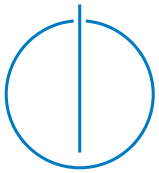
DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Master's Thesis in Informatics

Deep Learning for Sentiment Analysis

Ankit Bahuguna



DEPARTMENT OF INFORMATICS

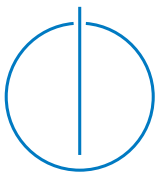
TECHNISCHE UNIVERSITÄT MÜNCHEN

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, September 15, 2015 Ankit Bahuguna

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

Acknowledgments **iii**

Abstract **iv**

1 Introduction **1**
1.1 Information Gathering, Opinions and Sentiment 1
1.2 Sentiment Analysis: An Introduction 1
1.3 Scope of the Problem 1
1.4 Applications 1

2 Sentiment Analysis **2**
2.1 Section 2
2.1.1 Subsection 2
2.2 Section 2

3 Deep Learning **4**
3.1 Section 4
3.1.1 Subsection 4
3.2 Section 4

4 Deep Learning and NLP: Word Representations and Language Modelling **6**
4.1 Section 6
4.1.1 Subsection 6
4.2 Section 6

5 Deep Learning: Modern Apporaches **8**
5.1 Section 8
5.1.1 Subsection 8
5.2 Section 8

6 Concatenated Word Representations for Sentiment Classification **10**
6.1 Section 10
6.1.1 Subsection 10
6.2 Section 10

7 Moving from Text to Speech: Emotion Recognition **12**
7.1 Section 12
7.1.1 Subsection 12
7.2 Section 12

8 Conclusion **14**
8.1 Section 14
8.1.1 Subsection 14
8.2 Section 14

Glossary **16**

Acronyms **17**

List of Figures **18**

List of Tables **20**

Bibliography **21**

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 information has grown severely rapidly over the years. Modern day companies 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 Section

Citation test [Lam94].

2.1.1 Subsection

See Figure 8.1.

Figure 2.1: An example for a figure.

2.2 Section

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

Table 2.1: An example for a simple table.

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

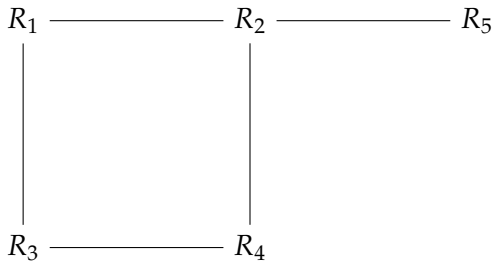


Figure 2.2: An example for a simple drawing.

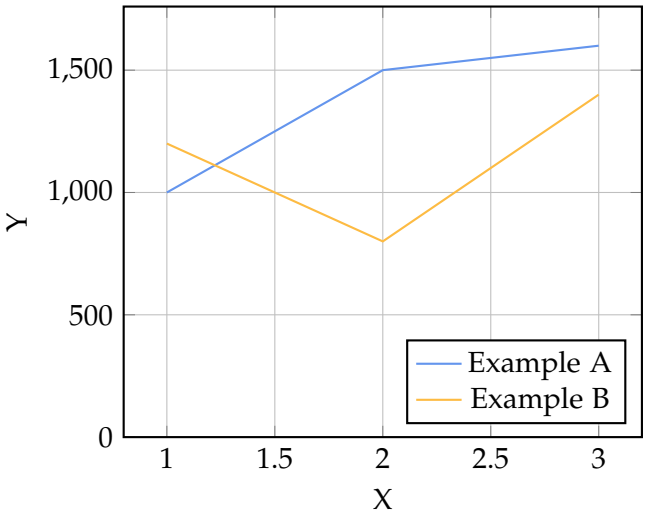


Figure 2.3: An example for a simple plot.

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

Figure 2.4: An example for a source code listing.

3 Deep Learning

3.1 Section

Citation test [Lam94].

3.1.1 Subsection

See Figure 8.1.

Figure 3.1: An example for a figure.

3.2 Section

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

Table 3.1: An example for a simple table.

