Mona Maghraby

بسم اللّه الرحمن الرحيم

مركز الشبكات وتكنولوجيا المعلومات

قسم التوثيق الإلكتروني

Mona Maghraby

جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

قسم

نقسم باللّه العظيم أن المادة التي تم توثيقها وتسجيلها

علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات

AIN SHAMS UNIVERSITY

FACULTY OF ENGINEERING

Computer Engineering and Systems

Improved Sentiment Analysis Approach Using Deep Learning

A Thesis submitted in partial fulfilment of the requirements of the degree of

Master of Science In Electrical Engineering

(Computer Engineering and Systems)

By

Sarah Abdul-Aziz Mahmud Abdu

Bachelor of Science In Electrical Engineering

(Computer Engineering and Systems)

Faculty of Engineering, Ain Shams University, 2018

Supervised By

Prof. Dr Ashraf Salem

Prof. Dr Ahmed Hassan Youssef

Cairo - (2022)

AIN SHAMS UNIVERSITY
FACULTY OF ENGINEERING
Computer and Systems

Improved Sentiment Analysis Approach Using Deep Learning

By
Sarah Abdul-Aziz Mahmud Abdu

Bachelor of Science In Electrical Engineering
(Computer Engineering and Systems)
Faculty of Engineering, University, 2018

Examiners' Committee

Name and Affiliation	Signature
Prof. Sherif Gamal Aly Computer Engineering and Systems American University in Cairo (AUC)	…….
Prof. Hossam El-Din Hassan Abd El Munim Computer Engineering and Systems Faculty of Engineering, Ain Shams University	…….
Prof. Ashraf Elfarghaly Salem Computer Engineering and Systems Faculty of Engineering, Ain Shams University	…….
Prof. Ahmed Hassan Youssef Computer Engineering and Systems Faculty of Engineering, Ain Shams University	…….

Date: 21 March 2022

Statement

This thesis is submitted as a partial fulfilment of Master of Science in Electrical Engineering, Faculty of Engineering, Ain shams University.

The author carried out the work included in this thesis, and no part of it has been submitted for a degree or a qualification at any other scientific entity.

Student name

Sarah Abdul-Aziz Mahmud Abdu

Signature

……..……

Date: 21 March 2022

Researcher Data

Name	: Sarah Abdul-Aziz Mahmud Abdu
Date of birth	: 10/9/1994
Place of birth	: Egypt
Last academic degree	: Bachelor in Electrical Engineering
Field of specialization	: Computer and Systems
University issued the degree	: Ain Shams University
Date of issued degree	: July 2018
Current job	: Teaching Assistant at the Faculty of Engineering, Ain Shams University

Abstract

In the recent years, deep learning has emerged as a powerful machine learning technique to employ in the field of multimodal sentiment analysis; many deep learning models and various algorithms have been proposed in this field. This research has two main contributions. First, it tackles a comprehensive overview of the latest updates in the field of multimodal sentiment analysis. We present a sophisticated categorization of thirty-five state-of-the-art models, which have recently been proposed in the field, into eight categories based on the architecture used in each model. The effectiveness and efficiency of these models have been evaluated on the most two widely used datasets in the field, CMU-MOSI and CMU-MOSEI. After carrying out an intensive analysis of the results, we eventually conclude the most powerful architecture in multimodal sentiment analysis task. We also provide a brief summary of the most popular approaches that have been used to extract features from multimodal videos in addition to a comparative analysis between the most popular benchmark datasets in the field. We expect that these findings can help newcomers to have a panoramic view of the entire field and will guide them easily to the development of more effective models. Second, this research also proposes a novel framework for multimodal sentiment analysis using semi supervised GANs, which we name "MuSA-GAN". More specifically, we tailor Improved-GAN with one of the most robust models in multimodal sentiment analysis - MMUUBA model - in a single framework. As far as our knowledge, we are the first to employ semi-supervised GANs for this task. We evaluate the effectiveness and efficiency of the proposed framework and compare it against a broad range of state-of-the-art baselines on both CMU-MOSI and CMU-MOSEI datasets. The conducted experiments demonstrate that tailoring Improved-GAN with MMUUBA on either datasets can achieve unprecedented improvement in classification accuracy when compared against the performance of isolated MMUUBA. In particular, MuSA was found to improve the 7-class accuracy by 5.21% on CMU-MOSI and by 10.04% on CMU-MOSEI. Different from that, we explore a much more complex task of extremely low data regimes: besides the given small amount of labelled samples, no data from similar distributions is assumed to be available throughout training. After carrying out intensive analysis of the results, we can assert with confidence that MuSA can learn from the small set of labelled examples and somehow maximize the usage of unlabelled examples in order

to find semantically meaningful information that helps it characterize the underlying data distribution. Such information helps the classifier trained using only a few labelled examples to generalize to parts of the data distribution that it would otherwise have no information about. This research also tackles a comprehensive overview of the latest updates in the field of generative adversarial networks. Further, we present a sophisticated categorization of all semi-supervised GANs, which have recently been proposed, into two categories based on the architecture used in each model. We expect that these findings will provide helpful insights to the development of more effective models in the field of multimodal sentiment analysis.

