Deep Learning for Aspect-Based Sentiment Analysis of Government Mobile Apps Reviews

استخدام التعلم العميق لتحليل المشاعر المستند على السمات للتعليقات على التطبيقات الحكومية الذكية

by

BILAL YACOUB SIYAM

Dissertation submitted in fulfilment

of the requirements for the degree of

MSc INFORMATICS
(KNOWLEDGE AND DATA MANAGEMENT)
at

The British University in Dubai

September 2018

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

Signature of the student

Abstract

The number of smartphone users in the United Arab Emirates (UAE) has increased by 25% in the last four years (Statista.Com 2018c). Government entities in the UAE aim to provide their services through various digital channels, including smartphones among others. This offering will allow customers to use their preferred digital channels and cover the largest customers' segment. As life is accelerated and people are seeking a quick and efficient way to consume services, government entities become concerned in providing easily accessible services through smart applications. On the other hand, smartphone users are interested in considering others' opinions and reviews before downloading or using an application. Thus, it is essential for government entities to take into consideration these opinions and comments for the purpose of developing and improving their applications and providing the intended value to their customers. Sentiment Analysis (SA) is the process of analysing textual information and understanding the intent related to emotions, feelings and behaviours. SA is classified into three levels of analysis, namely document-level, sentence-level and aspect-level. This study focuses on aspect-based sentiment analysis (ABSA) of Dubai government mobile apps reviews. ABSA takes into account analysing various features or aspects stated in the review. This study aims at exploring the use of deep learning techniques in improving the performance of ABSA in government app reviews domain. Our approach uses Convolutional Neural Network (CNN) framework which is one of the deep learning approaches. Deep learning is chosen in this study over other approaches since it requires less human intervention and less effort in features engineering. In addition, it simulates how humans think in representing patterns and simplifying them. The approach utilized word embeddings to represent the reviews as vectors and inducted them into the input layer of the deep learning model. The proposed CNN framework consisted of three hidden layers: convolutional, pooling and classification. Further, several techniques and hyper-parameters have been explored as well as their effect on the accuracy of our framework. The performance measures showed that (GloVe) outperformed other word embeddings models. Additionally, utilizing variety of activation functions, namely Sigmoid, Tanh and Softmax enhanced the model performance by 6.5%. The proposed framework achieved high performance results with 89.42% accuracy. This study highlighted several advantages of using deep learning approach. For instance, it does not require much human experience in the domain of the problem compared to the rule-based approach. Further, it needs less effort in identifying features for training.

Keywords: Aspect-Based Sentiment Analysis; Deep Learning; Neural Networks; CNN; Word Embeddings.

ملخص

ازداد عدد مستخدمي الهواتف الذكية في دولة الإمارات العربية المتحدة بنسبة 25% في السنوات الأربع الأخيرة (Statista.Com 2018c)، والذي تماشى أيضاً مع اهتمام الجهات الحكومية في دولة الإمارات بتقديم خدماتها من خلال مختلف القنوات الرقمية، بما في ذلك الهواتف الذكية. أتاح هذا التوجه الحكومي للمتعاملين استخدام القنوات الرقمية المفضلة لديهم وتغطية شريحة أكبر من المتعاملين. ومع تسارع نسق الحياة واحتياج الناس لوسيلة سريعة وفعالة للحصول على الخدمات، فقد سارعت الجهات الحكومية بتوفير خدمات يسهل الوصول إليها من خلال التطبيقات الذكية. في المقابل، يهتم مستخدمو الهواتف الذكية بمعاينة آراء وتعليقات الآخرين قبل تحميل أو استخدام التطبيقات الذكية، وبالتالي فإنه من الضروري أن تأخذ الجهات الحكومية بعين الاعتبار هذه الآراء والتعليقات وتحليلها لغايات تطوير وتحسين تطبيقاتهم الذكية وتقديم القيمة المرجوة لمتعامليهم. تحليل المشاعر هو عملية تحليل المعلومات النصية وفهم النوايا المتعلقة بالمشاعر والأحاسيس والسلوكيات، ويصنف إلى ثلاثة مستويات من التحليل: مستوى المستند ككل، ومستوى الجملة، ومستوى السمة. تركز هذه الدراسة على تحليل المشاعر المستند على السمات للتعليقات على التطبيقات الحكومية الذكية في دولة الإمارات العربية المتحدة، ويأخذ هذه المستوى من التحليل بعين الاعتبار تحليل المميزات أو الجوانب المختلفة المذكورة في التعليقات، كما تهدف هذه الدراسة إلى استكشاف استخدام تقنيات التعلم العميق في تحسين أداء تحليل المشاعر المستند على السمات في مجال التعليقات على التطبيقات الحكومية الذكية. يوظف النهج المتبع في هذه الدراسة نموذج "الشبكة العصبية الملتفة" والذي يعد أحد نماذج التعلم العميق، وقد أختير نهج التعلم العميق بدلاً من المناهج الأخرى؛ لأنه لا يتطلب تدخلاً بشرياً ومجهوداً عالٍ في هندسة الميزات قبل إدراجها لنموذج التعلم العميق، بالإضافة إلى أنه يحاكي طريقة تفكير الإنسان في تمثيل الأنماط وتبسيطها. كما استخدم إطار العمل المقترح في الدراسة نماذج "تضمين الكلمات" لتمثيل التعليقات كمتجهات لإدخالها في طبقة المدخلات في نموذج التعلم العميق، ويتكون نموذج "الشبكة العصبية الملتفة" المقترح من ثلاث طبقات مخفية: الطبقة الملتفة، وطبقة التجميع، وطبقة التصنيف. علاوة على ذلك، تم استكشاف تأثير العديد من التقنيات والعوامل على دقة الإطار المقترح، فقد أظهرت مقاييس الأداء أن نموذج (GloVe) تفوق على نماذج "تضمين الكلمات" الأخرى، كما أن استخدام مجموعة متنوعة من وظائف التفعيل وهي (Sigmoid) و(Tanh) و(Softmax) أدى إلى تعزيز أداء النموذج بنسبة 6.5٪، وقد حقق الإطار المقترح نتائج أداء عالية بدقة تصل نسبتها إلى 89.42٪. وأبرزت هذه الدراسة عدة مزايا لاستخدام نهج التعلم العميق؛ على سبيل المثال، لا يتطلب الأمر الكثير من الخبرة البشرية في مجال المشكلة مقارنة بالنهج القائم على القواعد، كما يحتاج إلى جهد أقل في تحديد وهندسة الميزات المراد استخدامها في تدريب النماذج.

Acknowledgement

First and foremost, I thank Allah Almighty for His guidance and support to complete this dissertation.

The work presented in this dissertation would not have been possible without the help, advice, inspiration, and encouragement of many people.

I am grateful to Prof. Khaled Shaalan for his guidance, motivation and moral support. I was always able to reach out to him when I got any difficulties or queries regarding my study.

Many thanks to my family, my parents, my sisters and my brother, and my colleagues for all their support, prayers, patience, motivation and infinite support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Description
SA	Sentiment Analysis
ABSA	Aspect-Based Sentiment Analysis
NLP	Natural Language Processing
CL	Computational Linguistics
AI	Artificial Intelligence
NN	Neural Network
ANN	Artificial Neural Network
DNN	Deep Neural Network
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
IAN	Interactive Attention Networks
XML	Extensible Mark-up Language
JSON	JavaScript Object Notation
SVR	Support Vector Regression
SVM	Support Vector Machine
TF-IDF	Term Frequency - Inverted Document Frequency
UAE	United Arab Emirates
OTE	Opinion Target Expression
KNN	K-Nearest Neighbor
LR	Logistic Regression

MLP	Multi-layer Perception
BP	Back Propagation
BPN	Back Propagation Neural network

1. CHAPTER ONE: INTRODUCTION

During the past few decades, the world has witnessed great development in the field of science and technology. This may be related to the global needs for new techniques and innovations to improve offered services' quality. Sentiment Analysis (SA, hereafter) is one of the contemporary concepts that led to an extensive impact over the quality of services in both the private and public sectors (Siddiqui, Monem & Shaalan 2016). SA refers to the process of computing and analysing individuals' prospective concerning a specific topic. It represents one of the commonly widespread domains in text analytics (Siddiqui, Monem & Shaalan 2018). This process is of importance as organizations in the private and public sectors are very interested in knowing the individuals' points of views about the services of the provided products (Wagner et al. 2014).

The introduction of WEB 2.0 technology had helped the internet and social media users and product consumers in showing their attitudes and feelings as well as permit the generated available information to develop promptly on social media (Shaalan 2017). Public points of views on the internet have great contribution in the development of daily lives. This means that users' opinions and feedback on websites are a very significant integral source of obtaining information, particularly in analysing social media websites. SA assists in comprehending the individuals' attitudes towards a specific subject in different fields. Additionally, it aids in obtaining information about the service quality of a product and in automatically taking decisions concerning the essential aspects of the service. Moreover, it helps in briefing the different constructive and undesirable reviews relating to the service qualities. This means that SA is implemented to obtain information about the customers' perspectives concerning their text reviews about the quality of provided services of a particular topic (Kiritchenko et al. 2014).

Consequently, with the flow of data in the internet and especially in social networking sites, we often hear about the analysis of sentiment using machine language and note the orientation of elite researchers on this field. SA caught the attention of many scholars in this field since taking decisions basically depends on extracting public's opinions. Thus, it is significant to find out the individuals' points of views and opinions about the quality of certain e-services (Sanchez & Vazquez 2014).

SA can be defined as the process of analysing opinions and emotions to infer the apparent tendencies of the analysed data and classifying them as positive, negative or sometimes neutral. SA is an application of natural language processing (NLP) and computational linguistics (CL) to extract the feeling and sense (i.e., subjective view) about the target aspects of an entity. The analysis of emotions is intended to determine the position of the speaker and his impression on an issue in the studied text (Mikolov et al. 2013).

To conclude, the analysis of opinions and all the concepts related to it such as feelings, assessments, attitudes and emotions are entirely subject to study by SA techniques. SA targets to understand the individuals' sentiments towards a specific entity or topic. It helps capturing insights about the provided services by organizations in order to take data-driven decisions. This study intends to investigate the use of deep learning in Aspect-Based Sentiment Analysis (ABSA, hereafter) of government app reviews.

1.1. Emergence of Smart Apps

Social media has undergone many developments recently. This took the attention of many decision-makers in both public and private sectors. Social media is viewed as one of the most essential and effective trends in the information technology sector worldwide (Al Suwaidi, Soomro & Shaalan 2016). Social media is an essential method in the process of content and knowledge sharing (Alqaryouti, Alqudah & Shaalan 2016; Cambria 2016). The significance of social media results from the huge amount of users exceeding the 4 billion as well as the amount of time these users spend online (Schouten, Frasincar & de Jong 2017; Statista.Com 2018a).

The growing use of Smart Devices resulted in the prompt development in smart apps' stores. That is, smart phones' devices use has rapidly increased compared to earlier engagement. This grabbed the attention of decision-makers in both the private and public sectors resulting in the inclination to understand the opinions and views of internet users and bloggers (Addo-Tenkorang & Helo 2016).

Google Play and Apple Store are the most popular stores serving more than 5 million users. Figure 1-1 depicts the most common app stores (Statista.Com 2018b). Similarly, statistics showed that the number of applications downloaded in 2017 were about 178.1 billion and they are expected to reach 258.2 billion app downloads in 2022.

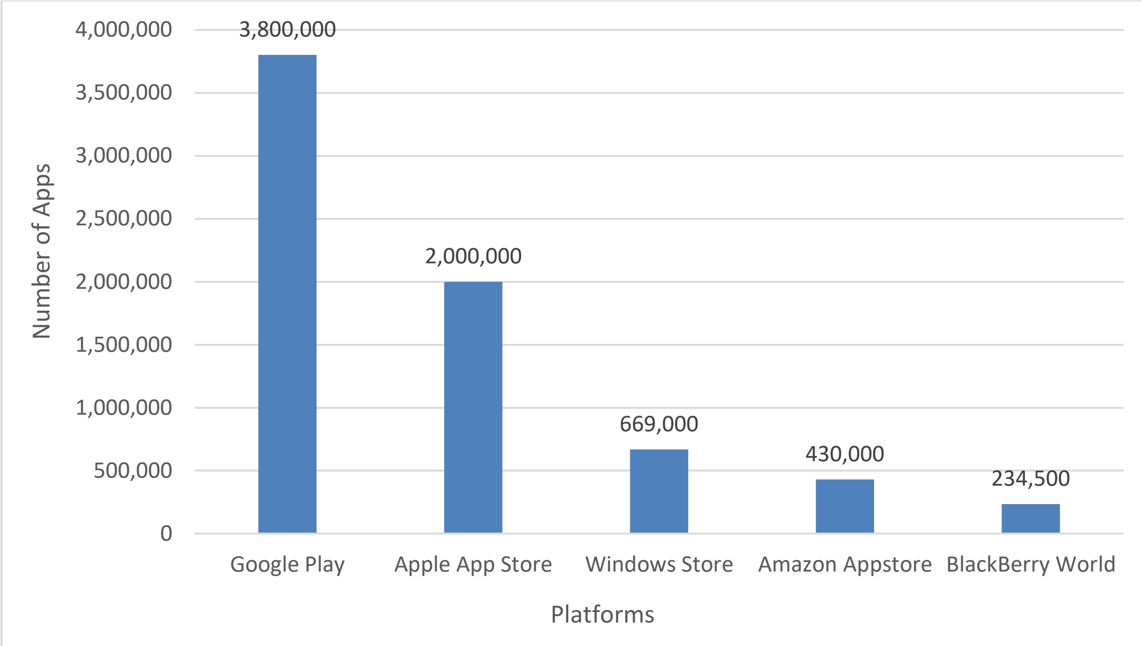

Figure 1-1: Number of apps available in leading app stores as of 1st quarter 2018

Lately, both private and public sectors companies have progressively emphasised on the significance of social media and smart apps to understand internet users and consumers requirements as a result of the high demands on more effective apps. For example, decision-makers in private and public sectors can reach *Likers* and *Followers* of a specific post or product on Facebook pages in seconds. This helps interacting and understanding the feedback of the page's *Followers* or *Likers* in short time. According to Asur and Huberman (2010) a posted tweet spreads on followers' profiles and Twitter's pages very quickly.

