State of the art in *Sentiment Analysis*

C. B. Hotz-Behofsits
0929002
christian.hotz-behofsits@tuwien.ac.at

Thomas Schmidtleitner
1025525
e1025525@student.tuwien.ac.at

Dominik Pichler
1026045
dominik.pichler@aon.at

Matthias Reisinger
1025631
matthias.reisinger@web0.at

Florian Taus
0627918
florian.taus@hotmail.com

ABSTRACT

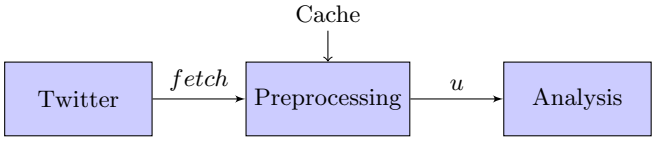
The extensive growth of user-generated content in social networks and the common usage of emoticons and hashtags has introduced new possibilities to classify these information. In this paper, we provide an overview of the state-of-the-art regarding sentiment analysis of Twitter messages. The usefulness of existing lexical ressources and special expressions like smileys or hashtags is evaluated and an general introduction to sentiment analysis is provided as part of the introduction.

1. INTRODUCTION

Within the last years Twitter and similar social media platforms grew considerably in terms of user numbers and mainstream fame. Twitter has about 284 Mio. active users, generating approximately 500 Mio. posts[1] on a single day[2]. These impressive numbers show the amout of information generated by the crowd. Furthermore Twitter is able to reach a vast number of potential customers especially for consumer businesses. As a result the named services gain more and more acceptance as opinion platforms and powerful tools for both, opinion leader as well as pollister. This trend influenced various marketing and sales strategies and created a new market for companies specialiced on sentiment analysis of tweets, like tweetfeel[3], Social Mention[4] and Twitratr[5].

The aim of sentiment analysis is to determine and try to measure positive and negative feelings, emotions and opinions written in a text. English as language allows to express the same intent in different ways. The main challenge therefore consists in abstracting the intention of the writer from the grammatic and language specific rules.

In general sentiment analysis can be splitted into 2 steps: a preprocessing and a constitutive sentiment analysis phase. In Twitter sentiment analysis there is another step right bevor preprocessing: fetching the data from the application programming interface (API). This task is not trivial, because the official API is limited regarding the datasets fetched within certain intervals. This leads to the requirement of a caching mechanism to avoid fetching the same tweets multiple times and allow the usage of more datasets.



While sentiment analysis of conventinal ressources like news papers or articles is quite well investigated the special research area of sentiment analysis in terms of social network posts is relatively new. Excpecially the possibility of priorizing keywords[6] and the common usage of emoticons lead to advanced possibilities to categorize feelings of posts.

2. PREPROCESSING

The Data-Preprocessing-Process is an essential part of sentiment-Analysis. Its goal is to prepare data for the sentiment analysis and remove unnecessary parts.[1] Unnecessary Parts are:[1]

1. Remove URLs, Special Characters
2. Filter Unnecesarry Words
3. Remove Retweets

Besondere an twitter: Hashtags, Mentions, Smileys Tokenizen

3. DATA AND METHODS

4. CONCLUSIONS

APPENDIX

[1] posts are also known as tweets
[2] `https://about.twitter.com/company`, Effective 12.11.2014
[3] `www.tweetfeel.com`
[4] `www.socialmention.com`
[5] `www.twitratr.com`
[6] known as hashtags

A. HEADINGS IN APPENDICES

References

[1] Govardhan Hemalatha Saradhi-Varma. "Preprocessing the Informal Text for efficient Sentiment Analysis". In: *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)* 1 (2012), pp. 58–61.