset. English translation of the first sentence is "can you tell me how to get to the opera"

Table 4.3: Selected samples from MADAR dataset.

corpus	CORPUS-6	CORPUS-26
number of samples	11117	7877

4.3.3 HAAD Dataset

The HAAD dataset is considered the first accessible dataset for Arabic ABSA. The HAAD dataset contains 1513 Arabic book reviews and has 2838 aspect terms. Each review was annotated with tasks T1, T2, T3, and T4. Our study focuses on two tasks, aspect term polarity (T2) and aspect category polarity (T4). The dataset was partitioned initially, but we decided to rely on our random split since it is not balanced.
In aspect term polarity, the distribution of which over the sentiment polarity classes (Positive, Negative, and Neutral) in training, validation, and testing sets are summarized in Table 4.4.
figure 4.3 shows an example from MADAR dataset

Table 4.4: Our random splitting of HAAD dataset for T2.

Label	Training set	Validation set	Test set
Positive	1054	37	285
Negative	972	52	263
Neutral	118	5	24
Total	2144	94	572

In aspect category polarity, we relied on the division shown in Table 4.5

Table 4.5: Our random splitting of HAAD dataset for T4.

Label	Training set	Test set	overall dataset
Positive	586	137	723
Negative	590	161	751
Neutral	16	3	19
conflict	18	2	20
Total	1210	303	1523

```xml
<sentence id="622">
    <text>كتاب مميز اعجبني العديد من المقالات وبالذات فى النصف الاخير من الكتاب</text>
    <aspectTerms>
        <aspectTerm term="كتاب" polarity="positive" from="0" to="4"/>
        <aspectTerm term="المقالات" polarity="positive" from="27" to="35"/>
    </aspectTerms>
    <aspectCategories>
        <aspectCategory category="الاسلوب" polarity="positive"/>
    </aspectCategories>
</sentence>
```

Figure 4.3: An example of HAAD dataset. English translation of the sentence is "special book, I liked many articles, especially in the last half of the book"

4.3.4 The Arabic Hotel Reviews Dataset

The Arabic Hotel Reviews Dataset was submitted at SemEval-2016 to support ABSA's multilingual task, which covers 8 languages and 7 domains. There are 4902 testing tuples and 19,226 training tuples in this dataset. This dataset allowed for both text-level and sentence-level annotations; however, this study only focuses at the latter. As mentioned in [8], this dataset supported 3 tasks, we will focus on task T3 (aspect sentiment classification). This dataset was originally split into two parts: a training set and a test set; for the validation set, we chose 100 samples from the training set. The table 7 shows the number of samples in each label. figure 4.4 shows an example from a Arabic Hotel Reviews dataset

Table 4.6: Our randomly splitting for The Arabic Hotel Reviews dataset .

Label	Training set	Validation set	Test set
Positive	5747	72	1426
Negative	3119	22	784
Neutral	654	6	162
Total	9520	100	2372

```xml
</sentence>
<sentence id="1606:1">
    <text>خدمات مميزه موقع ممتاز وقريب خصوصا لمن معهم نساء، وذلك لقربه من بوابة مصلى النساء، نظافة المكان</text>
    <Opinions>
        <Opinion target="خدمات" category="SERVICE#GENERAL" polarity="positive" from="0" to="5"/>
        <Opinion target="موقع" category="LOCATION#GENERAL" polarity="positive" from="12" to="16"/>
        <Opinion target="المكان" category="LOCATION#GENERAL" polarity="positive" from="87" to="93"/>
    </Opinions>
</sentence>
<sentence id="1606:2">
```

Figure 4.4: An example of Arabic Hotel Reviews dataset. English translation of the sentence is "excellent services and a great location, especially for those going with women because it is close to the women's prayer gate."