Smart Applications websites help applicants and internet users circulate posts and stay tuned with peers and friends worldwide any time. This also assists them in expressing their feeling and emotions when interacting with posts as well as access the opinions of others. The analysis of opinions and all the concepts related to them such as feelings, assessments, attitudes and emotions are entirely subject to study by the techniques of SA. The field of opinion analysis has evolved in some influential areas in NLP and data mining due to the vast amount of information shared over the internet (Poria et al. 2016).

Consequently, the evolution of advanced methods has raised the awareness of private and public organizations concerning the need of internet users and followers (Zaslavsky & Georgakopoulos 2015). Social media users can express their views using smart applications related to a specific post or products on varied topics and interests throughout Facebook, Twitter, Google Plus, YouTube, Instagram and LinkedIn. The content of these reviews do not follow grammar or spelling rules and usually uses abbreviations. However, this kind of language is easy to use and understand by most social media users. This sheds light on a considerable number of linguistic issues including grammar, meaning and syntax that may hinder automatic data extraction from reviews' sources (Alkhatib & Shaalan 2018). SA provides emphasis on text analysis depending on the implementation of semantic networks. This permits the efficiency in NLP data depending on greater base of information and methods (Qazi et al. 2017).

In brief, smart apps are now considered as an essential channel to consume services by customers. Additionally, smart apps users post considerable amount of reviews on daily basis to express their opinions and interests toward the quality of the provided services. This grabbed the attention of service providers in both private and public sectors to understand their customers' opinions and feedback in order to develop and enhance these services.

1.2. Dubai Government Directions and Strategies

As a result of the excessive demand on efficient channels for providing influential electronic and smart services, public sector services' provisioning has changed because of the great development in technology and rapid proliferation of smart phones communicators. That is, due to the influence of technological trends and consumers' needs, public sector has inventive strategies in order to shift towards providing competitive quality and advanced smart services (Wagner et al. 2014; Hai et al. 2015; Kim et al. 2016).

This may be viewed in the United Arab Emirates (UAE) experience in using smart technologies as directed by the Vice-President and Prime Minister of the UAE and Ruler of Dubai, His Highness Shaikh Mohammad bin Rashid Al Maktoum, who urged all the government's parties to utilize and adopt the system of smart services instead of Electronic Government services in order to support the activation of what is called the Smart Government. This vision embraced the benefits of the community members when using available governmental services through smart phones (GulfNews 2013). The emergence of smart governments reveals the significance of being updated with the latest technological developments and consumers' expectations.

In 2014, the government of Dubai announced an initiative to figure out the publics' contentment with the provided government services in order to detect the consumers' satisfaction towards the administrative services. The consumers have the ability to rate their satisfaction and write their comments about the services. This assists in improving the quality of the facilities to achieve the public happiness and rate of competitiveness amongst governmental sectors in different fields (KhaleejTimes 2014). This exceptional governmental initiative reflects the significance of evaluating and examining the reactions and responses of customers concerning the governmental services as a main issue for imminent developments that represent users' expectations. Public sectors' entities capitalize a great deal of resources to obtain customers' comments and responses. SA is very significant for public organizations as they incline to know the quality of services from customers' perspectives in order to consider providing high quality services. The adoption of smart services has been implemented in many other countries, such as Malaysia, who aim to apprehend the public demands, reflecting the importance of SA (Hasbullah et al. 2016).

Recently, the government of Dubai set a target that intends to raise the percentage of users who benefit from government services through mobile devices up to 80%, forcing developers to provide more mobile services with optimum user-friendliness.

Similarly, in October 2017, the UAE has launched the first "UAE Strategy for Artificial Intelligence" directed by the UAE Vice President and Prime Minister and Ruler of Dubai His Highness Sheikh Mohammed bin Rashid Al Maktoum (Sheikhmohammed.ae 2017). This is built on top of the successful phase of Smart Government program and it raises the bar to a new level of innovation. The strategy aims to explore and invent future opportunities to innovate government services through AI (uaeai.ae 2017).

1.3. The Concept of Sentiment Analysis

Since 2000, and for the first time in humans' history, the information amount and registered data have taken a vast amount in social networks and forums, which was an evidence of the rapid and simultaneous growth of communication methods (Tang et al. 2014). With the flow of data in the internet and especially in social networking sites, we often hear about the analysis of sentiment using machine language and note the orientation of elite researchers on this field. SA is an NLP and CL application used to extract the feeling of the target entity. The analysis of emotions is intended to determine the position of the speaker and his impression on an issue, including positive or negative polarity in the text under study (Anand & Naorem 2016). Opinion analysis has been considered as one of the most effective aspects under the categories of data mining and NLP.

Some researchers refer to SA as Opinion Mining (OM) (Kumar & Jaiswal 2016; Zhang & Liu 2016; Yadollahi, Shahraki & Zaiane 2017). They consider the process of analysing opinions as the process of studying people's views, behaviour and expressions towards certain issues. These issues can be an event, product, service or people. Both SA and OM are intersecting and useful in the same sense. Researchers have agreed that these two terms have the same meaning and have convergent objectives despite some differences between them. The purpose of OM is to extract and analyse people's views on a specific issue, while SA is the identification of emotion revealed in the text and then analysis (Chiu et al. 2015).

Programmers and linguists have developed algorithms that deal with the sentence and automatically analyse it to recognize emotions through words. Often, these programs deal with polarized words. Polarization refers to the classification of words into negative, positive or neutral. The result of this automatic analysis is called emotion analysis. Some ambitious systems go beyond polarization to address specific emotions such as happiness or anger (Al-Smadi et al. 2018).

1.5. Research Motivation

The huge amount of information in social media and websites attracted the attention of many decision-makers in private and governmental organisations. This raised the importance of reviewing the customers’ opinions and points of views about the quality of services and products. Thus, this topic is believed of utmost importance due to the following points:

- Unstructured large amount of data containing citizens’ opinions about government smart applications is available.
- The need to deal with such unstructured opinionated data.
- The need to have an automated and effective way to understand the sentiment of the public towards government smart applications.
- Present research in ABSA focuses on rule-based approach and traditional machine learning models approaches. However, deep learning approaches is understudied.
- Current studies show that the methods dealing with ABSA for government apps reviews are limited.

1.6. Research Objectives

This study aims at investigating the use of deep learning in ABSA of government app reviews that include a variety of aspects. It intends to achieve the following objectives:

- To extract sentiments toward different aspects in mixed-opinion reviews and variety of aspects.
- To explore an effective way to utilize deep learning techniques in ABSA of government app reviews.

1.7. Research Questions

In order to fulfil the objectives of this research, we aim to answer the following research questions:

- Q1. How can deep learning be utilized in ABSA of government app reviews that include a variety of aspects?
- Q2. Can deep learning techniques improve the performance of aspect-based sentiment classification for government app reviews?

1.8. Dissertation Structure

This dissertation consists of six chapters, including this introductory chapter. Chapter two begins by demystifying ABSA and deep learning as well as the importance of employing and analysing both concepts. In addition, it investigates the related literature relevant to this topic. The third chapter is concerned with the methodology used for this study. Chapter four discusses and analyses the results and performance of the experiments as well as tying up the various theoretical and practical implications. The fifth section provides the answers for the research questions. Finally, chapter six concludes the main areas covered in this dissertation and includes a discussion of the limitation and future prospects.

2. CHAPTER TWO: REVIEW OF LITERATURE

The emergence of modern specialized data collection techniques has led to a large-scale accumulation of data in various fields. Traditional data query methods have become inadequate methods for extracting useful information from large databases of various types. The ability of humans to store this vast amount of information without using techniques that help them understand data and extract knowledge from it is limited. Hence, it is required to analyse sentiment or opinions using the techniques of NLP, computer-linguistics and textual analysis to detect positive and negative feelings or neutral ones towards the subject of the text. SA is used on a scale in the areas of marketing, customer service and others (Baecchi et al. 2016).

SA generally aims at identifying the speaker or writer's feelings on a document. Bin et al. (2017) state that SA is used to define the position of the speaker on some topics or guide the overall context of the text. The attitude may be related to judgment or assessment, emotional state, or the intended emotional communication. However, it is important to identify sentiments toward specific aspects instead of the overall context which allows organizations to benefit more from SA outcomes. This is called Aspect-Based Sentiment Analysis (ABSA). ABSA takes into account classifying reviewers opinions toward different features or aspects of the entity presented. This is important since the person can express different feelings about features of the same entity.

This research utilizes Deep Learning techniques to solve various challenges and limitations of rule-based and statistical machine learning approaches. Deep learning is a new research area that deals with the creation of theories and algorithms that allow the machine to learn through modelling neurons of the human body. It is a branch of science that deals with artificial intelligence (AI) and one of the branches of machine learning (Deng & Yu 2014).

In this regard, this chapter aims at reviewing and discussing significant concepts of ABSA and its importance. Additionally, more discussion about deep learning framework is introduced in this chapter as well as related literature relevant to this topic.

2.1. Aspect-Based Sentiment Analysis

SA helps organizations and brand owners measure the satisfaction of their customers with a particular product by using a more rapid, comprehensive and credible analysis of views than traditional questionnaires. Using an analysis of opinions, the beneficiary can reveal the feelings and opinions of the general public about a specific product or subject (Muhammad, Wiratunga &

Lothian 2016). Each review consists of different elements that can be explicitly shown in the review or implicitly understood from the context. Liu (2010) stated that a review consists of five elements, namely sentiment target (entity), entity aspect, the sentiment on the aspect, sentiment source, and the time. Figure 2-1 illustrates these five elements that formulate a sample review.

Figure 2-1: Review sample demonstrating different elements that formulate a review

Wiebe, Wilson & Cardie (2005) pointed out that SA includes three levels of analysis: Document level, sentence level and aspect level. At the document level analysis, the document is treated as a single piece of information, and the objective is to classify the document positively or negatively based on the emotions detected.

In the sentence level analysis, the goal is to categorize the sentiments shown in the sentence. The work is of two phases. The first phase is concerned with determining whether a sentence contains feelings. In the second phase, the feelings, if any, are classified into negative or positive.

Lastly, in the level of aspect analysis, the main task includes the classification of feeling, considering specific features or aspects of the entity presented. The task at this level includes two phases. The first phase includes the identification of the entity and its features, while the second phase consists of the analysis and classification of emotions associated with each feature. One individual can express different feelings about different features for the same entity. This level of analysis is called ABSA.

Each of these levels of analysis has its own advantages and limitations (Cambria et al. 2013). The experiment conducted in this research focuses on the analysis at the level of aspects. Various studies were conducted at this level. However, it is still considered a challenging task and has its limitation (Poria et al. 2016). On the other hand, in order to allow organizations to benefit from SA outcomes, it is crucial to drill down to this level of analysis. In the example illustrated in Figure 2-1, the reviewer expressed that he has a positive opinion towards the user interface feature in a specific app. Hence, the opinion expressed was toward only a specific feature, not toward the whole product. So while the reviewer had expressed a positive opinion towards one feature, he may state that he does not like another feature, like the performance. In this case, different polarities will be identified towards different aspects of the same product. This can help the organization to consider enhancements on specific features.

Several attempts were experimented to enhance the performance of ABSA. Al-Smadi et al. (2018) conducted a study that aimed to analyse the hotels Arabic reviews through an enhanced model of ABSA by using supervised machine learning. To achieve the study objective, the researchers employed a set of classifiers including: Bayes Networks, K-Nearest Neighbor (KNN), Decision Tree, and Naïve Bayes. The authors trained a set of classifiers with morphological, syntactic, and semantic features to address the research tasks, namely: (a) T1:Aspect Category Identification, (b) T2:Opinion Target Expression (OTE) Extraction, and (c) T3: Sentiment Polarity Identification. The study used a data set from SemEval 2016. The results of the study indicated that the supervised learning approach outperforms related work evaluated using the same data set. More precisely, evaluation results showed that all classifiers in the proposed approach outperform the baseline approach, and the overall enhancement for the best performing classifier (SVM) is around 53% for T1, 59% for T2, and 19% in T3.

