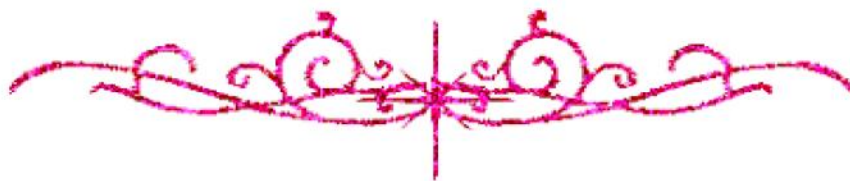


بسم الله الرحمن الرحيم





شبكة المعلومات الجامعية التوثيق الالكتروني والميكرو فيلم



جامعة عين شمس

التوثيق الإلكتروني والميكروفيلم

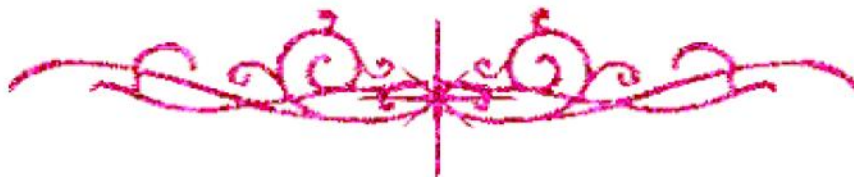
قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأقراص المدمجة قد أعدت دون أية تغيرات



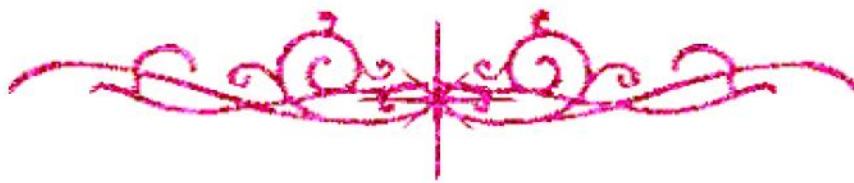
يجب أن

تحفظ هذه الأقراص المدمجة بعيدا عن الغبار



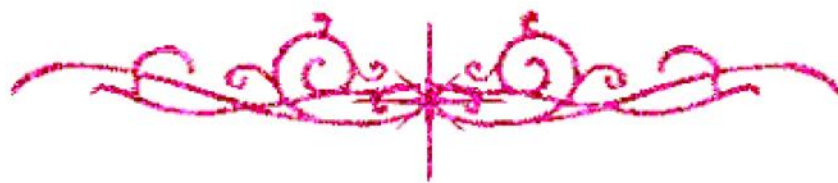


بعض الوثائق الأصلية تالفة





بالرسالة صفحات
لم ترد بالأصل





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

Deep Learning for Traits Detection Using Social Networks Interactions

A Thesis submitted to Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, in partial fulfillment of the requirements for the degree of Master of Science in Computer and Information Sciences.

By:

Eman Mohammed Hamdi El-Sayed Ibrahim

Supervised by:

Prof. Dr. Mostafa Mahmoud Aref

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

Dr. Sherine Rady Abdel Ghany

Associate professor at Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University

Cairo, Egypt
2020

Acknowledgment

First, I am grateful to Almighty ALLAH (S.W.T), the most Generous and the most Merciful. I thank God for providing me with strength and patience to complete this work.

Second, I would like to express my sincere gratitude to my supervisors; Prof. Dr. Mostafa Aref for his support, encouragement and guidance and Dr. Sherine Rady for all the support, patience and special supervision.

Third, I would like to thank my family and my husband for the love and care they give me constantly. Thank you for being with me through the good and the bad times.

My dear friends who have kept on encouraging me to get this work done; Eman Reda, Fatima El-Zahraa, Ghada Hamed, Hadeer Elsaadawy, Hanan Hindy and Marwa Salah. It was a precious journey with your companionship.

At last, I would like to thank all my professors, colleagues and students who believed in me. Thank you for being in my life.

Abstract

Social media networks are one of the main platforms used on a daily basis by millions of people. Feelings, emotions, and opinions are expressed by posting text, images, and videos to express self or to communicate with other people. Using text, the words reflect emotions and indicate behavior towards different topics. Detecting emotions and sentiments helps in many directions such that marketing, political orientation and product reviews. A huge amount of textual data is currently available. There is a need for detecting emotions from social media networks which enhances the machine understanding of humans' perspective.