4.4 Experimental Series 1

Using MADAR dataset, we trained our normalization model only on the Cairo city dialect.The obtained model's accuracy is equal to 91.1%. We want to see how our two pre-trained language models BERT perform after replacing the Egyptian samples in the HAAD and SemEval-2016 datasets with those predicted by our normalization model. Table 4.7 shows few examples of Egyptian samples from HAAD dataset and their corresponding normalized samples predicted by our normalization framework:

Table 4.7: Examples before and after normalization.

Sentence	Corresponding normalized sentence
الكتاب ده خلاني افكر في حاجات كتير عمري ما فكرت فيها قبل كدة و عجبني اسلوب الكتابة و موضوعاته كلها	الكتاب هذا جعلني افكر في أشياء كثير عمري ما فكرت فيها قبل . و أعجبني اسلوب الكتابة و موضوعاته كلها
أسلوب سارة درويش الجميل اللي بقيت بعرفه من غير ما أبص مين اللي كاتب :)	أسلوب سارة درويش الجميل الذي بقيت أعرفه من غير ما أنظر مين الذي كاتب .:
بجد قصة حلوة اوى بس شبه ما وراء الطبيعة كتير و كانت ممكن تتحط فيها	نعم قصة حلوة جدا لكن شبه ما وراء الطبيعة جدا و كانت ممكن توجد فيها
الكتاب ده حلو بس	الكتاب هذه حلو فقط
قايم كله على فكرة التناسخ اللى كان الهنود بيؤمنوا بيها . مش فى ديننا الكلام دة	قايم كله على فكرة التناسخ الذي كان الهنود يؤمنون بها . ليس فى ديننا الكلام هذا

In experiment 1, we did not normalize all sentences but only sentences in the Egyptian dialect. To get Egyptian dialect sentences, we use CAMeL tools [39]. It is a collection of free and open-source tools for processing natural Arabic language developed by the CAMeL Lab. It has several tools for pre-processing, including Dialect Identification. We list the obtained results with and without normalization in Tables 4.8 and 4.9.

Table 4.8: Performance results on T2 using HAAD and SemEval-2016 datasets with 1st experiment.

dataset	HAAD	SemEval-2016
Without normalization	73.42	83.76
Normalization 1	**74.85**	**83.81**

Table 4.9: Performance results on T4 using HAAD dataset for NLI-M and QA-M methodes with 1st experiment.

Methode	NLI-M	QA-M
Without normalization	75.90	75.08
Normalization 1	**76.89**	**75.49**

Tables 4.8 and 4.9 reports the performance of the models using the accuracy metric. The aspect term polarity model has achieved the best accuracy on both HAAD and SemEval-2016 datasets compared to the results without normalization. As well as, the Aspect category polarity model showed a good performance on the HAAD dataset with normalization for both NLI-M and QA-M methods, unlike without normalization.

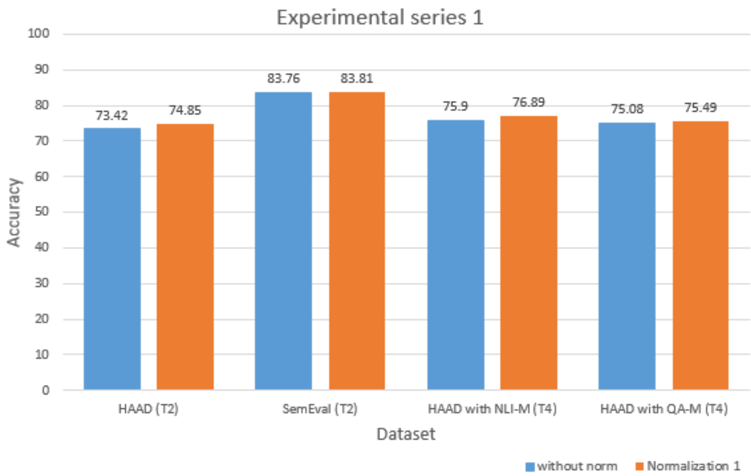

Figure 4.5: Performance of our models after and before normalization of the experimental series 1.