Laskari and Sanampudi (2016) conducted a study that aimed at identifying the aspects of the target entities and the sentiment expressed for each aspect. To achieve this objective, the authors first extracted the aspect terms and grouped them into aspect categories, and then determined the polarity of the aspect terms and aspect categories of each sentence. To run this system, the study used train data sets of different products, such as restaurants, laptop and hotels reviews. Additionally, the researchers collected data from other resources like tripadviser.com and amazon.com. Various techniques were used for data pre-processing, such as normalization of data, stop words and removing punctuations. The results of the study showed that the proposed system is good but cannot be applied to all the data sets, since non-matching aspect words in cross domains framework has not been defined.

Pateria (2016) proposed an approach that relies on predicted probability estimates from first level to second level, especially at incorporate neighbour information stage. The study used data from SemEval 2016 Workshops to evaluate the proposed model. The study achieved a better performance compared to base-line models, with a prediction accuracy in a range of 83 to 88%.

Brychcín, Konkol and Steinberger (2014) aimed to identify the field of some target entities and the sentiment expressed for each field. In order to achieve the study objective, the researchers provided a system that relied on supervised machine learning. This system is strictly constrained and depends on a training data. The researchers used this type of data since it's the only available source of information. Additionally, the study used an unsupervised method for LDA and semantic spaces latent semantics discovery. The model relied on sentiment vocabularies in order to extend the system. The system was applied in two areas; restaurants and laptops. The results of the study indicated that the proposed approaches achieved very good results. The constrained versions were always above average, often by a large margin. The unconstrained versions were ranked among the best systems.

Castellucci et al. (2014) presented a system built through using Kernel Methods during the framework of Support Vector Machine. The first phase, Aspect Term Extraction, was tackled and designed through SVM as a sequential tagging task. Other phases include: Aspect Term Polarity, Aspect Category and Aspect Category Polarity detection, designed as a classification problem in order to linearly combine multiple kernels to generalize a set of linguistic information. The results of the competition showed that the proposed system achieves outperforming results compared to other competitors.

Hamdan, Bellot and Béchet (2014) conducted a study to find the polarities and determining the categories of the sentence and the polarity of each category for the filed terms that exists in each sentence. To achieve these objectives, four subtasks were performed; aspect terms extraction, aspect term polarity detection, category detection, and category polarity detection. The system was tested using the training data provided by SemEval 2014 ABSA task, which includes two data sets. The first data set contains 3,000 sentences of restaurant reviews annotated by the aspect terms, their polarities, their categories and the polarities of each category. The second data set contains 3,000 sentences of laptop reviews annotated just by the aspect terms and their polarities. The evaluation results showed a significant improvement in accuracy and f-measure compared to baseline studies.

In this section, we described the basic elements that form a review, namely sentiment target (entity), entity aspect, the sentiment on the aspect, sentiment source, and the time. Additionally, different levels of analysis in the domain of SA were introduced focusing on the analysis at the level of aspect, which is the focus of this research. We stated that in order to allow organizations to benefit of SA outcomes, it is crucial to drill down to this level of analysis. ABSA takes into account classifying reviewers opinions towards different features or aspects of the entity presented. This is important since the person can express different feelings about different features for the same entity. Finally, we reviewed several researches in the domain of ABSA as well as different techniques used in each attempt.

2.2. Demystifying Deep Learning

Deep Learning is a new research area that deals with the creation of theories and algorithms that allow the machine to learn by modelling neurons of the human body. It is a branch of science that deals with AI and one of the branches of machine learning (Deng & Yu 2014). Deep learning has developed since 2006 within the Machine Learning's field. Deep learning techniques have evolved over the past years affecting many fields and information processing in traditional and modern forms within the basic concepts of machine learning and AI (Lee et al. 2017).

Machine learning in general is not new as it was originated in the mid-20th century. Where English athlete Alan Turing proposed a vision of AI and the machine learning. Over the next decades, several AI techniques have emerged and disappeared. One of these techniques is neural networks (NNs). The algorithms that support deep learning play a pivotal role in the recognition of images and robotic vision. They are inspired by the neurons that make up the human brain. NNs include neural layers that are interconnected to form several interconnected layers. The more layers the deeper the network (Baldi, Sadowski & Whiteson 2014). There is also unsupervised learning approaches which uses, for instance, images without a description of them and here the NN must recognize the different patterns in the image to begin to identify any image that may contain these patterns (Bastien et al. 2012).

The applications of deep learning are now imbedded in banks, communication and storage services, where computers recognize the extent to which customers are satisfied with their enormous numbers of services. Deep learning is also used to identify good or non-paying clients, recognize biomedical diagnosis, and detect drugs, fingerprints, handprints and DNA. It is currently widely used in email filtering for spam, translation services, classifications, and text-to-speech conversion (Bengio 2013).

In addition, deep learning uses algorithms with different representations to model complex relationships within data, thereby defining high-level features and concepts (Xu et al. 2017). Moreover, deep learning is acquired at different layers conforming to their abstraction, and implements artificial neural networks (ANNs). Layers in these arithmetical learning representations are consistent with different high-level concepts described by lower-level concepts (Yanmei & Yuda 2015).

Severyn and Moschitti (2015) revealed that the methods of deep learning in ANNs are being presented by two proper examples: Multi-Layer Perceptrons (MLPs) and Back Propagation (BP). MLPs is a class of feed-forward ANN supplied with many hidden layers. On the other hand, BP became common in the 1980s and was a famous algorithm for learning networks' coefficients. However, BP by itself did not perform well enough with learning networks that consisted of a huge number of hidden layers, as both the networks' difficulty and depth were increased.

Additionally, word embedding is a significant field in NLP that uses the principle of deep learning. It relies on the method of semantic analysis by extracting the meaning from the text to understand natural language. Moreover, word embedding creates a map of similar words to help deep learning predict text’s meaning by recognizing the contextual similarity of the words (Mehta, Parekh & Karamchandani 2018). Hence, word embedding is a method to predict and understand the meaning of the sentence from its words. This method depends on connecting the words that have a relationship. For example, in a sentence that contains (eat and juice) words, we can find fruit or vegetables words like (apple, potato, and orange), while we cannot find other words like (plane and car).

Researchers have developed several frameworks for deep learning to overcome different challenges. One of the most famous frameworks is Convolutional Neural Networks (CNNs). CNNs do not depend on parsers and compute vectors in sub-phrasing process. In this regard, unified NN architecture was proposed to be used in different NLP tasks (Srivastava et al. 2014). Another famous framework is Recurrent Neural Networks (RNNs). RNNs are sequential approaches that form the network of the previous and present state of input. This led to the emergence of propagation and its effect at the end of the sequence. During forward computation, gradients calculated at each time step are propagated backward for error calculations (Mikolov et al. 2013).

Multi-label classification (MLC) is an important learning problem that is expected to take the hidden correlation of the labels under consideration for learning algorithm. Extracting the hidden correlation is usually a difficult venture. Also, MLC of data took into consideration difficult problems. This is because of complex data that is sometimes hard to infer statistics from for non-mutually exclusive classes (Maxwell et al. 2017).

In summary, deep learning uses computers to test the logarithms and programs and learns to improve and develop them by itself. As mentioned earlier, deep learning deals with several layers of processing nonlinear and naturally occurring information; depending on how structures and techniques are used (formation, definition or classification) (Ruder, Ghaffari & Breslin 2016).

In this section, we highlighted that deep learning is a branch of AI and part of the machine learning family that simulates the processes of the human brain in learning data representations and creating patterns to use in the process of making decisions. In addition, we briefly discussed some famous deep learning architectures, such as CNNs and RNNs. These architectures are widely used in several domains and applications such as spam detection, translation, classifications, and text-to-speech conversion.

2.3. Deep Learning Approaches in Sentimental Analysis

Researchers have demonstrated various approaches and models using deep learning in SA. Some of these models utilize CNN framework (Santos & Gatti 2014; Ouyang et al. 2015; Severyn & Moschitti 2015; Yanmei & Yuda 2015; Xu et al. 2017; Chakravarthy et al. 2018), while other models utilize RNN framework (Al-Smadi et al. 2017; Chakravarthy et al. 2018). Additionally, some researchers tested other deep learning models, such as Interactive Attention Networks (IAN) and Long Short-Term Memory (LSTM) network (Lee et al. 2017; Ma et al. 2017; Ma, Peng & Cambria 2018). Moreover, word embedding models are employed in many of these researches (Santos & Gatti 2014; Ouyang et al. 2015; Severyn & Moschitti 2015; Al-Smadi et al. 2017; Lee et al. 2017; Ma et al. 2017; Xu et al. 2017; Chakravarthy et al. 2018).

Chakravarthy et al. (2018) examined word embedding techniques and implemented a model to achieve optimal accuracy. The researchers employed a set of classifiers including a hybrid architecture consisting of CNNs and RNNs to implement the DL model on the SAR14 and Stanford Sentiment Treebank data sets. The results of the study indicated that RNN cannot handle long term dependency.

Similarly, Ma, Peng and Cambria (2018) conducted a study to investigate individual's opinions and sentiments towards aspects. The study adopted a completely unique method that tackles the challenges of ABSA tasks SA by exploiting common sense data. LSTM network with a hierarchal attention mechanism consisting of an aspect-level attention and a sentence-level attention were augment. Common sense information of sentiment-related ideas is incorporated into the end-to-end training of a deep neural network (DNN) for sentiment classification. In order to tightly integrate the common sensible information into the repeated encoder an associate extension of LSTM, termed Sentic LSTM, was proposed. The experiment was conducted on two public free data sets showing that the mixture of the planned attention design and Sentic LSTM outdid progressive strategies in ABSA tasks.

Moreover, Al-Smadi et al. (2017) presented Associate in Nursing approaches to tackle ABSA challenges of Arabic Hotels' reviews. Couple approaches of RNN and support vector machine (SVM) in conjunction with lexical, word, syntactic, morphological, and linguistics options were used. The projected approaches were evaluated using the SemEval 2016 data set for hotels' reviews. Analysis results showed that the SVM approach outperforms the opposite deep RNN approach. However, considering the execution time needed for coaching and testing the models, the deep RNN was quicker.

Lee et al. (2017) analysed the individuals opinion about the Deutsche Bahn AG. In order to achieve this objective the study used three methods: sentence embedding, an ensemble of classifiers for two sub-tasks, and state-of-the-art sequence taggers for two other sub-tasks. The data consisted of (22,000) German messages collected from different sources, including social media, web pages, and training word. This experiment indicated that the models performed better than the baseline.

Furthermore, Ma et al. (2017) examined the special treatment of target and contexts as well as their ability to learn representations via interactive learning. To achieve these objectives, the study proposed an IAN framework to interactively learn attentions within the contexts and targets, and produce the representations for targets and contexts separately. The IAN model represented a target and its collective context in an effective way that is useful to sentiment classification. The experimental done through victimization SemEval 2014 data sets. The results showed that there is an effect of the proposed model on the special treatment of target and contexts.

Moreover, Xu et al. (2017) provided a deep learning model for multilingual aspect-based SA, through using a CNN for extraction phase and SA. The proposed system consists of a vector for each word embedding and applies a convolution on it. In addition, the proposed system is an extension of the CNN-rand structure used by Kim. The study used the Yelp data set that includes restaurants and portable computer reviews. The result showed that feature-based SVM performs comparably well on the Yelp reviews, demonstrating the importance of powerful illustration for side level sentiment classification.

Further, Ouyang et al. (2015) conducted a study that aimed to represent a framework consisting of seven layers in order to analyse the sentiments of the sentences. This framework relies on CNN and Google Word2vec for SA and vector representation. The study used many methods to progress the correctness and generalizability of the proposed approach, including Dropout technology, Normalization and Parametric Rectified Linear Unit (PReLU). The study used a data set from "rottentomatoes.com" in order to verify the proposed framework. The data set consisted of movie review excerpts. After comparing the proposed framework with previous approaches like Matrix-Vector RNN (MV-RNN) and RNN, the results indicated that the proposed model outperformed the previous models by achieving 45.5 % accuracy.

In another study, Severyn and Moschitti (2015) utilized deep learning techniques for SA of tweets. The aim of the study was to initialize the load of parameters of CNN to train the model accurately while avoiding the need of adding new feature. A neural language was employed to initialize the word embedding and was trained by a huge unsupervised cluster of tweets. For additional purification, NN was employed to train the embedding on large supervised corpus, a standard. To initialize the network, antecedent embedded words and parameters were used, having same design and training on the supervised corpus of SemEval 2015. The activation functions employed are square measure activations, sentence matrix pooling, Softmax and convolutional layers. To train the network, random gradient descent (SGD) and non-convex operate improvement rules were used, while back propagation (GBP) algorithms were used to calculate the gradients. Dropout techniques were adapted to enhance the NNs regularization. The deep learning model was applied on the message level and the phrase level from SemEval 2015 to predict polarity. The model was found to perform well on term of accuracy.

