

Emoji Prediction for Text Messages

Carsten Draschner - 2719095

Maren Pielka - 2468882

Jonas Weinz - 2571421

Motivation

- In addition to the widely used auto correction for texts and prediction of next words (i.e SwiftKey) propose also Emojis etc.
- Usage of Emojis are related to topics occurring in text message as well as to the writers sentiment
- Also Emojis are common to indicate writers sentiments
 - labeling for text messages
 - with some noise and possible misunderstandings
 - irony, sarcasm
 - insider

Emoji Prediction - Topic vs Sentiment

Topic 🚗 🚑 🚚 🎮 🎯 ⑧

- close to occurring words in text
- less interpretation
- no need for large intelligence

Sentiment 😄 😊 😁

- not easy to interpret sentiment directly from text
- need of intelligence
- Machine learning and mining of text messages
- prediction can be tuned by reinforcement learning
 - i.e. by recognizing selected emoji after providing a set of best fitting emojis

Preprocessing Data

Preprocessing

For the emoji specifications:

- Lowercasing
- Stopword removal

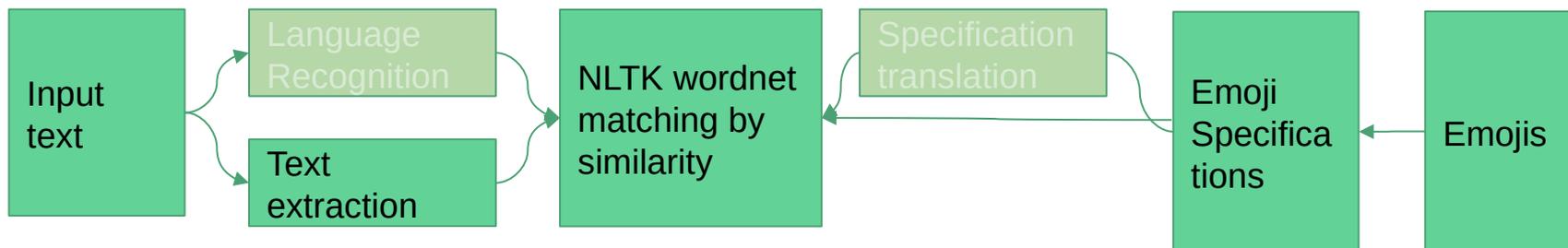
For the text corpus:

- removed retweets
- removed URLs
- replaced emojis, usernames and hashtags with keywords and added them as own labels

	EMOJI	text	lang	LINKED_USER
0	[🔥, 🍌]	fashionbombdally's photo <EMOJI> <EMOJI>🍌	en	[]
1	[👤]	It's scary how on point my horoscope be <EMOJI>...	en	[]
2	[😞]	Wooooaaahhh <EMOJI>	en	[]
3	[📷]	<EMOJI> vivalc!l: Portraits by Zhao Guojing an...	en	[]
4	[@, @]	<USER> I wanna know too<EMOJI> <EMOJI>	en	[@hiphoppHles]
5	[👤, ❤️]	<USER> veda was yoda too <EMOJI> <EMOJI>	en	[@WizMommma]
6	[😞]	I'm less stressed about turning 30 now <EMOJI>...	en	[]
7	[🔋]	Full charged. <EMOJI>	en	[]
8	[😞]	<USER> That's part of the problem, (they) <USE...	en	[@SeaDimon, @Isarsour]
9	[😞, 😞, 😞]	<USER> got me emotional there<EMOJI> <EMOJI><EM...	en	[@lan_khetye]