4.5 Experimental Series 2

As in the first experiment, we will only replace the sentences with the Egyptian dialect, but this time our normalization model will be trained on all PADIC and MADAR datasets. Obtained accuracy of training in this experiment was 95%. The results are shown in the following tables.

Table 4.10: Performance results on T2 using HAAD and SemEval-2016 datasets with 2nd experiment.

dataset	HAAD	SemEval-2016
Without normalization	73.42	83.76
Normalization 1	**74.85**	83.81
Normalization 2	74.77	**84.65**

Table 4.11: Performance results on T4 using HAAD dataset for NLI-M and QA-M methodes with 2nd experiment.

Methode	NLI-M	QA-M
Without normalization	75.90	75.08
Normalization 1	**76.89**	75.49
Normalization 2	76.50	**76.48**

Tables 4.10 and 4.11 present the models performance for the second experiment in this study. We can see that the performance of the aspect term polarity and aspect category polarity has increased with the SemEval-2016 dataset and the QA-M method, respectively. Meanwhile, the models have achieved lower results with HAAD and NLI-M methods for T2 and T4, respectively, compared to the first experiment.

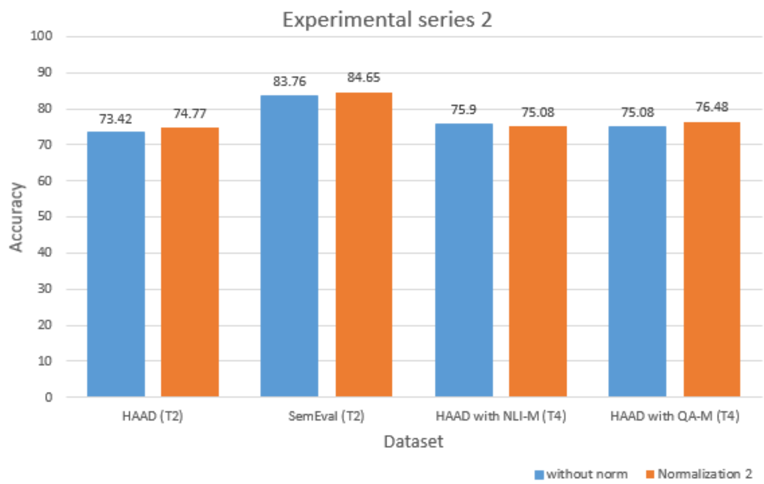

Figure 4.6: Performance of our models after and before normalization of the experimental series 2.

4.6 Experimental Series 3

In this experiment, we will replace all the sentences in HAAD and semeval-2016 with the ones that our normalization model predicted. As in the second experiment, our normalization model will be trained on all PADIC and MADAR datasets with the same achieved accuracy of 95% . The results are shown in the next two tables:

Table 4.12: Performance results on T2 using HAAD and SemEval-2016 datasets with 3rd experiment.

dataset	HAAD	SemEval-2016
Without normalization	73.42	83.76
Normalization 1	**74.85**	83.81
Normalization 2	74.77	**84.65**
Normalization 3	66.39	79.76

Table 4.13: Performance results on T4 using HAAD dataset for NLI-M and QA-M methodes with 3rd experiment.

Methode	NLI-M	QA-M
Without normalization	75.90	75.08
Normalization 1	**76.89**	75.49
Normalization 2	76.50	**76.48**
Normalization 3	75.42	75.24

Tables 4.12 and 4.13 show the models performance for all experiments. The best-obtained result is with experiment 1 and experiment 2 for HAAD and SemEval-2016 datasets for aspect term polarity, respectively. Meanwhile, for aspect category polarity, the best-achieved result is with experiment 1 and experiment 2 for NLI-M and QA-M, respectively.