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

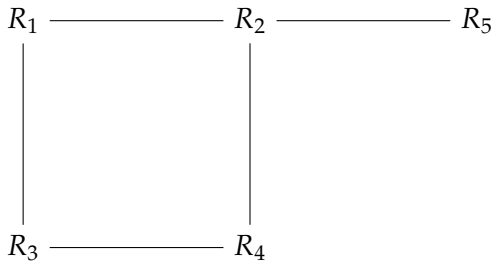


Figure 3.2: An example for a simple drawing.

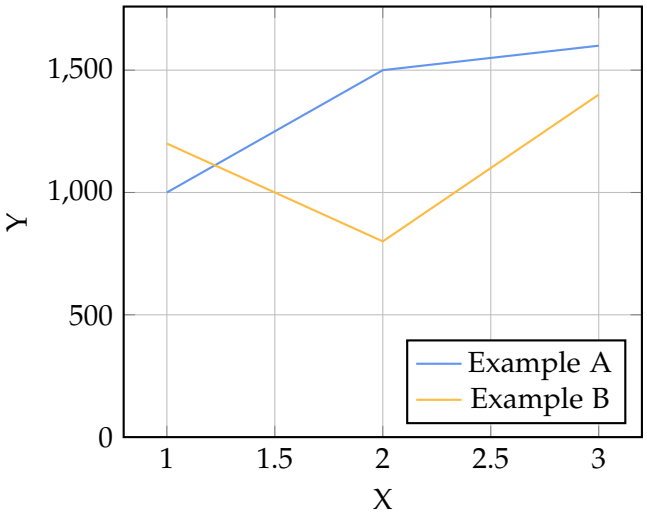


Figure 3.3: An example for a simple plot.

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

Figure 3.4: An example for a source code listing.

4 Deep Learning and NLP: Word Representations and Language Modelling

4.1 Section

Citation test [Lam94].

4.1.1 Subsection

See Figure 8.1.

Figure 4.1: An example for a figure.

4.2 Section

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

Table 4.1: An example for a simple table.

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

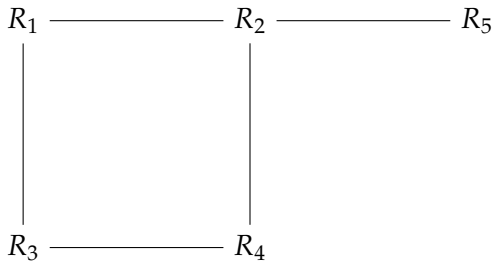


Figure 4.2: An example for a simple drawing.

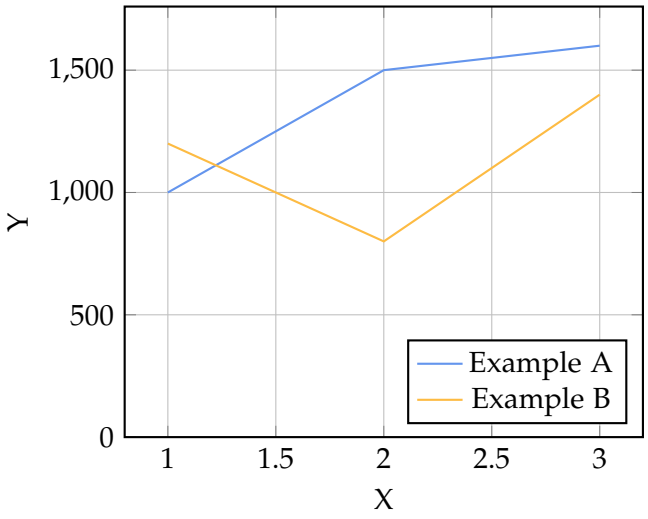


Figure 4.3: An example for a simple plot.

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

Figure 4.4: An example for a source code listing.

5 Deep Learning: Modern Apporaches

5.1 Section

Citation test [Lam94].

5.1.1 Subsection

See Figure 8.1.

Figure 5.1: An example for a figure.

5.2 Section

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

Table 5.1: An example for a simple table.