Yanmei and Yuda (2015) proposed an outline of SA concerning Micro-blogs. The aim of this effort was to urge the opinions and attitudes of users regarding hot events by exploitation CNN. The employment of CNN overcomes the matter of express feature extraction and learns implicitly through coaching information. To gather the information from the target, the input universal resource locator and centred crawler were used. A thousand micro-blog comments were collected as a corpus and divided into 3 labels; 274 neutral emotions, three hundred negative emotions and 426 positive emotions. The planned model has been compared with the previous studies that used CRF and SVM. However, the performance proves that the planned model is affordable and efficient to boost the accuracy in terms of sentiment analysis.

Santos and Gatti (2014) aimed to effectively analyse sentiments in Twitter messages and movies' reviews. They proposed CNN that exploits from character-level to sentence-level information in order to analyse sentiments of short reviews or tweets. The proposed CNN model consisted of two convolutional layers and inducted with vectors generated by word embeddings. When considering the fault tolerance ability, the back propagation neural network (BPN) is chosen to compute gradients of the network. Segmentation of words, indexes calculations, CNN trainings and performance evaluations were conducted. Training sets were extracted from the data sets. The accuracy of model is 86.4% which indicates that the proposed technique has increased the performance of classification and saved training time.

Table 2-1 summarizes reviewed papers discussed in this section as well as the deep learning model used, the selected data set, special tasks performed and the results of each experiment.

Paper	**Deep leaning model**	**Data set**	**Special Tasks**	**Results**
(Chakravarthy et al. 2018)	Hybrid: CNNs and RNNs	SAR14 and Stanford Sentiment Treebank	Word embedding	Not Reported
(Ma, Peng & Cambria 2018)	LSTM	SemEval 2015 and SentiHood (Yahoo! Answers) (Saeidi et al. 2016)	Incorporating External Knowledge Base: termed Sentic LSTM	Acc. (SemEval): 76.47% Acc. (SentiHood): 89.32%
(Al-Smadi et al. 2017)	RNN	SemEval 2016 (Arabic Hotels' reviews)	Extracting syntactic, morphological, and semantic features. Word embedding	Acc. (RNN): 87%
(Lee et al. 2017)	LSTM	Deutsche Bahn AG	Embeddings: Word2Vec, Sent2Vec, and SIF	Macro F1: 55.4%
(Ma et al. 2017)	IAN	SemEval 2014 (Restaurant and Laptop)	Word embeddings	Acc. (Restaurant): 78.6% Acc. (Laptop): 72.1%
(Xu et al. 2017)	CNN	Yelp Data set (Restaurant and Computer)	Word embeddings	Acc. (Restaurant): 68.34% Acc. (Computer): 76.9%
(Ouyang et al. 2015)	CNN	rottentomatoes.com	Word embeddings and 7-layers architecture model for CNN	Acc.: 45.5%

(Severyn & Moschitti 2015)	CNN	SemEval 2015 (Twitter)	Word embeddings	Acc. (Phrase-level): 84.79% Acc. (Twitter-level): 64.59%
(Yanmei & Yuda 2015)	CNN	Micro-blog	classifier based on SVM/RNN	Acc.: 69%
(Santos & Gatti 2014)	CNN	Twitter messages and movies' reviews	Word embeddings. In addition, it exploits from character-level to sentence-level information	Acc.: 86.4%

Table 2-1: A Summary Table for Mentioned Deep Learning Approaches

3. CHAPTER THREE: METHODOLOGY

This dissertation pursues an experimental approach employing deep learning techniques. The purpose of this approach is to handle the task of aspect-based sentiment classification for government app reviews that include a variety of aspects. Deep learning is an arithmetic model that enables computers or machines to understand NLP with sufficient data to realize patterns and relationships based on logical reasoning. With the abundance of data and information obtained by the computer, the machine is supposed to "learn" or acquire new experience on how to communicate (Li et al. 2014).

The development of any sentimental application depends mainly on sentiment resources (Liu 2010). This means that the availability of data sets related to a customers’ reviews about a specific product or topic is necessary for conducting an assessment of their sentiments. Hence, this experiment is conducted on a high-quality and manually annotated data set. This data set resulted by an accurate annotation process described in (Alqaryouti, Siyam & Shaalan 2018) research. The original data set consists of 11,912 of Dubai government mobile apps reviews. This data set will be discussed further in the following section in addition to the pre-processing tasks that took place to incorporate the data set in the suggested deep learning approach.

This study uses a CNN framework which is a class of deep learning approach. CNN utilizes various multi-layer perceptrons built in a way to minimize pre-processing and feature engineering. The suggested CNN framework in this experiment consisted, in addition to the input and output layers, of three hidden layers: convolutional layer, pooling layer and classification layer. Further, word embeddings were incorporated in this experiment as the language representation model and a technique of learning review features.

This chapter discusses in details the reasons behind utilizing deep learning approach in the proposed methodology. Further, the data set used in in this experiment is described in details, in addition to the tools and technologies that supported designing and testing of different parts of this research. Also, it shows various pre-processing tasks performed as well as how reviews are represented and inducted into the deep learning model. Consequently, this chapter provides a discussion on the experiments that have been conducted and illustrates the proposed CNN framework. Figure 3-1 summarizes different stages of the experiment methodology.

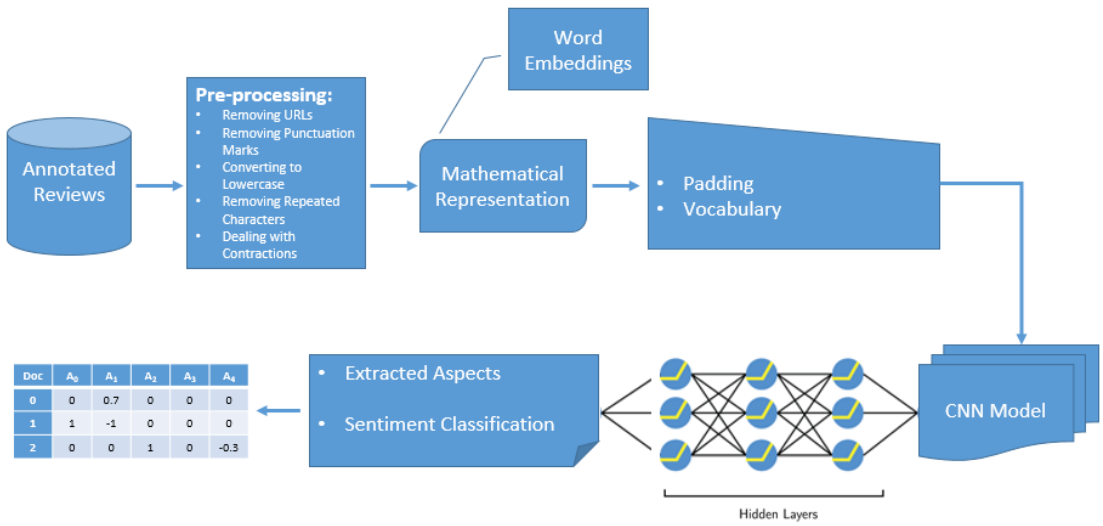

Figure 3-1: Summary of the Experiment Methodology

3.1. Data Set

This research adopted a high-quality manual annotated data set built as part of Alqaryouti, Siyam and Shaalan (2018) work in order to check if it is possible to improve the performance of aspect-based sentiment classification using deep learning. This data set comprised reviews of a considerable number of government apps used in the UAE, particularly in Dubai government. Although this manual approach consumes a lot of time and demands devoted resources, it is used because of its accuracy compared with the automated technique (Liu 2012).

The data set covers reviews of 60 smart government apps on both Apple Store and Google Play Store. The elicited reviews comprised the entire information gathered and posted during the years 2013 and 2016. The reason behind this amount of reviews, that 2013 formulates the year once His Highness ruler Mahammad bin Rashid declared the launch of the "Smart Government". The information collected comprised a collection of data as well as a substantial range of reviews within which each review shows the app name, store name, app ID, review ID, language, star rating, date, author, subject, and body.

The original data set comprises 11,912 reviews. The reviews were collected from Google Play and Apple stores. As a research procedure, the researcher stored the reviews in Microsoft Excel format. Figure 3-2 depicts the percentage of reviews based on the review language. This included both reviews in English Language (76%) and Arabic Language (20%) as well as Other

Languages (4%). Therefore, this experiment considered only English reviews in order to cover the majority of the reviews. In addition, the analysis of the data showed the most of the reviews were positive. Each review in the original data set was rated by the user based on a 5-stars rating. For simplicity, five and four stars rated reviews are considered positive in this experiment. The total number of positive reviews were about 7506 reviews. Three stars rated reviews are considered neutral and the total number of these reviews were about 691 reviews. Finally, one and two stars rated reviews were considered negative reviews and were about 3715 reviews. This indicates that individuals favour posting positive reviews compared to negative reviews.

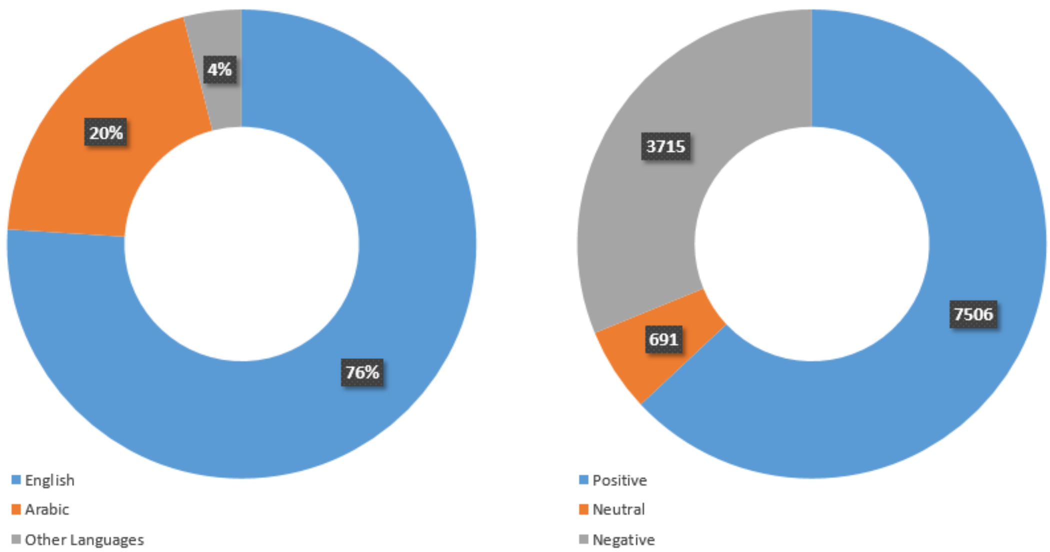

Figure 3-2: The Language and Star Rating Statistics of the Collected Reviews

During data analysis of the reviews, statistics showed that the average word count of the reviews was 9.23. In addition, the words count in the reviews ranged from 1 to 50. These short reviews show how mobile apps reviews is different than other domain such as movies reviews where users write longer reviews. Figure 3-3 illustrates the distribution of words count in the data set. This is important for specifying the padding sequence length of data. Padding will help in dealing with length variations of words counts in reviews.

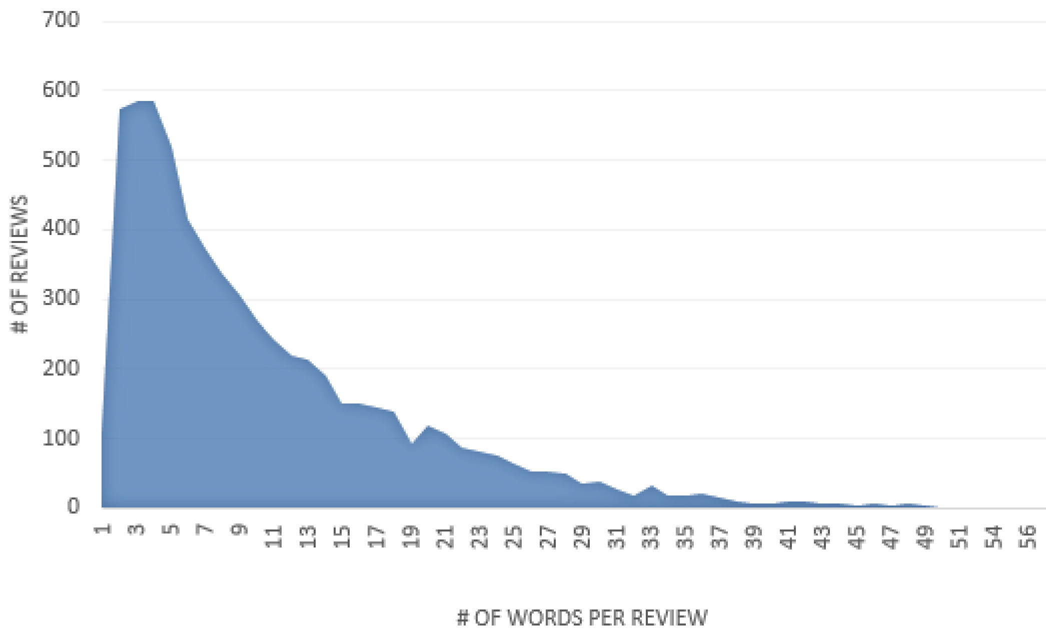

Figure 3-3: The distribution of words count in the data set

Alqaryouti (2017) emphasized on specific aspects that relate to achieving the customers' desires and needs. This can be achieved by providing an answer to the question "What will my customer remember about this experience?" Thus, Alqaryouti, Siyam and Shaalan (2018) suggested an original complete model to conduct their study. This model considers the state-of-the-art procedures as well as customers' aspects. Similarly, this experiment uses the same model, so it can benefit from the constructed data set and the annotation process done by Alqaryouti, Siyam and Shaalan (2018). This model contains five basic categories of aspects which are illustrated in **Error! Reference source not found.**.