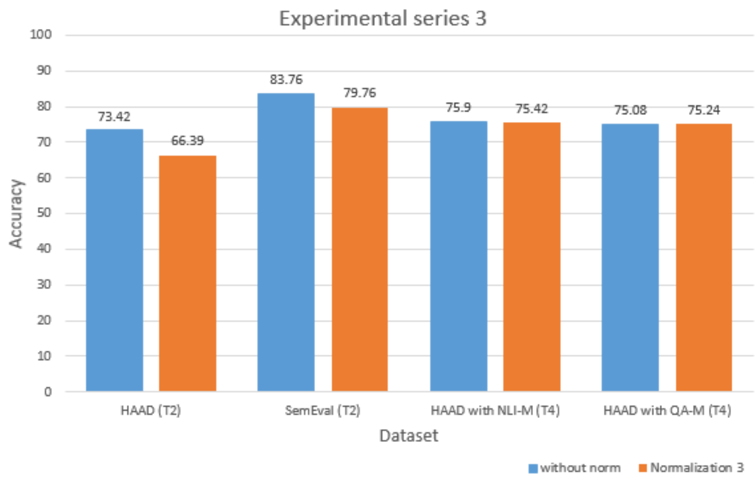

Figure 4.7: Performance of our models after and before normalization of the experimental series 3.

4.6.1 Results Discussion

Obtained results showed a good performance of both aspect term polarity and aspect category polarity models with the first and the second experiments when we applied dialect normalization only on Egyptian sentences in HAAD and SemEval-2016 datasets. On aspect term polarity, the best accuracy of HAAD is with the first experiment (74.85%) and SemEval-2016 with the second experiment(84.65%), on aspect category polarity, the best accuracy of the NLI-M method is with the first experiment (76.89%) and QA-M with the second experiment (76.48%). This improvement confirms that Egyptian words are considered out-of-vocabulary words in Bert's vocabulary, and reducing them will give us the best results. Meanwhile, in the third experiment, when we applied normalization to all datasets, We can notice that the performance had declined. We can justify this decrease because our normalization model may give us OOV words when we use it on MSA, and then the accuracy decreases.

4.7 Experimental Series 4

As a final experiment, we compare our best-obtained results of the three experiments and the previous existing works on both aspect term polarity and aspect category polarity.

4.7.1 Comparison Models:

IAN-BGRU [2]: a based bidirectional GRU interactive attention network model to identify aspect sentiment polarity.

BERT-Linear-pair [14]: a simple and effective BERT-based neural model with the use of sentence pair input on Arabic aspect sentiment polarity classification task.

AB-LSTM-PC [19]: an aspect-based LSTM for sentiment polarity classification which is implemented to handle the second task.

INSIGHT-1(CNN)[40]: a deep learning-based approach which employs a conventional neural network for aspect sentiment polarity.

Lexicon-based approach [5]: several lexicon-based approaches to handle aspect category polarity task.

As Table 4.14 shows, our models achieved the best results on both HAAD and SemEval-2016 datasets with our research tasks T2 and T4.

Table 4.14: Comparison of our best models with previous works.

Model	HAAD	SemEval-2016
T2: Aspect term polarity		
Abdelgawad et al.[2]	NA	83.98
Abdelgawad et al. [14]	73.23	NA
Al-Smadi et al. [19]	NA	82.6
Ruder et al.[40]	NA	82.7
Our model	**74.85**	**84.65**
T4: Aspect category polarity		
Obaidat et al.[5]	71	NA
Our model	**76.48**	NA

Chapter 5: Conclusions and Future Work

Sentiment analysis is one of the essential areas of NLP; it enables companies to improve their products by knowing customer opinions and provides governments with valuable insights about different aspects. Unlike English, a few studies were conducted on Arabic sentiment analysis. Most of this research has focused on sentence-level and text-level sentiment analysis, with only a few studies concerning aspect-level. The most critical challenge that can face the Arabic ABSA is the lack of dialectal resources and datasets that can be used to train the ABSA model. Furthermore, Dialectal Arabic is difficult to process because it breaks all grammatical rules, reducing ABSA model accuracy. Pre-trained language models such as BERT have shown remarkable effectiveness in the sentiment analysis field in recent years. For Arabic, the small and restricted number of available datasets covering multiple Arabic dialects reduces the BERT model in-vocabulary words, thus decreasing the model's performance. This study aimed to increase the effectiveness of the BERT model and reduce the out-of-vocabulary words by translating our dialectical text and transforming it into formal speech. Specifically, the addressed ABSA tasks in this research are aspect term polarity (Task T2) and aspect category polarity (Task T4). The experiments show that the normalized data has enhanced the model's performance and achieved the best results in comparison with other models.