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

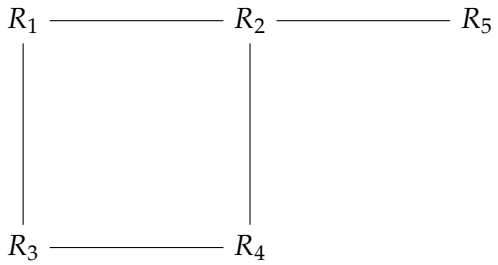


Figure 5.2: An example for a simple drawing.

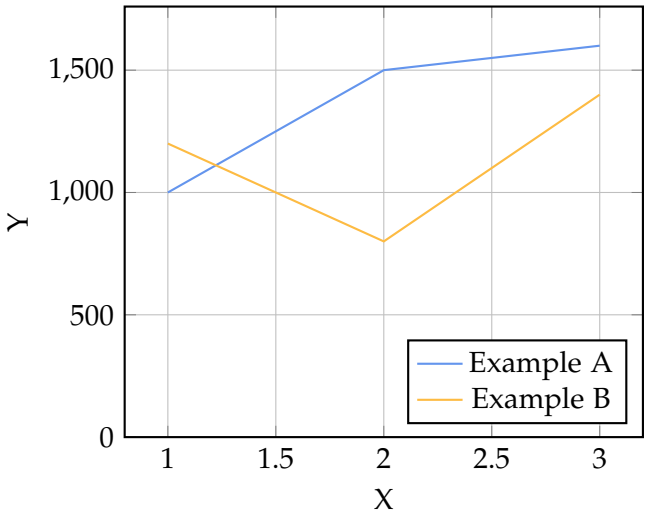


Figure 5.3: An example for a simple plot.

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

Figure 5.4: An example for a source code listing.

6 Concatenated Word Representations for Sentiment Classification

6.1 Section

Citation test [Lam94].

6.1.1 Subsection

See Figure 8.1.

Figure 6.1: An example for a figure.

6.2 Section

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

Table 6.1: An example for a simple table.

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

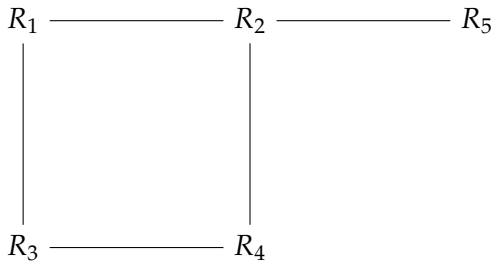


Figure 6.2: An example for a simple drawing.

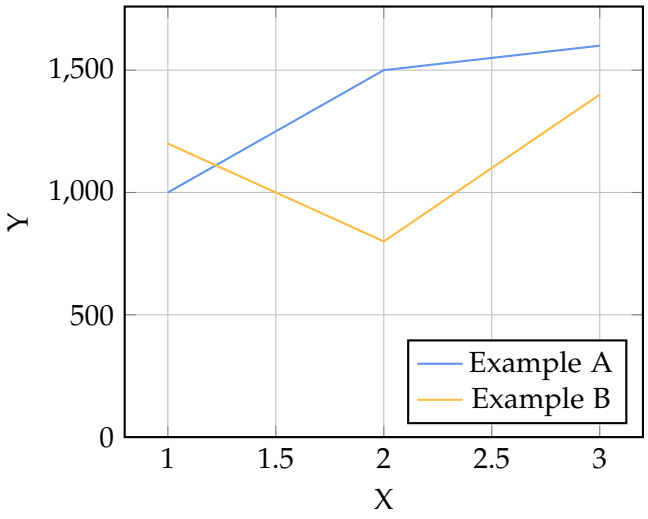


Figure 6.3: An example for a simple plot.

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

Figure 6.4: An example for a source code listing.

7 Moving from Text to Speech: Emotion Recognition

7.1 Section

Citation test [Lam94].

7.1.1 Subsection

See Figure 8.1.

Figure 7.1: An example for a figure.

7.2 Section

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

Table 7.1: An example for a simple table.

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

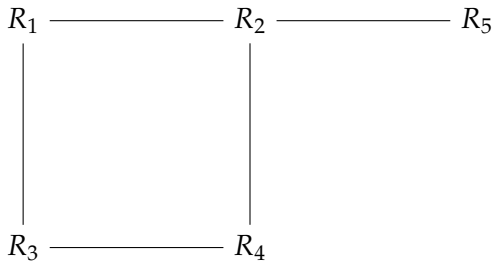


Figure 7.2: An example for a simple drawing.