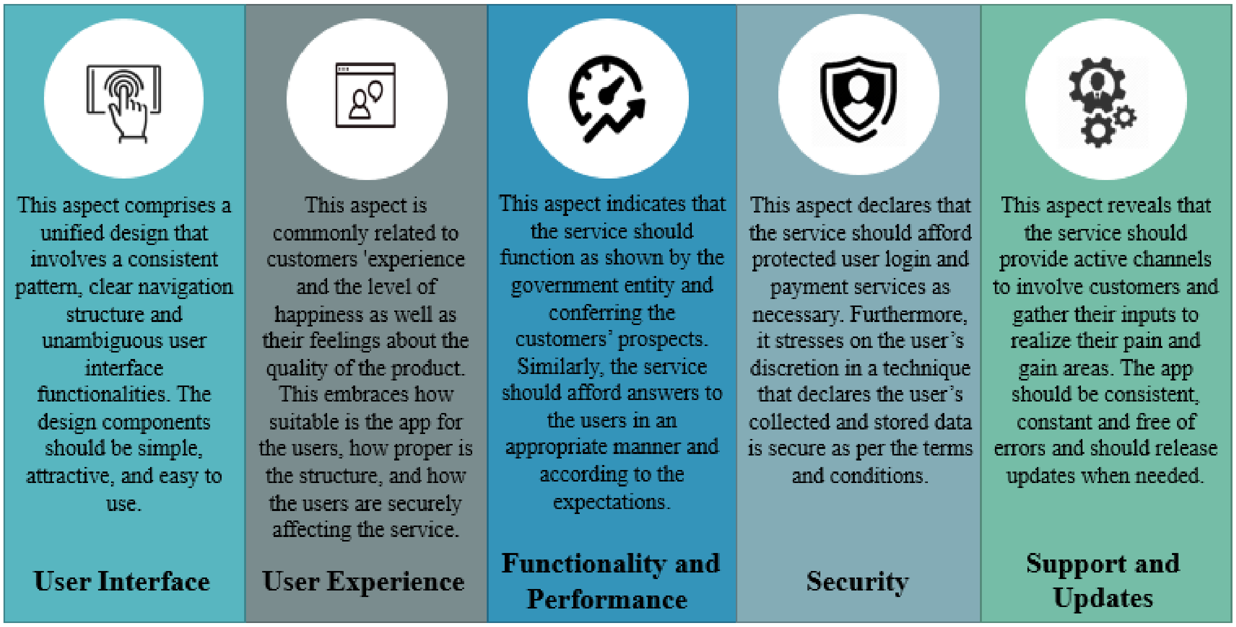

Figure 3-4: Basic categories of aspects used in our model (Alqaryouti, Siyam & Shaalan 2018)

The original data set is finally built in XML format that possesses a tag for every review. This tag comprises a review body, scores, subject, star rating, aspect terms, opinion words and the polarity score for every opinion word. Figure 3-5 depicts a sample review of the concluding original data set outcome in XML layout.

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<reviews>
    <review id="3662">
        <reviewbody>This App has very attractive design, but bad performance.</reviewbody>
        <reviewStarRating>4</reviewStarRating>
        <aspects>
            <aspect name="User Interface">
                <aspectTerm OpinionWord="attractive" term="UI" polarityStrength="0.354166667"/>
            </aspect>
            <aspect name="Functionality & Performance">
                <aspectTerm OpinionWord="bad" term="performance" polarityStrength="-0.604166667"/>
            </aspect>
        </aspects>
    </review>
</reviews>
</xml>
```

Figure 3-5: Sample review of the original data set that resulted from Alqaryouti (2017)

Since this experiment uses deep learning approaches, most of the features in the original data set are dropped. In addition, the review subject and body were combined together under one feature called "content" separated by a dot. Further, each review is labelled using a vector that describes the polarity of each aspect. For example, the differences of the label in the original data set and the resulted data set are illustrated in Figure 3-6.

Original Data set Label	[	User Interface	,	User Experience	,	Functionality & Performance	,	Security	,	Support & Update	]
	[	Positive	,	NA	,	Negative	,	NA	,	NA	]
Resulted Data set Label	[	1	,	0	,	-1	,	0	,	0	]

Figure 3-6: The representation differences between the original data set label and the resulted data set label

Finally, the resulted data set used in this experiment is in JSON format. Figure 3-7 shows a sample of review of the final data set.

