Ain Shams University
Faculty of Computer and Information Sciences
Computer Science Department

Unsupervised Emotion Detection From Text.

Thesis submitted as a partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences

By

Salma Mohamed Osama Elgayar

Teaching Assistant at Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Under Supervision of

Prof. Dr. Zaki Taha Fayed
Professor in Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

Dr. Abdelaziz Abdelhamid
Assistant Professor in Computer Science Department,
Faculty of Computer and Information Sciences,
Ain Shams University

January– 2019
Cairo

Acknowledgements

There are several people who stood by me through this journey of my research; my beloved husband, my great father and mother, this work would not have been accomplished without their support.

I would like also to express my appreciation and gratitude to all those who helped me to complete this research specially Professor Doctor Zaki Taha for his encouragement, guidance and kindness from day one until today has enabled me to not only understand my subject but also work on it with such enthusiasm. I would like to express my appreciation to Doctor abdel aziz for his constructive suggestions ideas during the planning and development of this research work, also his valuable feedback.

Abstract

Artificial intelligence is not only the ability of a machine to think or interact with end user smartly but also to act humanly and rationally. Emotion detection from text plays a key role in human-computer interaction which is a main field in affecting computing. Lately emotion detection from text has attracted the attention of many researchers. That is due to the great revolution of emotional data available on social and web applications of computers and much more in mobile devices. Although some past approaches focused on addressing emotion from text, but still there is less effort on completely unsupervised direction emotion detection from text.

This research proposes a completely unsupervised model for textual emotion detection using hybrid technique of lexicon and word embedding concept. The proposed model represents sentences and their meanings in terms of word vectors. To enhance the overall accuracy, emotion ratios were assigned to short sentences and word lexicon.

The proposed approach has been validated using the International Survey on Emotion Detection Antecedents and Reactions(ISEAR) and twitter datasets. The evaluation results show that the proposed approach successfully classifies ISEAR sentences based on hybrid technique of lexicon and word embedding with and overall accuracy of 81% which is pretty good result comparing to other unsupervised techniques.

List of Publications

1. Elgayar, Salma, Abdel Aziz A.Abdelhamid, and Zaki.T.Fayed. "Unsupervised Emotion Detection from text: Survey" In International Organization of Scientific Research (IOSR) Journal of Computer Engineering, Volume 19, Issue 4, pp 30-37. August 2017.

2. Elgayar, Salma, Abdel Aziz , and Zaki.T.Fayed "Unsupervised Emotion Detection from text using Word embedding", The Eighteenth Conference on Language Engineering, December 2018.

Contents

List of Tables

List of Figures

Abbreviations

ANEW	**A**ffective **N**orm **E**nglish **W**ords
CBOW	**C**ontinuous **B**ag **O**f **W**ords.
DAL	**D**ictionary **A**ffect **L**anguage
FN	**F**alse **N**egative.
FP	**F**alse **P**ositive.
ISEAR	**I**nternational **S**urvey of **E**motions **A**ntecedents and **R**eactions.
LIWC	**L**inguistic **I**nquiry **W**ord **C**ount .
LSA	**L**atent **S**emantic **A**nalysis.
NLP	**N**atural **L**anguage **P**rocessing
NMF	**N**on-negative **M**atrix **F**actorization.
OCC	**O**rtony **C**lore **C**ollins.
PAD	**P**leasure **A**rousal **D**ominance
PMI	**P**ointwise **M**utual **I**nformation
TF-IDF	**T**erm **F**requency **I**nverse **D**ocument **F**requency.
TN	**T**rue **N**egative.
TP	**T**rue **P**ositive.
PLSA	**P**robabilistic **L**atent **S**emantic **A**nalysis.
SVD	**S**ingular **V**alue **D**ecomposition.
UTF	**U**nicode **T**ransformation **F**ormat
VSM	**V**ector **S**pace **M**odel.

