DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Deep Learning for Sentiment Analysis

Ankit Bahuguna

DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Deep Learning for Sentiment Analysis

Deep Learning Methoden für Sentiment Analysis Probleme

Author:	Ankit Bahuguna
Supervisor:	Prof. Georg Groh
Advisor:	Prof. Georg Groh
Submission Date:	September 15, 2015

I confirm that this master's thesis in informatics is my own work and I have documented all sources and material used.

Munich, Germany, Ankit Bahuguna
September 15, 2015

Acknowledgments

I would like to wholeheartedly thank my supervisor and adviser **Mr. Georg Groh**, who advised my Master's Thesis at TUM and encouraged me the whole time. He played a pivotal role in me continuing the study in the area of my interest and help me chase the state of the art systems with utmost frivolity. I would also thank, **Prof. Hinrich Schütze**, with whom, I started this study during an interdisciplinary project at Ludwig Maximillian's University (München) in April 2014 . His supervision and guidance helped me to get started with the new area of Deep Learning and its application to problems in Natural Language Processing. Additionally, I would thank my previous guide, **Prof. Pushpak Bhattacharrya**,(Indian Institute of Technology, Bombay) who introduced the field of Natural Language Processing and helped kick-start my interest in the area and encouraged me continuously to aim higher and helped me open my mind to look at problems from various perspectives.

In the end, I would also like to thank my family, who gave me this opportunity to study at Technical University of Munich, Germany. Without them, nothing of this would be possible.

Abstract

Word vector representations or word embeddings, learned using deep neural network models prove extremely useful at capturing semantically similar words and phrases. Recent work in the area, utilize semantic word vector spaces and sentiment compositionality to classify sentiment better. One of the biggest constraints with these methods, is need of manually labeled sentiment for phrases in parse trees of sentences making its use challenging for languages with constrained resources.

Moreover, traditional methods do not efficiently leverage the use of semantic word vector spaces and word context in overall sentiment classification. In this report, we first present a broad overview of the problem, discuss traditional methods and then introduce Deep Learning based methods for Natural Language Processing which help in understanding the problem of Sentiment Analysis, better than traditional methods. Also presented is the analysis of concatenating various word vector representations which do not necessarily require phrase level labeling, comparisons of their results in various experimental configurations and showing improvement over traditional unsupervised baseline methods and state of the art results. In the same experiment, we also explore in-domain, out-of-domain and mixed-domain data towards training these word vectors and show how each one affects sentiment classification.

We also present promising future work, which may lead to better deep neural network based word representations for various NLP tasks.

Contents

1 Introduction

Romance should never begin with sentiment. It should begin with science and end with a settlement. — Oscar Wilde, An Ideal Husband

1.1 Information Gathering, Opinions and Sentiment

"*What other people think*", has been a very important element in the overall decision making process. This is more relevant in the modern era with the explosion of web and internet based services where the people have a new platform where they can voice their opinions and discuss about day to day things. With the advent of this platform, the subjective information too has grown severely rapidly over the years.
It is interesting to note that the social opinion of users impact a large corpus of audience. A simple case example is, that if a product is negatively perceived in the social circles of the internet, it will not be taken positively by a new user who is interested in purchasing the product. This also is hence a tool for organizations to understand how the product is performing in the public.
In the modern day, organizations have significant investments to process this vast amount of data and make sense of the important information contained in it, so that they can leverage their products and services and improve their overall user experience or increase chances of making more profit, by identifying key areas which are causing an indirect impact on the overall balance statement.

1.2 Sentiment Analysis: An Introduction

1.3 Scope of the Problem

1.4 Applications

2 Sentiment Analysis

2.1 Supervised Machine Learning Techniques

2.2 Unsupervised Machine Learning Techniques

2.3 Domain Adaptation

3 Deep Learning

3.1 Foundations: Neural Networks

3.2 From Neural Networks to Deep Neural Networks

3.3 Applications

4 Deep Learning and NLP: Word Representations and Language Modelling

4.1 Language Modelling

4.2 Word2Vec

4.3 Glove

4.4 Polyglot

5 Deep Learning: Modern Apporaches

5.1 RNTN : Recursive Neural Tensor Networks

5.2 MSDA : Marginalized Stacked Denoising Auto Encoders

5.3 CNN : Convoluted Neural Networks

5.4 LSTM : Long and Short Term Memory Networks

5.5 Comparative Evaluation

6 Concatenated Word Representations for Sentiment Classification

6.1 Problem

6.2 Experiments

6.3 Observations

6.4 Result

7 Moving from Text to Speech: Emotion Recognition

7.1 Conventional Methods

7.1.1 GMM: Gaussian Mixture Models

7.1.2 I-Vectors: Total Variability Matrix

7.2 Deep Neural Networks for Emotion Recognition

8 Conclusion

8.1 Section

Citation test [Lam94].

8.1.1 Subsection

See Figure 8.1.

Figure 8.1: An example for a figure.

8.2 Section

See Table 8.1, Figure 8.2, Figure 8.3, Figure 8.4.

Table 8.1: An example for a simple table.

A	B	C	D
1	2	1	2
2	3	2	3

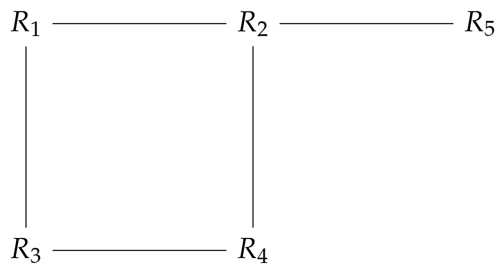

Figure 8.2: An example for a simple drawing.

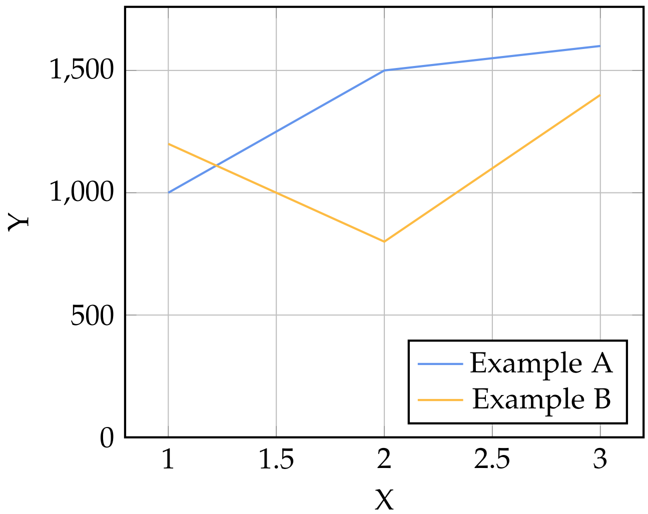

Figure 8.3: An example for a simple plot.

```
SELECT * FROM tbl WHERE tbl.str = "str"
```

Figure 8.4: An example for a source code listing.

Glossary

computer is a machine that. . . .

Acronyms

TUM Technische Universität München.

List of Figures

List of Tables

Bibliography

[Lam94] L. Lamport. *LaTeX : A Documentation Preparation System User's Guide and Reference Manual*. Addison-Wesley Professional, 1994.