This thesis contributes to detect positive and negative emotions from short text -tweets- by proposing a deep Convolutional Neural Network (CNN) using different types of word embeddings. CNN is the main building block of the proposed model and is responsible for extracting high-level features from low-level features. Word embeddings are the features fed to the model. The model is constructed by three CNN streams, where each CNN stream contains an embedding layer, a convolutional layer and a max-pooling layer. CNN streams are concatenated and followed by a fully connected layer for classifying text into a positive or a negative emotion class. The used textual features are different types of word embeddings including randomly initialized word embeddings and pre-trained word embeddings. The used pre-trained word embeddings are of different variants such as Word2Vec, Glove and fastText models. The word embeddings in both setups are

trainable and updated through the training phase. After training, the model learns relations between words and generates task-specific word embeddings.

The proposed model has been tested on the Stanford Twitter Sentiment (STS) dataset for classifying emotions. Experiments indicated that the achieved accuracy is 78.5% when using the randomly initialized word embeddings. The accuracy increases 3.6% when using fasttext pretrained word embeddings, 4.5% when using Glove pretrained word embeddings and 6.4% when using Word2Vec pretrained word embeddings. The best tuning for the model is when using Word2Vec pretrained word embeddings which achieves 84.9% accuracy. It is concluded that using CNNs in emotion detection from text is very promising as even when using randomly initialized word embeddings it achieves 78.5% accuracy without any external dataset. Also, not only randomly initialized word embedding can achieve good accuracy in emotion detection from text, it is proven that the power of the pretrained word embeddings helps to achieve a higher competitive accuracy in emotion detection from text.

Keywords: Deep Learning, CNN, Sentiment Analysis, Emotion Detection, Social Media Networks, Word Embeddings.

List of Publications

- 1- Eman Hamdi, Sherine Rady and Mostafa Aref, “A Survey on Mental Illness Detection using Language via Social Media Networks,” Proceeding of The Seventeenth Conference on Language Engineering (ESCOLE), Cairo, Egypt, 2017.
- 2- Eman Hamdi, Sherine Rady and Mostafa Aref, “A Convolutional Neural Network Model for Emotion Detection from Tweets,” Proceeding of The Fourth International Conference on Advanced Intelligent Systems and Informatics (AISI), vol. 845, pp. 337-346, Springer, Cairo, Egypt, 2018.
- 3- Eman Hamdi, Sherine Rady and Mostafa Aref, “A Deep Learning Architecture with Word Embeddings to Classify Sentiment in Twitter,” Proceeding of The Sixth International Conference on Advanced Intelligent Systems and Informatics (AISI), vol. 1261, pp. 115-125, Springer, Cairo, Egypt, 2020.

Table of Contents

Acknowledgment.....	II
Abstract.....	III
List of Publications.....	V
Table of Contents.....	VI
List of Figures	IX
List of Tables	XII
List of Abbreviations.....	XIII
Chapter1: Introduction	2
1.1 Motivation	4
1.2 Problem Definition.....	5
1.3 Research Objectives	5
1.4 Thesis Contributions	6
1.5 Thesis Organization	7
Chapter 2 : Background and Related Work.....	9
2.1 Background	9
2.1.1 Text Classification	9
2.1.2 Sentiment Analysis	11
2.1.3 Machine Learning	12
2.1.3.1 Supervised Learning.....	17
2.1.3.2 Unsupervised Learning.....	18
2.1.3.3 Semi-Supervised Learning	18
2.1.4 Artificial Neural Networks	19
2.1.5 Convolutional Neural Networks	21

2.1.6	Word Embeddings	25
2.1.4.1	Word2vec Word Embeddings	26
2.1.4.2	Glove Word Embeddings	26
2.1.4.3	fastText Word Embeddings	26
2.1.7	Transfer Learning	27
2.2	Related Work	29
2.2.1	Lexicon-Based Approach	30
2.2.2	Machine Learning Approach	30
2.2.2.1	CNN Working with Character-Level Features.....	34
2.2.2.2	CNN Working with Word-Level Features	38
2.2.2.3	CNN Working with Both Character-Level and Word-Level Features	43
Chapter 3 : A CNN Based Emotion Detection Model.....		51
3.1	Model Architecture	51
3.1.1	Text Pre-processing	53
3.1.1.1	Filtering Sentences	54
3.1.1.2	Tokenizing Sentences	54
3.1.1.3	Indexing Sentences	54
3.2	CNN Streams	55
3.2.1	Embedding Layer.....	55
3.2.1.1	Randomly Initialized Word Embeddings	57
3.2.1.2	Pretrained Word Embeddings.....	58
3.2.2	Convolutional and Max-pooling Layers	58
3.3	Fully connected layer	63
3.4	An Illustrative Example of a Sentence-level Execution	66