Chapter 1

Introduction

Chapter 1

Introduction

1.1 Overview

Recently, with the huge growth of communications through various types of Internet applications and sites there are enormous amount of data transformation, which can lead to many misunderstanding from the textual conversation due to the lack of emotional communication and interaction. The importance and potential role in emotion detection from text has been emerged.

Facial expressions and vocal tone, can be easier to some extent, to detect the underlying emotional tone of the speaker. When he/she state a sentence. But imagine all these aforementioned features were removed away and all we had were the words only [1], therefore Emotions are complex, ambiguous and easily misunderstood entities, and sophisticated.

In the survey in [2], authors presented the recent advances in emotion models, and techniques. to answer some of the research questions such as how people feel when they read written text, how writers could transfer their emotional feelings to the readers through the text, and what is the best way to write emotional text to send clear message?. This lead to the motivation of more human-computer interaction [3].

In this thesis, we are trying to answer to those questions, taking into consideration the emotion is a strong person feeling that describes a state of temper such as happiness, sadness, fear and so on.

1.2 Motivation

Most of current researches in emotion detection are mainly based on the supervised approach of the emotion detection from text [4, 5]. This approach usually requires large annotated training data regardless it needs a lot of time and effort to train large dataset to get high accuracy. In addition, the model that is trained on dataset of specific domain does not work well in different ones. On the other hand, unsupervised learning can give us a solution to these difficulties. Therefore some other contributions were directed to use unsupervised approach from text but only detecting polarity of text *(positive or negative)* such as the work presented in [6].

Our proposed model is based mainly on unsupervised approach to detect up to seven ISEAR's(International Survey Antecedents and Reactions) emotion categories, which is a hot topic nowdays as it have many real-word applications. The proposed approach is realized in an application which is capable of getting the percentage of each emotion presented in a sentence ("Soft information") or assigning the sentence to the most dominate emotion percentage ("hard information). It also obtains the sentimental analysis of negative emotions such as Disgust, Sadness,Anger , Shame and Fear categories or positive emotion such as Happiness category.

One of the common problems with current emotion detection techniques that they evaluate each word independently without considering the sentence context although the same word could have different emotion in a different context. Therefore they may assign wrong emotion label to a sentence. Solving this problem in this research by using word embedding calculations to calculate word vector based on the other context words in the same sentence.

1.3 Objectives

There are rapid work in emotion detection field of Natural Language Processing (NLP). New technologies has been inspired to be applied into real word applications; such as marketing, customers reviews, health care and even political direction.

Detecting emotions, opinion or sentiment from text, can be overlapping. However, some efforts in this domain were employed and resulted in many information retrieval applications [7]. This allows businesses, researchers, governments, politicians and organizations to care about peoples sentiments and their emotions. These gains play a key role in decision making processes such as:

- Businesses realized the potential of emotional advertisements as emotions affect users so they tend to buy brands and get more services.
- Corporations want to evaluate their services, products and customers text feedback.
- Emotional mining in text works for automatic answering and in chat box or dialog system based on your current mood.
- Sentiments analysis concerning voters and public opinion can be extracted from politicians tweets [8].
- E-learning, online classes teachers will be able to connect in a better way with their students by automatically identifying their current affective.
- As a result of social media we can easily using emotion detection from text to discover a person who is passing by deep depression through his online interaction and prevent him from suicide.

1.4 Methodology

The entire process can be summarized as sentence preprocessing, lexicon detection, calculation mean of sentence's vector and finally get the nearest cosine distance from each emotion's vector. The preprocessing task consists of minimization rules, replace apostrophes if exists, then apply tokenization, remove stop words, correcting wrong syntax words if found, remove non-english words, emoticon replacement, and finally word stemming, other processes will be explained later through thesis.

1.5 Contribution

This research uses context based approach by calculating word's emotional ratio according to its neighbor words before classifying the sentence.

We can summarize our main contributions of this research as follow:

- We present a completely unsupervised approach for emotional detection. Mainly working on sentence level, but can be also extended to paragraph and document level as well.
- The proposed methodology does not require any training dataset before classification.