

Natural Language Processing Emoticon Prediction

By: Gornishka, Loor, Qodari, Xumara

November 22, 2014

Contents

| | |
|------------------------------------|----------|
| Introduction | 2 |
| Assignment | 2 |
| Twitter emoticon prediction | 2 |
| Preprocessing | 2 |
| Model | 2 |
| Features | 2 |
| Results | 2 |
| Ubunto emoticon prediction | 3 |
| Preprocessing | 3 |
| Model | 3 |
| Features | 3 |
| Results | 3 |
| Conclusion | 4 |
| Sources | 4 |

Introduction

This report contains the implementation of an emoticon predictor, based on text. This is a project within Natural Language Processing (NLP), meaning many resources within the NLP community were consulted. This report discusses two different ways of prediction emotions. In the first case, emoticons are predicted based on a short message. This lead to the use of an average multiclass perceptron. In the second case, chat data is used to predict emoticons, based on the message as well as the previous post. This lead to the use of the Viterbi algorithm.

Assignment

As stated in the introduction, this report discusses emoticon prediction based on a short message and based on the previous message. In order to find small messages, Twitter was used. For the second part of the assignment, Ubuntu chat data was used.

Twitter emoticon prediction

In order to collect data from Twitter, a crawler was used. The Twitter crawler searched for training and test data, based on emoticons. These messages contain much information.

Preprocessing

Before any of the data can be used, the messages need to be preprocessed.

Model

Features

Results

Ubuntu emoticon prediction

Preprocessing

Before any of the data can be used, the messages need to be preprocessed.

Model

Features

Results

Conclusion

Files attached

- Code

Sources

- Some paper