Naive Approach

Naive Approach



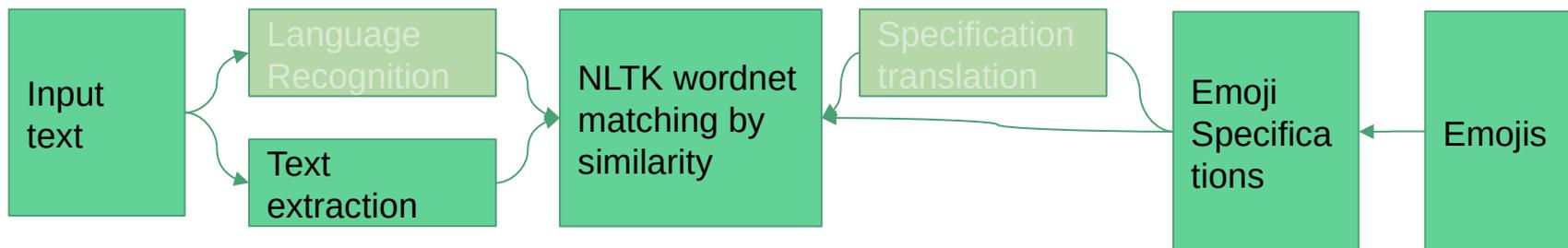
Wo steht
dein Auto?



I like playing
soccer



Naive Approach - Different language: German



Wo steht
dein Auto?

I like playing
soccer

WordNet not available in German
→ Maybe GermaNLTK could be
usable



Findings from the Naive approach

Positive

- The Naive approach is efficient in matching topic-related emojis to text which contains the respective specification
- I.e. “I like to drive my car”, “I like soccer” etc)



→ can roughly be used for topic dependent emoji prediction

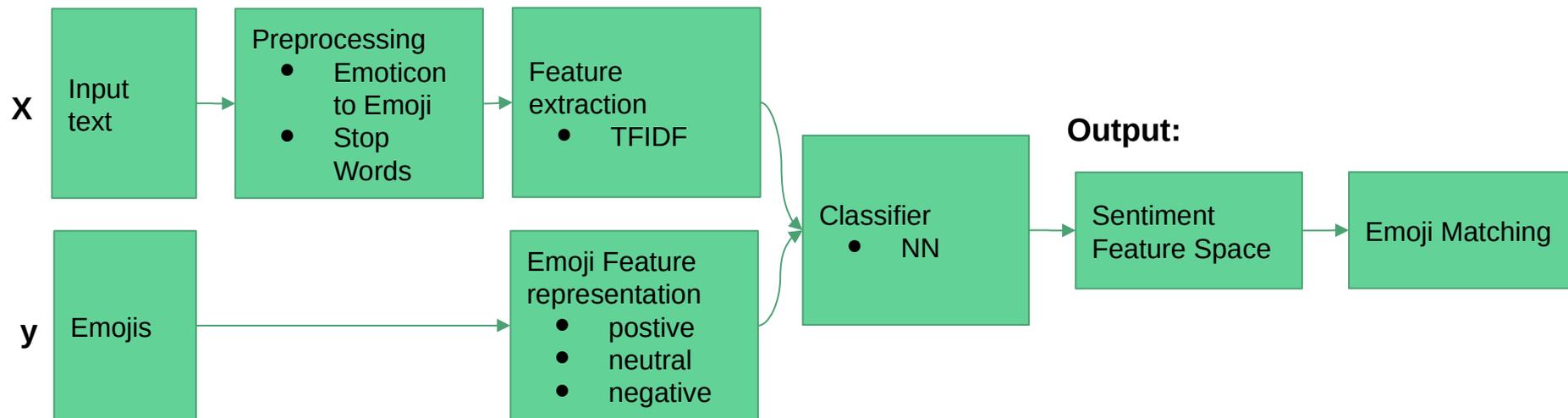
Negative

- It fails in recognizing emotion-related emojis (smileys) 😊
- sentiment is rarely related to used words in text

→ A more sophisticated approach is needed!

Advanced Approach

Advanced Approach: Outline



Emoji distance

The distance or the correlation of sentiment of a pair of Emoticons

Sentiment Space: Transform Emoji into sentiment space to interpret each emoji as representation of its cluster or group of similar emojis

- 1-dim: Negative/Positive
- 2-dim: Negative/Positive, Neutrality
- 3-dim: Negative, Positive, Neutrality
- 6-dim: Sadness Anger Joy Fear Surprise Disgust
- ...