```
{
  "reviewid": 5138,
  "features_content": "This App has very attractive design, but bad performance.",
  "Original_Rating": 4,
  "labels_index": "[1, 0, -1, 0, 0]"
}
```

Figure 3-7: Sample review of the final data set used in this experiment in JSON format

3.2. Tools and Technologies

In order to achieve the objectives of this research, the experiment employed a number of tools and technologies, including Python 3.6; Tensorflow 1.8 +; Numpy and Gensim, as described in Figure 3-8.

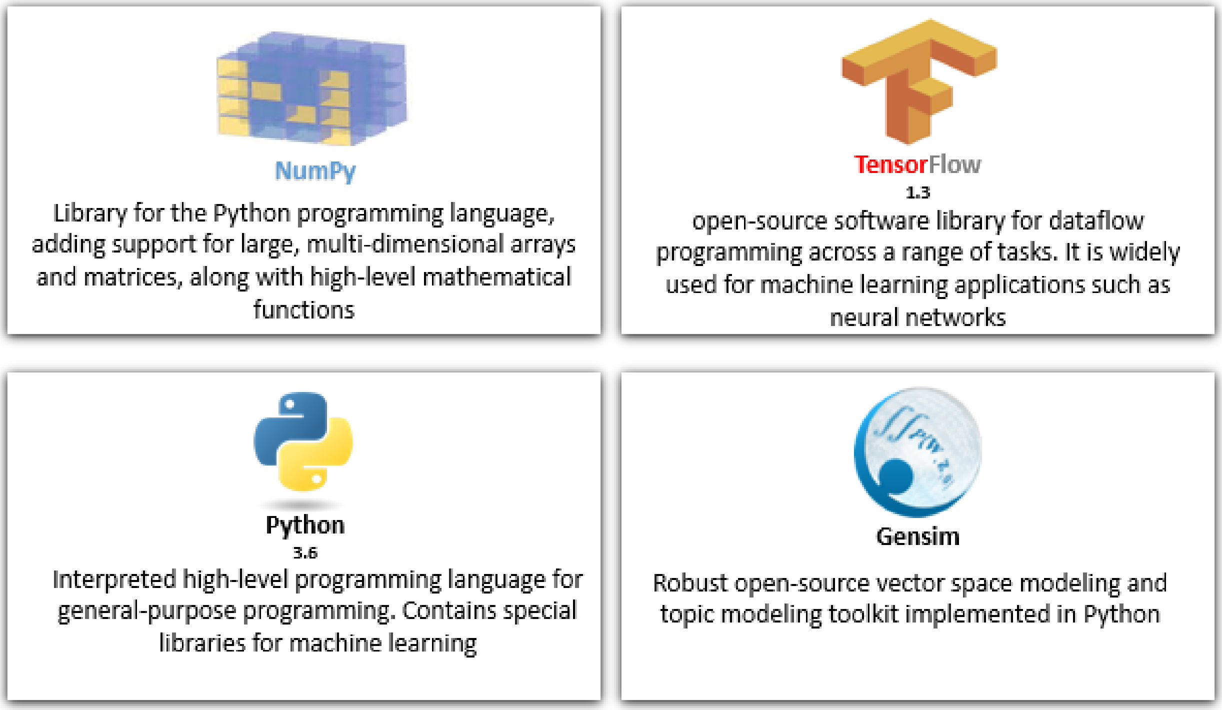

Figure 3-8: Tools and technologies used in this experiment

According to Rossum (1991) Python is one of the most effective programming languages used in different general functions. This language focuses on the code readability and the use of whitespace technique. One of the particular features of Python language is that it does not have specific libraries for machine learning such as Scipy and Numpy. These libraries are very significant for linear algebra process to recognize Kernel Methods of machine learning. This language is well-known for being accessible in machine learning algorithms and syntax related-aspects. Further, there are many other open-sources software beneficial in programming different tasks such as Tensorflow. This software is a symbolic math library. It is used in machine learning applications including NN.

Additionally, Tensorflow is very common in Google search, production and internal use. This was launched by Apache 2.0. On the other hand, NumPy is another library that is used for Python programming language. It is an add-value library that provides great multi-dimensional arrays and matrices. This library considers the enormous assortment of high-level mathematical functions in the operations of arrays. NumPy as a library was originated by Hugunin and many developers worked in developing it. In 2005, Oliphant established a new generation NumPy. This new version has some incorporated features such as Numarray. Another open-source vector space and topic modelling toolkit used in Python is Gensim. It implements NumPy, SciPy and optionally Cython in the performance process. This open-source vector space is used to treat large text collections. It uses data streaming and effective incremental algorithms to distinguish it from relevant scientific software applicable in this field.

3.3. Pre-processing

It is essential to conduct text pre-processing tasks in NLP domain especially when the text is pulled from social media platforms. Social media uses informal texts that includes many abbreviations, emotions and slangs. Mobile application reviews are not an exceptional case. That is, many pre-processing steps are used to take out any unnecessary information from reviews. It is very significant to apply text-processing techniques to deal with NLP tasks.

Initially, entities such as URLs are removed from the text because they do not provide any contribution to the given sentiment. Additionally, all the punctuation marks are taken away and all words are converted to lowercase. Stojanovski et al. (2015) attempted to provide actual meaning instead of using abbreviations by creating a social media abbreviations' dictionary. However, the function of such replacement has not shown any development. Moreover, the shortening of elongated words, found very commonly in social media reviews, is another pre-processing step. Social media users apply shortening to elongate words to bring focus on specific aspects or statement. The length words differ from one another at large extent but the same meaning is provided in either way. For instance, reviewing the word "happy" that is produced by another person as "haaaaapy". Since users do not have the interest to distinguish between these tokens, every elongated word is shortened to the maximum of about three-characters repetitions. Finally, informal contractions are modified to the formal form. For instance, "don't" is changed to "do not" and "they're" changed to "they are". **Error! Reference source not found.** summarizes the pre-p rocessing tasks that took place in this experiment.

Figure 3-9: Summary of pre-processing tasks took performed in this experiment

3.4. Word Vectorization and Language Models

Finding an illustration of acceptable features for the used data set is one among the foremost challenging elements in machine learning and NLP. This is because of the varied structures in human language at the level of words, sentences or reviews in the case of this research. One approach is to represent the review as a concatenation of the words in which NLP deals with them conventionally as indices. This method is called "one-hot vector" representation. This representation includes an outsized vector with zeroes and ones. Several problems arise when applying such approach. This is due to the weak scale of vectors when the vocabulary develops. In addition, words do not contain a semantic relationship.

Another famous technique for language modeling and feature learning in NLP is word embedding. This technique is useful in learning embedded words in an organized way throughout large text corpora and word representation to a lower dimensional vocabulary similar to what illustrated in Fig. 11. The vocabulary consists of n number of unique words. Each word in the vocabulary is mapped to a vector of real numbers. Each of these vectors has the same dimension or length.

this	v_1	0.9	0.3	-0.1	0.01	0.34	...	...	0.56
app	v_2	0.01	0.34	-0.34	0.4	0.9	...	...	0.01
has	v_3	-0.34	0.4	0.56	0.3	-0.1	...	...	0.4
very	v_4	0.34	-0.34	0.4	0.8	-0.34	...	...	0.7
attractive	v_5	0.21	-0.34	0.4	0.01	0.34	...	...	0.78
⋮					⋮				
theme	v_n	0.78	0.34	-0.34	0.4	-0.34	...	...	0.32

Figure 3-10: Vocabulary sample generated using word embedding technique

The main benefit is to provide word's similarities in terms of its syntactic structure and semantic meaning based on the closest texts or sentences. Each token in the review is organized as embedded words that can be compared. Earlier to the NN training, words embeddings are accumulated and researched. A look-up table is generated in a random way by initializing word embeddings. Gatti and Santos (2014) and Ifrim, Shi and Brigadir (2014) asserted that pre-trained word embeddings achieved better performance, which makes it suitable for this role. It is worth mentioning that two types of embedding words are being evaluated; GloVe Word Embeddings and Sentiment Specific Word Embeddings (Mikolov et al. 2013; Pennington, Socher & Manning 2014; Tang et al. 2014). Table 3-1 describes the word embeddings used in this research.

	SSWE	**GloVe**
Source	Tweets	Tweets
Corpora Tokens Count	10M	20B
Dimension	50	200
Vocabulary	137K	1.2M

Table 3-1: Pre-Trained word embeddings used in this research

Table 3-1 illustrates the length of feature vector described by the “Dimension”, whereas the total count of word's token in the data set used in the training of word embeddings’ is described by “Corpora Tokens Count”. Moreover, the “Vocabulary” is what describes the number of unique words. Table 3-2 describes the number of tokens existing in the look-up table for each word embedding in the data set.

Dataset Dictionary	**# of matched tokens**
SSWE	4,365
GloVe	4,432
Dataset	4,449

Table 3-2: Dictionaries in the Data set and Range of Matched Words' Tokens

Many different issues may arise when using word embeddings for sentimental analysis tasks in mobile apps reviews. This is because mobile reviews use informal language compared to other word embeddings that depend on corpora training and formal language. Consequently, there will be a lot of missing tokens resulting in meaningless representation. Based on this, word embeddings are trained on Twitter data to overcome such problem since informal language is employed for written reviews. In addition, the word embeddings model is trained using an unsupervised approach where no sentiment data is encoded within the model. Thus, words that expose similar settings such as the word “happy” and “sad” will be considered adjacent words based on cosine similarity. This similarity problem is approached by applying pre-trained word embeddings and updating them throughout implementing BP. BP helps encoding the classification error and sentiment regularities into the trained model in which “happy” will be no more be similar to “sad”, which allows building more constructive representations for words existing in the vocabulary.

Each review in the data set is represented by a sequence of vectors based on the utilized vocabulary. Since the lengths of reviews are different in most cases, we can crop the longer reviews and pad the shorter ones. In this experiment, we have decided that the padding length is 35 words based on the distribution of words count as illustrated in Figure 3-3 earlier. Figure 3-11 shows a padding sample for short reviews. On the other hand, any review with more than 35 words will be cropped.

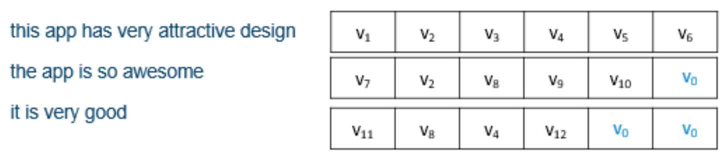

Figure 3-11: Padding sample for short reviews

3.5. Convolutional Neural Network

The present research uses CNN framework, which is a class of deep learning approach. CNN utilizes various multi-layer perceptrons built in a way to minimize pre-processing and feature engineering. The proposed CNN framework in this experiment consisted, in addition to the input and output layers, of three hidden layers: convolutional layer, pooling layer and classification layer as shown in Figure 3-12.

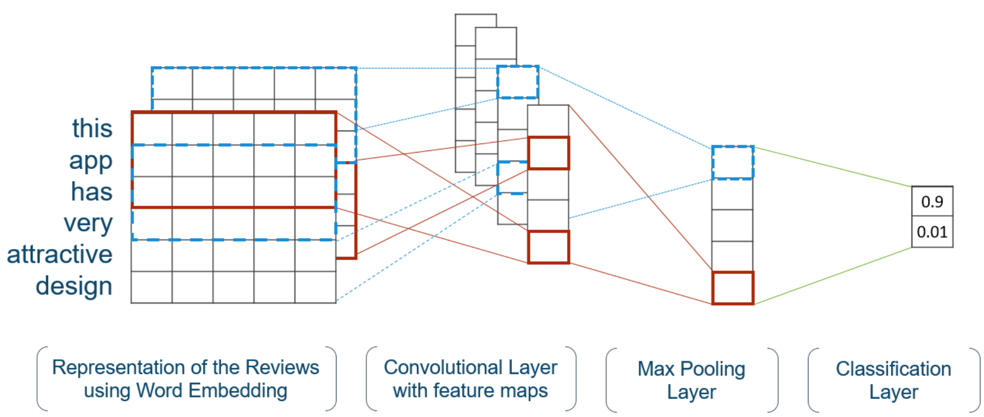

Figure 3-12: CNN framework used in this experiment

3.5.1. Convolutional Layer

The variable size of sentences is considered when using CNNs with pooling operation. Additionally, CNNs give enough attention to words’ order and context. This differs from the application of CNNs image recognition that considers the convolutional layers' as multiple levels. That is, one single layer is sufficient for understanding language and expressive features. For instance, the review of *(r)* with the length of *(n)* tokens is considered in such context. Using a proper padding at the end of the review is required because the filters within the convolutional layer are utilized in a sliding window manner. Padding length is referred to as *(h/2)* representing the filter's window size *(h)*. The look-up table will be constructed before the network training. Each word and token are presented. Hence, $L \in R^{k\times|V|}$, refers to the look-up table and *"K"* relates to the word vector's dimension and *"V"* stands for vocabulary. The mapping process of each word and token is conducted based on word embeddings' projection. Thus, the projection vector is $wi \in R^k$ *as* indicated in the following formula.

$$x = \{w_1, w_2, \dots w_n\}$$

As a result, various filters are used by the convolutional layer with different window sizes *(h)*, which depends on the presented feature maps as well as the review. This dissertation uses a window sizes of 2 and 3, since the average number of tokens in the reviews is 9.23. The feature maps can be achieved by the used repeated function applications throughout the sub-regions or windows in the review. Figure 3-13 illustrates the convolutional layer with two filters. The one-dimensional filter is convoluted with the inserted review, which adds a leaning concept and eventually uses a function that is non-linear. As a result, this experiment used hyperbolic tangent as an activation function.

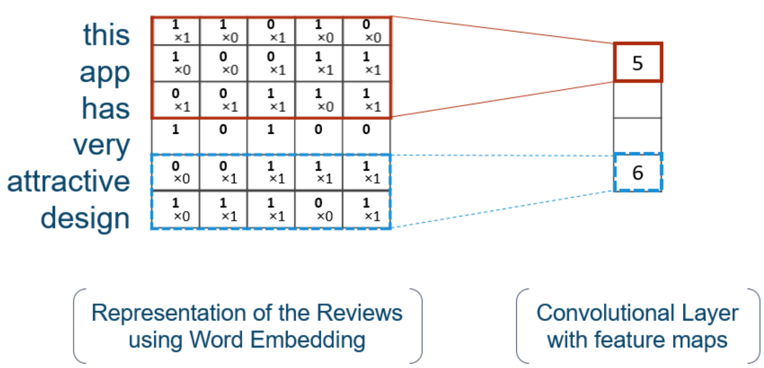

Figure 3-13: Illustration of the convolutional layer with two filters

3.5.2. Max-pooling Layer

Similar to the convolutional layer, the max-pooling layer creates an output based on the dimension of the input sequence. This layer releases the virtual importance of each vector and keeps the sentence size unchanged. Max-pooling takes the maximum value of each group of neurons from the previous layer.

3.5.3. Classification Layer

The classification layer is used to determine whether each extracted aspect in the review is positive or negative. In this research, two classifying algorithms were examined in order to compare their performance. The algorithms are Softmax regression classifier and support vector regression (SVR).

3.6. Summary

Deep learning is an arithmetic model that enables computers or machines to understand NLP with sufficient data to realize patterns and relationships based on logical reasoning. With the abundance of data and information obtained by the computer, the machine is supposed to "learn" or acquire new experience on how to communicate (Li et al. 2014). In this chapter, we described in details the data set selected for this experiment. The data set covers reviews of 60 smart government apps, resulting in 11,912 reviews. The pre-processing tasks performed as well as how reviews are represented and inducted into the deep learning model using word embeddings. Further, the CNN framework utilized in the methodology was demonstrated. The suggested CNN framework in this experiment consisted, in addition to the input and output layers, of three hidden layers: convolutional layer, pooling layer and classification layer. The results of different experiments of the model hyper-parameters along with the suggested pre-trained word embeddings are discussed in the next chapter.

Finally, it is important to highlight the reasons deep learning approach and CNN framework were chosen over traditional machine learning or rule-based approaches. Deep learning requires less human intervention and less effort in features engineering. It simulates neurons in the human body. Hence, it more capable and flexible in learning and representing the world as nested hierarchy of concepts while diving deeper into hidden layers. In each layer, the representation gets more abstract and simpler. Furthermore, deep learning can handle massive amounts of data compared to traditional machine learning approaches (Schmidhuber 2015).

4. CHAPTER FOUR: DISCUSSION OF EXPERIMENTAL RESULTS

As discussed in chapter three, the proposed framework utilizes pre-trained word embedding models as a language representation method. Furthermore, this framework consists of different layers. Each of these layers uses various hyper-parameters and activation functions that have various effect on the overall performance. In this chapter, we target to experiment several techniques and parameters to create precise results and to guarantee the in-depth understanding of the needs and expectations of customers. Further, we discuss the theoretical and practical implications of this study.

4.1. Discussion of Experimental Results