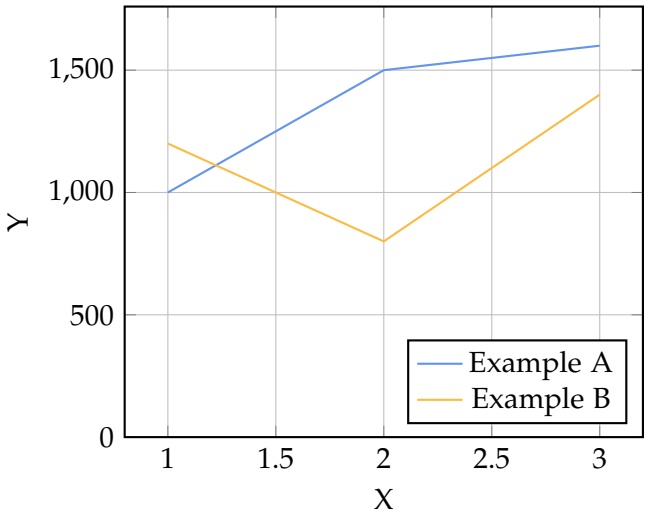


Figure 7.3: An example for a simple plot.

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

Figure 7.4: An example for a source code listing.

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

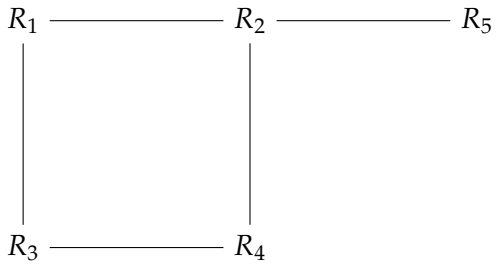


Figure 8.2: An example for a simple drawing.

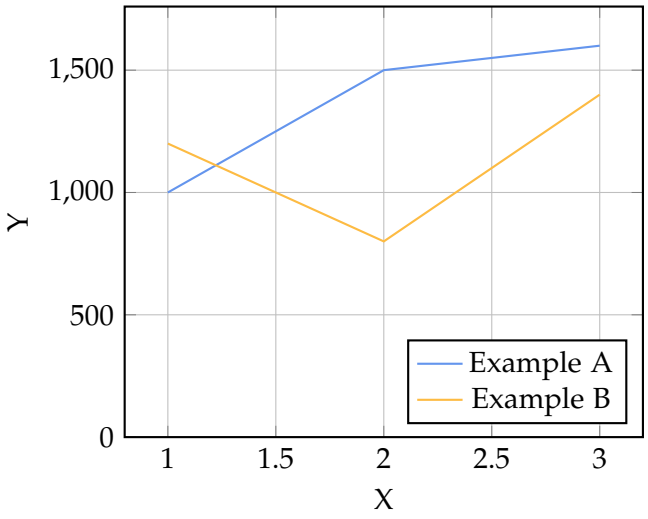


Figure 8.3: An example for a simple plot.

```
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

2.1 Example figure 2
2.2 Example drawing 3
2.3 Example plot 3
2.4 Example listing 3

3.1 Example figure 4
3.2 Example drawing 5
3.3 Example plot 5
3.4 Example listing 5

4.1 Example figure 6
4.2 Example drawing 7
4.3 Example plot 7
4.4 Example listing 7

5.1 Example figure 8
5.2 Example drawing 9
5.3 Example plot 9
5.4 Example listing 9

6.1 Example figure 10
6.2 Example drawing 11
6.3 Example plot 11
6.4 Example listing 11

7.1 Example figure 12
7.2 Example drawing 13
7.3 Example plot 13
7.4 Example listing 13

8.1 Example figure 14
8.2 Example drawing 15
8.3 Example plot 15
8.4 Example listing 15

List of Tables

2.1 Example table 2
3.1 Example table 4
4.1 Example table 6
5.1 Example table 8
6.1 Example table 10
7.1 Example table 12
8.1 Example table 14

Bibliography

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