Chapter 4 : Experimental Results and Model Evaluation.....	77
4.1 Dataset.....	77
4.2 Experiments.....	81
4.2.1 Evaluation metrics	81
4.2.2 Model Settings	83
4.2.1.1 Randomly Initialized Word Embeddings	84
4.2.1.2 Pre-trained Glove Word Embeddings	85
4.2.1.3 Pre-trained Word2vec Word Embeddings.....	85
4.2.1.4 Pre-trained fast-Text Word Embeddings.....	85
4.2.3 Results and Discussion	86
4.2.3.1 Training and Validation Results	86
4.2.3.2 Testing Results	96
Chapter 5 : Conclusion and Future Work.....	102
5.1 Conclusion.....	102
5.2 Future work	103
Appendix A: Twitter Sentiment Analysis Datasets.....	104
References.....	107

List of Figures

Figure 2-1 Flow Chart of Traditional Technique for Programming	13
Figure 2-2 Flow Chart of Machine Learning Technique for Programming	14
Figure 2-3 Training Phase in Machine Learning in Text Classification.....	16
Figure 2-4 Testing Phase in Machine Learning in Text Classification	17
Figure 2-5 The Artificial Neuron	19
Figure 2-6 The Multilayer Neural Network Structure	21
Figure 2-7 The Basic Architecture of a CNN	23
Figure 2-8 How a Kernel of Convolutional Layer Is Applied on Input to Generate a Feature Map	24
Figure 2-9 Machine Learning without Transfer Learning	28
Figure 2-10 Transfer Learning	29
Figure 2-11 A Convolutional Neural Network Model.....	35
Figure 2-12 A Deep Learning Model applied on An Example Sentence	37
Figure 2-13 A Model Architecture with Two Channels for an Example Sentence	39
Figure 2-14 A Deep Learning Architecture For Sentiment Classification	42
Figure 2-15 An Overview of a Deep Learning System	46
Figure 3-1 The Main Phases of The Proposed Emotion Detection Model	52
Figure 3-2 The Block Diagram of The Model	53
Figure 3-3 Text Pre-processing.....	53
Figure 3-4 The Embedding Layer.....	57
Figure 3-5 A Single CNN Stream Convolutional and Max-pooling Layers	59
Figure 3-6 ReLU Activation Function	60
Figure 3-7 Convolutional and Max-pooling Layers of The First CNN Stream....	63
Figure 3-8 The Sigmoid Fully Connected Layer	64
Figure 3-9 Sigmoid Activation Function	64
Figure 3-10 Dropout Neural Net Model.	65
Figure 3-11 First CNN Stream Filter of Size [3*300] Strides by 1 on The Sentence Matrix	72

Figure 3-12 Second CNN Stream Filter of size [5*300] Strides by 1 on The Sentence Matrix	73
Figure 3-13 Third CNN Stream Filter of size [7*300] Strides by 1 on The Sentence Matrix	74
Figure 3-14 Illustration on How the Sentence is Processed through the CNN Streams.....	75
Figure 4-1 Training Vs Validation Accuracy Graph Using Random Word Embeddings	87
Figure 4-2 Training Vs Validation Loss Graph Using Random Word Embeddings	87
Figure 4-3 Training Vs Validation Accuracy Graph Using Glove Wikipedia Word Embeddings	89
Figure 4-4 Training Vs Validation Loss Graph Using Glove Wikipedia Word Embeddings	89
Figure 4-5 Training Vs Validation Accuracy Graph Using Glove Twitter Word Embeddings	90
Figure 4-6 Training Vs Validation Loss Graph Using Glove Twitter Word Embeddings	90
Figure 4-7 Training Vs Validation Loss Graph Using fastText Wikipedia Word Embeddings	91
Figure 4-8 Training Vs Validation Loss Graph Using fastText Wikipedia Word Embeddings	91
Figure 4-9 Training Vs Validation Loss Graph Using fastText Crawl Wikipedia Word Embeddings	92
Figure 4-10 Training Vs Validation Loss Graph Using fastText Crawl Wikipedia Word Embeddings	92
Figure 4-11 Training Vs Validation Loss Graph Using Word2Vec Word Embeddings	93
Figure 4-12 Training Vs Validation Loss Graph Using Word2Vec Word Embeddings	93