The experiment adopts the CNN framework to handle the task of ABSA in the mobile apps reviews domain. This task is defined as a classification problem and requires a labelled data set. Also, it is considered as a multi-label multi-class classification problem where three labels can be assigned to each aspect of the review, namely positive, negative or neutral sentiment. As stated in chapter three, this study used the well-constructed data set of 11,912 of Dubai government mobile apps reviews. This data set resulted from a high-quality manual annotation process conducted by Alqaryouti, Siyam and Shaalan (2018). In order to train and test the proposed model, the data set is split into two parts: 70% of the data inducted to train the model and 30% kept unseen for testing purposes similar to Alqaryouti (2017). This approach allows us to compare the performance results between rule-based approach followed in Alqaryouti's (2017) study and our experiment results which utilizes a deep learning approach.

This study explores different experiment settings; the effect of various pre-trained word embeddings, and the impact of hyper-parameters and activation functions on the accuracy of the proposed framework. Pre-trained word embedding models is used to generate a vocabulary as text representation to deal with some of the various difficulties in ABSA in particular and NLP tasks in general. As highlighted in section 3.4 in the previous chapter, two pre-trained word embedding models are tested and evaluated in this experiment; GloVe Word Embeddings and SSWE.

Our proposed model includes many parameters in need of adjustment to achieve better results. For instance, the batch size is adjusted to 256 and filters are set to window sizes of 2 and 3. The decay rate of the parameters in the network is set to 0.95. Moreover, the dropouts for the layers is set to 0.5 and the number of the hidden units in the convolutional layer is adjusted to 100. Table 4-1 summarizes the values of the main parameters of the network.

Parameter	**Value**
Learning rate	0.001
Padding length	35
Batch size	1024
Filters with window sizes	2, 3
Hidden units	100
Dropout keep probability	0.5
Decay rate	0.95
Decay steps	100
Number of training epochs	100

Table 4-1: Values of the main parameters of the network

In addition, we have conducted various experiments to figure out whether the activation function in the convolutional layer must be a hyperbolic Tanh or rectified linear activation function. Hyperbolic tangent activation function provides a better result compared to the rectified linear activation function.

Also, many varied classifying algorithms are experienced in the final layer of the network, namely SVR and Softmax regression. This study also explored the capability of combining Softmax regression and a SVR along with sigmoid function or with a hyperbolic tangent.

4.2. Performance Evaluation

In our experimental approach, we have adopted the confusion matrix technique in order to calculate the performance measures, including accuracy, precision, recall and F-measure for both aspect extraction and sentiment classification. **Error! Reference source not found.** defines each e lement in the confusion matrix employed to calculate the performance of aspect extraction regardless of the sentiment classification results. On the other hand, **Error! Reference source not f ound.** defines the elements required to measure the performance of sentiment classification of the extracted aspects.

		Predicated Value	
		Extracted	Not Extracted
Actual Value	Relevant	TP True Positive (Extracted and Relevant)	FP False Positive (Not Extracted but Relevant)
	Irrelevant	FN False Negative (Irrelevant but Extracted)	TN True Negative (Irrelevant and Not Extracted)

Table 4-2: Definition of Aspect Extraction Confusion Matrix Elemnts

		Predicated Value	
		Calculated	Calculated Incorrectly
Actual Value	Relevant	TP True Positive (Relevant and Calculated Correctly)	FP False Positive (Relevant but Calculated Incorrectly)
	Irrelevant	FN False Negative (Irrelevant but Calculated)	TN True Negative (Irrelevant and Not Calculated)

Table 4-3: Definition of Sentiment Classification Confusion Matrix Elements

Additionally, the following formulas are used to calculate different performance measures, namely precision, recall, F-measure and accuracy.

$$Precision = \frac{TP}{(TP + FP)}$$

$$Recall = \frac{TP}{(TP + FN)}$$

$$F - measure = \frac{2\ \times Precision\ \times Recall}{Precision\ + Recall}$$

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FP)}$$

Figure 4-1 illustrates the concluded performance measures with varied word embeddings and combinations of functions. According to the performance results, the use of the word embedding (GloVe) achieved better results than SSWE. The better performance of GloVe might be due to the use of two activation functions in the network model: the sigmoid and the hyperbolic tangent.

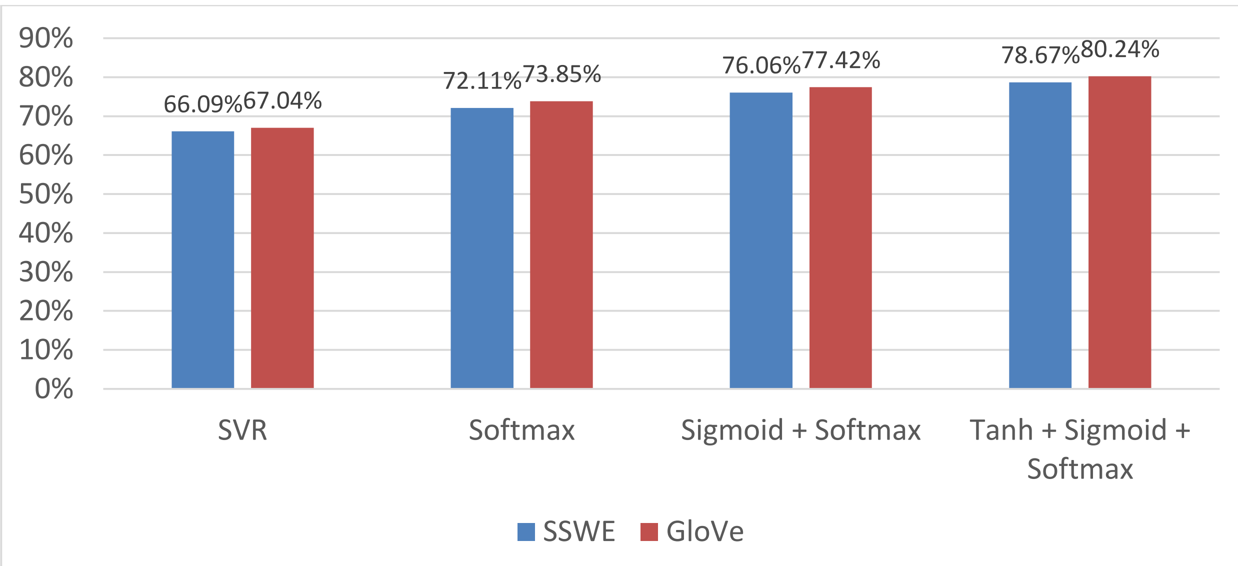

Figure 4-1: F-measure with varied word embeddings and activation functions

On the other hand, the evaluation results show that SVR implication decreases the performances rather than increases them, unlike the use of Softmax. Since the performance measures are lower when using SVR combined with other functions, SVR is not combined or reported with other functions. Combining other functions with Softmax regression classifier consistently improved the performance using both of the aforementioned word embeddings. Table 2-1 shows the results of different performance measures for various parameters settings.

Parameters	**Results**

Word Embedding	SVR	Softmax	Sigmoid	Tanh	Precision	Recall	F-measure	Accuracy
GloVe	√	-	-	-	73.98%	61.28%	67.04%	83.28%
GloVe	-	√	-	-	81.89%	67.24%	73.85%	86.42%
GloVe	-	√	√	-	85.07%	71.03%	77.42%	88.01%
GloVe	-	√	√	√	**87.31%**	**74.23%**	**80.24%**	**89.42%**
SSWE	√	-	-	-	73.01%	60.37%	66.09%	82.96%
SSWE	-	√	-	-	80.13%	65.55%	72.11%	85.79%
SSWE	-	√	√	-	83.70%	69.70%	76.06%	87.48%
SSWE	-	√	√	√	85.74%	72.68%	78.67%	88.78%

Table 4-4: The results of different performance measures for various parameters settings

Additionally, the evaluation showed significantly higher results when evaluating aspect extraction task alone regardless the sentiment classification results. Table 4-5 demonstrates these differences in performance.

	Precision	**Recall**	**F-measure**	**Accuracy**
Aspect Extraction	90.27%	77.17%	83.21%	90.69%
Aspect Extraction and Sentiment Classification	87.31%	74.23%	80.24%	89.42%

Table 4-5: Differences in performance results when evaluating aspects extraction task alone compared with the overall score

Finally, we have compared the performance of our deep learning approach with the rule-based approach by Alqaryouti (2017). This comparison is presented in Table 4-6. The results shows that the proposed approach by Alqaryouti (2017) slightly outperformed the deep learning approach in this experiment. This can be due to the fact that deep learning performs better with massive amount of data compared to the relatively small amount of data that was used to train the model, which did not exceed eight thousand reviews consisting of short sentences with few number words, averaging (9.23) words per review. However, there are several advantages of using deep learning. For instance, it does not necessitate human experience in the domain of the problem. In addition, it does not require much effort in identifying features for training compared with rule-based or traditional machine learning approaches.

	Precision	Recall	F-measure	Accuracy
Deep learning approach in this experiment	87.31%	74.23%	80.24%	89.42%
Rule-based approach by Alqaryouti (2017)	92.25%	79.59%	85.46%	91.64%

Table 4-6: Comparison of different performance measures between the rule-based approach by Alqaryouti (2017) and deep learning approach in this experiment

4.3. Theoretical and Practical Implications

The findings resulting from this experiment have several theoretical and practical implications. The theoretical implications include addressing of ABSA tasks and challenges through deep learning techniques, which is not widely investigated. In addition, this is the first attempt that employs a CNN framework to handle the task of ABSA for government app reviews. Further, considering the culture of Dubai citizens, mobile apps reviews are usually very short and include a variety of aspects in the same sentence. This raise the challenge level of extracting these aspects and classifying the sentiments.

On the other hand, this model can benefit the owners of mobile applications in general and government entities in particular. Due to the fast emerging in technologies, organizations need to listen to the public voice in an effective manner in order to attract customers to consume their services. Worldwide strategies are directing entities to focus on customers and design their services from their perspective. Our proposed model can have an essential role toward understanding citizens' points of view about offered services. It is an effective way to identify weakness points to focus on enhancing them in order to offer better services that can generate more revenues as a result.

4.4. Summary

This chapter discussed several techniques and parameters settings in order to create precise results and to guarantee the in-depth understanding of the needs and expectations of customers. Initially, the data set was divided into training and testing sets. Further, several experiments with different configurations were explored. The effect of various pre-trained word embeddings, hyper-parameters and activation functions were found to have a great effect on the performance. In addition, the confusion matrix approach was selected to evaluate the model performance. The performance results showed that (GloVe) outperformed other word embeddings models, in addition to the utilization of variety of activation functions, such as Sigmoid, Tanh and Softmax.

The performance results for the proposed deep learning approach were compared with the rule-based approach proposed by (Alqaryouti 2017). Finally, the theoretical and practical implications of this study were discussed.

5. CHAPTER FIVE: THE ANSWERS OF THE RESEARCH QUESTIONS

This chapter provides answers to the research questions along with the reference to the corresponding parts and details in the dissertation.

5.1. Research Question One

Question: How can deep learning be utilized in ABSA of government app reviews that include a variety of aspects?

Answer: As discussed in Chapter 4, the proposed CNN framework shows significant results in handling the task of ABSA in the mobile apps reviews domain. The framework utilizes pre-trained word embedding models as a language representation method. It was found that the word embedding (GloVe) outperformed other word embedding models. Since mobile reviews use informal language, it was recommended to use a word embedding model that is trained on a similar type of language such as Twitter, which is the case for (GloVe) Word Embedding model. In addition, what helps GloVe in achieving the best results is the use of two activation functions in a feed-forward manner: the sigmoid and the hyperbolic. Further, the proposed CNN framework consisted, in addition to the input and output layers, of three hidden layers: convolutional layer, pooling layer and classification layer. Different aspects of hyper-parameters and activation functions have been explored as well as their effect on the accuracy of the described framework. The evaluation process showed how the utilization of variety of activation functions, such as Sigmoid, Tanh and Softmax improved performance measures. To conclude, this approach shows the ability to create precise results that guarantee the in-depth understanding of customers' demands.

5.2. Research Question Two

Question: Can deep learning techniques improve the performance of aspect-based sentiment classification for government app reviews?

Answer: As discussed in chapter 4, the proposed deep learning model has achieved promising results depending on the use of word embeddings and different layers. It is important to state that this is the first attempt that employs a CNN framework to handle the task of ABSA for government app reviews. In addition, deep learning techniques are not widely investigated in ABSA domain.

The results showed that the proposed approach by (Alqaryouti 2017) slightly outperformed the deep learning approach in this experiment. This can be due to the fact that deep learning performs better with massive amount of data compared with the amount of data used to train the model, which did not exceed eight thousand reviews consisting of short sentences with few number words, averaging (9.23) words per review. However, there are several advantages of using deep learning. For instance, it does not necessitate human experience in the domain of the problem. In addition, it does not require much effort in identifying features for training compared with rule-based or traditional machine learning approaches.

6. CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS

This chapter discusses the conclusions and recommendations of the dissertation as well as the future work needed to be conducted in this regard.

6.1. Conclusion

Nowadays, public sectors are spending a lot of money to better understand their customers' demands. It is vital for public sectors to confirm that they offer the best services that fulfil the customers' needs. The process of analysing sentiments and opinions toward government services quality is subject to study by SA techniques. Further, ABSA can promote the process of understanding users' sentiments about different aspects of the provided services in an automated way in order to identify strengths and weaknesses of each aspect of the service. This study aimed to investigate the use of deep learning approach in ABSA of government app reviews that include a variety of aspects.

CNN framework, which is one of deep learning approaches, was selected for this study over traditional machine learning and rule-based approaches. Deep learning requires less human intervention and less effort in features engineering. It simulates how humans think and represent patterns. It has the ability and flexibility to learn and represent the world as a nested hierarchy of concepts while diving deeper into hidden layers. In each layer, the representation gets more abstract and simpler. Furthermore, deep learning can handle massive amounts of data compared to traditional machine learning approaches (Schmidhuber 2015).

The data set that was selected for this research includes 11,912 reviews of 60 smart apps on both Apple iOS and Google Play Stores offered by Dubai government to its citizens. The approach utilized word embeddings to represent the reviews and inducted them into the input layer of the deep learning model. The proposed CNN framework consisted of three hidden layers: convolutional layer, pooling layer and classification layer. Further, several techniques and hyper-parameters have been explored as well as their effect on the accuracy of our framework. The performance measures showed that (GloVe) outperformed other word embeddings models. Additionally, utilizing variety of activation functions, namely Sigmoid, Tanh and Softmax enhanced the model performance.