Future goals involve generalizing our study on the rest of the aspect-based sentiment analysis tasks. Improving the normalization performance to make it more flexible and practical and enhancing the results by focusing on more dialectical Arabic is among the most important future directions of this work.

References

[1] M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect-based sentiment analysis using bert," in *Proceedings of the 22nd nordic conference on computational linguistics*, pp. 187–196, 2019.

[2] M. M. Abdelgwad, T. H. A. Soliman, A. I. Taloba, and M. F. Farghaly, "Arabic aspect based sentiment analysis using bidirectional gru based models," *Journal of King Saud University-Computer and Information Sciences*, 2021.

[3] M. Al-Smadi, O. Qawasmeh, B. Talafha, and M. Quwaider, "Human annotated arabic dataset of book reviews for aspect based sentiment analysis," in *2015 3rd International Conference on Future Internet of Things and Cloud*, pp. 726–730, IEEE, 2015.

[4] Z. Obied, A. Solyman, A. Ullah, A. Fat'hAlalim, and A. Alsayed, "Bert multilingual and capsule network for arabic sentiment analysis," in *2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, pp. 1–6, IEEE, 2021.

[5] I. Obaidat, R. Mohawesh, M. Al-Ayyoub, A.-S. Mohammad, and Y. Jararweh, "Enhancing the determination of aspect categories and their polarities in arabic reviews using lexicon-based approaches," in *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–6, IEEE, 2015.

[6] O. Oueslati, E. Cambria, M. B. HajHmida, and H. Ounelli, "A review of sentiment analysis research in arabic language," *Future Generation Computer Systems*, vol. 112, pp. 408–430, 2020.

[7] W. Xue and T. Li, "Aspect based sentiment analysis with gated convolutional networks," *arXiv preprint arXiv:1805.07043*, 2018.

[8] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *International workshop on semantic evaluation*, pp. 19–30, 2016.

[9] N. Liu and B. Shen, "Aspect-based sentiment analysis with gated alternate neural network," *Knowledge-Based Systems*, vol. 188, p. 105010, 2020.

[10] X. Li, L. Bing, W. Zhang, and W. Lam, "Exploiting bert for end-to-end aspect-based sentiment analysis," *arXiv preprint arXiv:1910.00883*, 2019.

[11] X. Li, L. Bing, P. Li, and W. Lam, "A unified model for opinion target extraction and target sentiment prediction," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 6714–6721, 2019.

[12] B. Xu, X. Wang, B. Yang, and Z. Kang, "Target embedding and position attention with lstm for aspect based sentiment analysis," in *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*, pp. 93–97, 2020.

[13] T. E. Trueman, E. Cambria, *et al.*, "A convolutional stacked bidirectional lstm with a multiplicative attention mechanism for aspect category and sentiment detection," *Cognitive Computation*, vol. 13, no. 6, pp. 1423–1432, 2021.

[14] M. M. Abdelgwad, "Arabic aspect based sentiment analysis using bert," *arXiv preprint arXiv:2107.13290*, 2021.

[15] M. Al-Ayyoub, H. Al-Sarhan, M. Al-So'ud, M. Al-Smadi, and Y. Jararweh, "Framework for affective news analysis of arabic news: 2014 gaza attacks case study.," *J. Univers. Comput. Sci.*, vol. 23, no. 3, pp. 327–352, 2017.

[16] M. M. Ashi, M. A. Siddiqui, and F. Nadeem, "Pre-trained word embeddings for arabic aspect-based sentiment analysis of airline tweets," in *International Conference on Advanced Intelligent Systems and Informatics*, pp. 241–251, Springer, 2018.