Thesis Summary

The thesis is divided into an introduction and 6 chapters.

The introduction presents the importance of multimodal sentiment analysis (visual, acoustic and linguistic) and the role of deep learning in solving this problem. Then it summarizes the contribution of the thesis in solving the problem by proposing a new framework to solve the problem of multimodal sentiment analysis using generative adversarial networks.

The first chapter tackles a comprehensive overview of the latest updates in the field of multimodal sentiment analysis. Then it presents a sophisticated categorization of thirty-five state-of-the-art models, which have recently been proposed in video sentiment analysis field, into eight categories based on the architecture used in each model. The effectiveness and efficiency of these models are evaluated on the most two widely used datasets in the field, CMU-MOSI and CMU-MOSEI.

The second chapter tackles a comprehensive overview of generative adversarial networks (GANs) and some of its successors (DCGAN and CGAN) in order to help readers to have a panoramic view of the entire field.

The third chapter gives an overview of semi-supervised learning and semi-supervised GANs. It categorizes all semi-supervised GANs, which have recently been proposed, into two categories based on the architecture used in each model, manifesting the strengths and weaknesses of each architecture.

The forth chapter proposes a novel framework for multimodal sentiment analysis using semi supervised GANs, which we name "MuSA-GAN". The details the architecture of the proposed framework and the corresponding training algorithm are discussed in this chapter.

In the fifth chapter, the effectiveness and efficiency of the proposed framework are evaluated on the most popular datasets in the field, CMU-MOSI and CMU-MOSEI.

The sixth chapter summarizes the results that the researcher has found and presents the conclusions that researcher has found based on her experiments. It also gives

recommendations regarding the future work that can be done in order to improve the results obtained.

Acknowledgment

I would like to deeply thank my supervisors, Prof. Dr. Ashraf Salem and Prof. Dr. Ahmed Hassan Yousef for their unconditional support and motivation. Without their assistance and dedicated involvement in every step throughout the process, this research would have never been accomplished. I would also like to show my endless gratitude to my family. Getting through my dissertation required more than academic support, and I have to thank my family for listening to me, encouraging me and having to tolerate me over the past three year.

March 2022

Table of Contents

Introduction

Sentiments play a very important role in our daily lives. They help us to communicate, learn and make decisions, that's why over the past two decades, AI researchers have been trying to make machines capable of analyzing human sentiments. The early efforts in sentiment analysis have focused on textual sentiment analysis where only words are used to analyze the sentiment. However, textual sentiment analysis is insufficient to extract the sentiment expressed by humans; the meaning of words and sentences spoken by speakers often changes dynamically according to the non-verbal behaviours [5]. For example if someone said the word 'Amazing', it could express negative sentiment if it is accompanied by a sarcastic laugh or sarcastic voice.

Intensive research over many years have shown that multimodal systems are more efficient in recognizing the sentiment of a speaker than unimodal systems. The way humans naturally communicate and express their emotions and sentiments is usually multimodal: the textual, audio, and visual modalities are concurrently fused to extract information conveyed during communication in an effective way. A survey of multimodal sentiment analysis published in 2015 [6] reported that “multimodal systems were consistently (85% of systems) more accurate than their best unimodal counterparts, with an average improvement of 9.83%”. Also, several surveys published later suggest that textual information is not sufficient for predicting human sentiments especially in cases of sarcasm or ambiguity [7-10]. For example, it is impossible to recognize the sentiment of a sarcastic sentence “Great" as negative considering only the textual information. However, if the system can access the visual modality, it can easily detect the unpleasant gestures of the speaker and would classify it with the negative sentiment polarity. Similarly, acoustic features play important roles in the correctness of the system. In 2018 Poria et al. [11] also introduced an intuitive explanation of the improved performance in the multimodal scenario by visualizations of MOSI dataset [12] where both unimodal features and multimodal features are used to give information regarding the dataset distribution (see Fig. 0-1). For the textual modality only, comprehensive clustering can be seen with substantial overlap. However, this overlap is reduced in multimodal [11].

With the recent growth in social media platforms and the advances in technology, people started recording videos and uploading them on social media platforms like YouTube or Facebook to inform subscribers about their views. These videos may be