The accuracy of the described framework showed promising results where it has reached to 89.42% in term of accuracy. However, slightly better results were achieved using the rule-based approach proposed by Alqaryouti (2017). Deep learning performs better with massive amount of data, while the amount of data used to train the proposed model was less than eight thousand reviews. In addition, the majority of these reviews consisted of few numbers of words, which were on average (9.23). On the other hand, this study highlighted several advantages of using deep learning approach. For instance, it does not require much human experience in the domain of the problem compared to the rule-based approach. Further, it needs less effort in identifying features for training.

6.2. Limitations and Future Prospects

Recently, deep learning became an interesting topic that attracted researchers and organizations to benefit from its abilities. As highlighted in this research, deep learning illustrated promising capabilities to address ABSA tasks. However, there is still room for improvement and further investigation to achieve better performance results. For instance, acquiring a larger data set in the domain of government apps reviews was one of the main limitations in this experiment. Data sources of customers' opinions are now increasing rapidly. This requires to design a methodology to acquire this data in an efficient way. On the other hand, the platforms and technologies available are still limited and require large scale of computing resources. There is a lack of proper analysis capabilities to identify weakness points in ANN layers by employing visualization methods that show how the model is able to compose meanings in SA task. As extension to this research, we would like to explore the impact of adopting other deep learning techniques and parameters on the performance. For instance, utilizing a hybrid framework that benefits from the capabilities of both RNN and CNN.

References

Addo-Tenkorang, R. & Helo, P. T. (2016). Big data applications in operations/supply-chain management: A literature review. *Computers & Industrial Engineering*. Elsevier, vol. 101, pp. 528–543.

Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y. & Gupta, B. (2017). Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of Computational Science*. Elsevier B.V.

Al-Smadi, M., Talafha, B., Al-Ayyoub, M. & Jararweh, Y. (2018). Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. *International Journal of Machine Learning and Cybernetics*. Springer, pp. 1–13.

Alkhatib, M. & Shaalan, K. (2018). 'The key challenges for arabic machine translation'. , in *Intelligent Natural Language Processing: Trends and Applications*. Springer, pp. 139–156.

Alqaryouti, O. (2017). Aspect-Based Sentiment Analysis for Government Smart Applications Customers' Reviews. The British University in Dubai (BUiD).

Alqaryouti, O., Alqudah, A. & Shaalan, K. (2016). Effective Knowledge Management in Projects: Characteristics and Integration. *Information and Knowledge Management*, pp. 27–33.

Alqaryouti, O., Siyam, N. & Shaalan, K. (2018). A Sentiment Analysis Lexical Resource and Dataset for Government Smart Apps Domain. *International Conference on Advanced Intelligent Systems and Informatics*. Springer, pp. 230–240.

Anand, D. & Naorem, D. (2016). Semi-supervised Aspect Based Sentiment Analysis for Movies Using Review Filtering. *Procedia Computer Science*. Elsevier, vol. 84, pp. 86–93.

Asur, S. & Huberman, B. A. (2010). Predicting the future with social media. *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*. IEEE Computer Society, pp. 492–499.

Baecchi, C., Uricchio, T., Bertini, M. & Del Bimbo, A. (2016). A multimodal feature learning approach for sentiment analysis of social network multimedia. *Multimedia Tools and Applications*. Springer, vol. 75(5), pp. 2507–2525.

Baldi, P., Sadowski, P. & Whiteson, D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature communications*. Nature Publishing Group, vol. 5, p.

4308.

Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I., Bergeron, A., Bouchard, N., Warde-Farley, D. & Bengio, Y. (2012). Theano: new features and speed improvements. *arXiv preprint arXiv:1211.5590*.

Bengio, Y. (2013). Deep learning of representations: Looking forward. *International Conference on Statistical Language and Speech Processing*. Springer, pp. 1–37.

Bin, L., Quan, L., Jin, X., Qian, Z. & Peng, Z. (2017). Aspect-Based Sentiment Analysis Based on Multi-Attention CNN. *Journal of Computer Research and Development*, vol. 54(8).

Brychcín, T., Konkol, M. & Steinberger, J. (2014). Uwb: Machine learning approach to aspect-based sentiment analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 817–822.

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*. IEEE, vol. 31(2), pp. 102–107.

Cambria, E., Schuller, B., Liu, B., Wang, H. & Havasi, C. (2013). Statistical approaches to concept-level sentiment analysis. *IEEE Intelligent Systems*. IEEE, vol. 28(3), pp. 6–9.

Castellucci, G., Filice, S., Croce, D. & Basili, R. (2014). Unitor: Aspect based sentiment analysis with structured learning. *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pp. 761–767.

Chakravarthy, A., Desai, P., Deshmukh, S., Gawande, S. & Saha, I. (2018). HYBRID ARCHITECTURE FOR SENTIMENT ANALYSIS USING DEEP LEARNING. *International Journal of Advanced Research in Computer Science*, vol. 9(1).

Chiu, C., Chiu, N.-H., Sung, R.-J. & Hsieh, P.-Y. (2015). Opinion mining of hotel customer-generated contents in Chinese weblogs. *Current Issues in Tourism*. Taylor & Francis, vol. 18(5), pp. 477–495.

Deng, L. & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in Signal Processing*. Now Publishers, Inc., vol. 7(3–4), pp. 197–387.

GulfNews. (2013). *Dubai e-Government renamed Dubai Smart-Government* [online]. [Accessed 15 January 2018]. Available at: https://gulfnews.com/business/economy/dubai-e-government-renamed-dubai-smart-government-1.1199985.

Hai, Z., Chang, K., Cong, G. & Yang, C. C. (2015). An association-based unified framework for mining features and opinion words. *ACM Transactions on Intelligent Systems and*

Technology (TIST). ACM, vol. 6(2), p. 26.

Hamdan, H., Bellot, P. & Béchet, F. (2014). Supervised methods for aspect-based sentiment analysis. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 596–600.

Hasbullah, S. S., Maynard, D., Chik, R. Z. W., Mohd, F. & Noor, M. (2016). Automated content analysis: A sentiment analysis on malaysian government social media. *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication*. ACM, p. 30.

Ifrim, G., Shi, B. & Brigadir, I. (2014). Event detection in twitter using aggressive filtering and hierarchical tweet clustering. *Second Workshop on Social News on the Web (SNOW), Seoul, Korea, 8 April 2014*. ACM.

KhaleejTimes. (2014). *Dubai Municipality implements happiness index to evaluate services* [online]. [Accessed 15 January 2018]. Available at: https://www.khaleejtimes.com/nation/government/dubai-municipality-implements-happiness-index-to-evaluate-services.

Kim, E. H.-J., Jeong, Y. K., Kim, Y., Kang, K. Y. & Song, M. (2016). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. *Journal of Information Science*. SAGE Publications Sage UK: London, England, vol. 42(6), pp. 763–781.

Kiritchenko, S., Zhu, X., Cherry, C. & Mohammad, S. (2014). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pp. 437–442.

Kumar, P. & Jaiswal, P. (2016). A Comparative Study on Sentiment Analysis and Opinion Mining. *International Journal of Engineering & Technology IJET*, vol. 8(2), p. 938.

Laskari, N. K. & Sanampudi, S. K. (2016). Aspect Based Sentiment analysis. *IOSR Journal of Computer Engineering*, vol. 18(2), pp. 24–28.

Lee, J.-U., Eger, S., Daxenberger, J. & Gurevych, I. (2017). UKP TU-DA at GermEval 2017: Deep Learning for Aspect Based Sentiment Detection. *Shared Task on Aspect-based Sentiment in Social Media Customer Feedback*, p. 22.

Li, C., Xu, B., Wu, G., He, S., Tian, G. & Hao, H. (2014). Recursive deep learning for sentiment analysis over social data. *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume*

02. IEEE Computer Society, pp. 180–185.

Liu, B. (2010). Sentiment Analysis and Subjectivity. *Handbook of natural language processing*, vol. 2, pp. 627–666.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*. Morgan & Claypool Publishers, vol. 5(1), pp. 1–167.

Ma, D., Li, S., Zhang, X. & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. *IJCAI International Joint Conference on Artificial Intelligence*, pp. 4068–4074.

Ma, Y., Peng, H. & Cambria, E. (2018). Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM. *Proceedings of AAAI*.

Maxwell, A., Li, R., Yang, B., Weng, H., Ou, A., Hong, H., Zhou, Z., Gong, P. & Zhang, C. (2017). Deep learning architectures for multi-label classification of intelligent health risk prediction. *BMC bioinformatics*. BioMed Central, vol. 18(14), p. 523.

Mehta, A., Parekh, Y. & Karamchandani, S. (2018). 'Performance evaluation of machine learning and deep learning techniques for sentiment analysis'. , in *Information Systems Design and Intelligent Applications*. Springer, pp. 463–471.

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Muhammad, A., Wiratunga, N. & Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-based systems*. Elsevier, vol. 108, pp. 92–101.

Ouyang, X., Zhou, P., Li, C. H. & Liu, L. (2015). Sentiment analysis using convolutional neural network. *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*. IEEE, pp. 2359–2364.

Pateria, S. (2016). Aspect Based Sentiment Analysis using Sentiment Flow with Local and Non-local Neighbor Information. *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pp. 2635–2646.

Pennington, J., Socher, R. & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543.

Poria, S., Chaturvedi, I., Cambria, E. & Hussain, A. (2016). Convolutional MKL based multimodal emotion recognition and sentiment analysis. *Data Mining (ICDM), 2016 IEEE 16th International Conference on*. IEEE, pp. 439–448.

Qazi, A., Tamjidyamcholo, A., Raj, R. G., Hardaker, G. & Standing, C. (2017). Assessing consumers' satisfaction and expectations through online opinions: Expectation and disconfirmation approach. *Computers in Human Behavior*. Elsevier, vol. 75, pp. 450–460.

Rossum, G. (1991). Python: An Object Oriented Programming Language.

Ruder, S., Ghaffari, P. & Breslin, J. G. (2016). Insight-1 at semeval-2016 task 5: Deep learning for multilingual aspect-based sentiment analysis. *arXiv preprint arXiv:1609.02748*.

Saeidi, M., Bouchard, G., Liakata, M. & Riedel, S. (2016). SentiHood: targeted aspect based sentiment analysis dataset for urban neighbourhoods. *arXiv preprint arXiv:1610.03771*.

Sanchez, F. S. & Vazquez, A. M. (2014). Sentiment Analysis for e-Services. *2014 IIAI 3rd International Conference on Advanced Applied Informatics*. IEEE, pp. 42–47.

Santos, C. N. dos & Gatti, M. (2014). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. *Coling-2014*, pp. 69–78.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*. Elsevier, vol. 61, pp. 85–117.

Severyn, A. & Moschitti, A. (2015). Twitter sentiment analysis with deep convolutional neural networks. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, pp. 959–962.

Shaalan, K. (2017). Dataset Built for Arabic Sentiment Analysis. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2017*. Springer, p. 406.

Sheikhmohammed.ae. (2017). *Mohammed bin Rashid launches UAE Strategy for Artificial Intelligence* [online]. [Accessed 15 June 2018]. Available at: https://sheikhmohammed.ae/en-us/news/details?nid=25218&cid=.

Siddiqui, S., Monem, A. A. & Shaalan, K. (2016). Sentiment Analysis in Arabic BT - Natural Language Processing and Information Systems. E. Métais, F. Meziane, M. Saraee, V. Sugumaran, and S. Vadera (eds). . Cham: Springer International Publishing, pp. 409–414.

Siddiqui, S., Monem, A. A. & Shaalan, K. (2018). 'Evaluation and enrichment of arabic sentiment analysis bt - intelligent natural language processing: trends and applications'. , in K. Shaalan, A. E. Hassanien, and F. Tolba (eds). . Cham: Springer International Publishing, pp. 17–34.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*. JMLR. org, vol. 15(1), pp. 1929–1958.

Statista.Com, S. P. (2018a). *Global digital population 2018* [online]. [Accessed 11 June 2018]. Available at: https://www.statista.com/statistics/617136/digital-population-worldwide/.

Statista.Com, S. P. (2018b). *Number of apps in leading app stores 2018* [online]. [Accessed 15 June 2018]. Available at: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/.

Statista.Com, S. P. (2018c). *Smartphone users in the United Arab Emirates 2014-2019* [online]. [Accessed 28 June 2018]. Available at: https://www.statista.com/statistics/494619/smartphone-users-in-united-arab-emirates/.

Stojanovski, D., Strezoski, G., Madjarov, G. & Dimitrovski, I. (2015). Twitter sentiment analysis using deep convolutional neural network. *International Conference on Hybrid Artificial Intelligence Systems*. Springer, pp. 726–737.

Al Suwaidi, H., Soomro, T. R. & Shaalan, K. (2016). Sentiment Analysis for Emiriti Dialects in Twitter. *Sindh University Research Journal-SURJ (Science Series)*, vol. 48(4).

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T. & Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1555–1565.

uaeai.ae. (2017). *UAE Artificial Intelligence Strategy* [online]. [Accessed 15 June 2018]. Available at: http://www.uaeai.ae/en/.

Wagner, J., Arora, P., Cortes, S., Barman, U., Bogdanova, D., Foster, J. & Tounsi, L. (2014). DCU: Aspect-based Polarity Classification for SemEval Task 4. *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, (SemEval), pp. 223–229 [online].Available at: http://link.springer.com/10.1007/978-3-642-19460-3.

Wiebe, J., Wilson, T. & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*. Springer, vol. 39(2–3), pp. 165–210.

Xu, L., Lin, J., Wang, L., Yin, C. & Wang, J. (2017). Deep convolutional neural network based approach for aspect-based sentiment analysis. *Adv Sci Technol Lett*, vol. 143, pp. 199–204.

Yadollahi, A., Shahraki, A. G. & Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*. ACM, vol. 50(2), p. 25.

Yanmei, L. & Yuda, C. (2015). Research on Chinese Micro-Blog Sentiment Analysis Based on Deep Learning. *Computational Intelligence and Design (ISCID), 2015 8th International Symposium on*. IEEE, pp. 358–361.

Zaslavsky, A. & Georgakopoulos, D. (2015). Internet of things: challenges and state-of-the-art solutions in internet-scale sensor information management and mobile analytics. *Mobile Data Management (MDM), 2015 16th IEEE International Conference on*. IEEE, pp. 3–6.

Zhang, L. & Liu, B. (2016). 'Sentiment analysis and opinion mining bt - encyclopedia of machine learning and data mining'. , in C. Sammut and G. I. Webb (eds). . Boston, MA: Springer US, pp. 1–10.