[17] S. Al-Dabet, S. Tedmori, and A.-S. Mohammad, "Enhancing arabic aspect-based sentiment analysis using deep learning models," *Computer Speech & Language*, vol. 69, p. 101224, 2021.

[18] A.-S. Mohammad, M. M. Hammad, A. Sa'ad, A.-T. Saja, and E. Cambria, "Gated recurrent unit with multilingual universal sentence encoder for arabic aspect-based sentiment analysis," *Knowledge-Based Systems*, p. 107540, 2021.

[19] M. Al-Smadi, B. Talafha, M. Al-Ayyoub, and Y. Jararweh, "Using long short-term memory deep neural networks for aspect-based sentiment analysis of arabic reviews," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 8, pp. 2163–2175, 2019.

[20] M. Al-Smadi, O. Qawasmeh, M. Al-Ayyoub, Y. Jararweh, and B. Gupta, "Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of arabic hotels' reviews," *Journal of computational science*, vol. 27, pp. 386–393, 2018.

[21] J. Yin, X. Jiang, Z. Lu, L. Shang, H. Li, and X. Li, "Neural generative question answering," *arXiv preprint arXiv:1512.01337*, 2015.

[22] A. Javaloy and G. García-Mateos, "Text normalization using encoder–decoder networks based on the causal feature extractor," *Applied Sciences*, vol. 10, no. 13, p. 4551, 2020.

[23] R. Nallapati, B. Zhou, C. Gulcehre, B. Xiang, *et al.*, "Abstractive text summarization using sequence-to-sequence rnns and beyond," *arXiv preprint arXiv:1602.06023*, 2016.

[24] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.

[25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[26] Z. Tu, Z. Lu, Y. Liu, X. Liu, and H. Li, "Modeling coverage for neural machine translation," *arXiv preprint arXiv:1601.04811*, 2016.

[27] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," *arXiv preprint arXiv:1704.04368*, 2017.

[28] I. Lourentzou, K. Manghnani, and C. Zhai, "Adapting sequence to sequence models for text normalization in social media," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, pp. 335–345, 2019.

[29] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," *Advances in neural information processing systems*, vol. 28, 2015.

[30] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[31] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.

[32] A. Safaya, M. Abdullatif, and D. Yuret, "Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media," in *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pp. 2054–2059, 2020.

[33] C. Sun, L. Huang, and X. Qiu, "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence," *arXiv preprint arXiv:1903.09588*, 2019.

[34] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Bert post-training for review reading comprehension and aspect-based sentiment analysis," *arXiv preprint arXiv:1904.02232*, 2019.

[35] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, *et al.*, "Huggingface's transformers: State-of-the-art natural language processing," *arXiv preprint arXiv:1910.03771*, 2019.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[37] H. Bouamor, N. Habash, M. Salameh, W. Zaghouani, O. Rambow, D. Abdulrahim, O. Obeid, S. Khalifa, F. Eryani, A. Erdmann, *et al.*, "The madar arabic dialect corpus and lexicon.," in *LREC*, 2018.

[38] K. Meftouh, S. Harrat, S. Jamoussi, M. Abbas, and K. Smaili, "Machine translation experiments on padic: A parallel arabic dialect corpus," in *Proceedings of the 29th Pacific Asia conference on language, information and computation*, pp. 26–34, 2015.

[39] O. Obeid, N. Zalmout, S. Khalifa, D. Taji, M. Oudah, B. Alhafni, G. Inoue, F. Eryani, A. Erdmann, and N. Habash, "CAMeL tools: An open source python toolkit for Arabic natural language processing," in *Proceedings of the 12th Language Resources and Evaluation Conference*, (Marseille, France), pp. 7022–7032, European Language Resources Association, May 2020.

[40] S. Ruder, P. Ghaffari, and J. G. Breslin, "Insight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis," *arXiv preprint arXiv:1609.02748*, 2016.