Emoji distance

The distance or the correlation of sentiment of a pair of Emoticons

Sentiment Space: Transform Emoji into sentiment space to interpret each emoji as representation of its cluster or group of similar emojis

- 1-dim: Negative/Positive
- 2-dim: Negative/Positive, Neutrality
- **3-dim: Negative, Positive, Neutrality**
- 6-dim: Sadness Anger Joy Fear Surprise Disgust
- ...

Emoji representation

- 3-dimensional feature space
- The vector entries correspond to the relative occurrence of the emoji in different contexts.
- The possible contexts are “Positive”, “Negative” and “Neutral”.

We obtained this approach and the values from the work of *Novak et. al., 2015*.

Emoji/ Sentiment			
Positive	0.269113 1	0.754660	0.162962
Negative	0.480122 3	0.052905	0.555555
Neutral	0.250764	0.192434 2	0.281481

Problem specification and error function

- Using the 3-dimensional vector representation
 - Distances Example
 - Error function:
 - Euclidean distance of the prediction to the correct outcome in the feature space.
- regression problem

Original Emoji	Emojis with closest distance to the original emoji (rank 1 to 7)
	      
	      

Current Project Status

Training

- very first run so far:
 - using keras
 - NN with three hidden layers (10000, 5000, 2500 Neurons)
 - 3D output in sentiment space, mean squared error as loss function
 - RELU as activation function
- trained on smaller twitter subset and only with the 20 most used emojis:

- 🤔😬😍😞😁😂🙌🙏😄😇😁😍😘😙😚😛😜😝😞😟😠😡😢😣😤😥😦😧😨😩😪😫😬😭😮😯😰😱😲😳😴😵😶😷😸😹😺😻😼😽😾😿😺😻

```
clf = train(max_size=100000, layers=[(10000, 'relu'), (5000, 'relu'), (2500, 'relu'), (y1[0].shape[0], None)], n_iter=3)
```

Train on 18861 samples, validate on 4716 samples

Epoch 1/3

18861/18861 [=====] - 1288s 68ms/step - loss: 0.0185 - val_loss: 0.0164

Epoch 2/3

18861/18861 [=====] - 1282s 68ms/step - loss: 0.0103 - val_loss: 0.0164

Epoch 3/3

18861/18861 [=====] - 1333s 71ms/step - loss: 0.0063 - val_loss: 0.0166

(Very) First Results

- interactive tests:

sentence	prediction
"I am so sad!"	
"I am so happy!"	
"I love you so much!"	

- on test dataset:

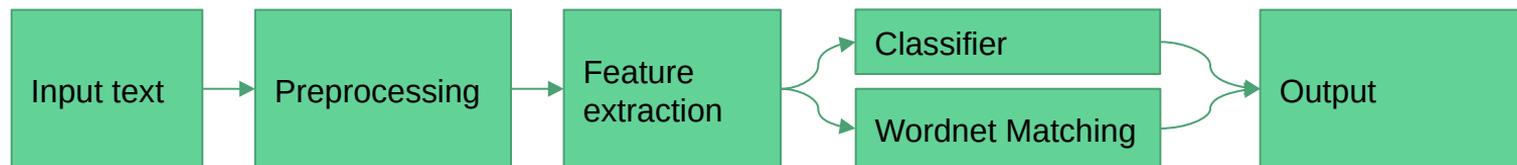
text	teacher	predict
I feel like I care so much more in every situa...		
I did not meat to add that '2' there...havent ...		
never...		
Lmao on me!!! Wtf was he supposed to say		
This dude always help me get through my schoo...		

- unfortunately still very random 😞

Prospects and further Ideas

Prospects and further ideas

- Combine naive and advanced approach into a model that can deal with both topic-related and emotion-related emojis



Where is
your car?



I am so
happy today



Prospects and further ideas (1)

- Design Feature Vector
 - Add more preprocessing options
 - Stemming, Lemmatization, emoticon emoji replacement
- Compare different classifiers
 - Hyperparameter Tuning
- Compare different learning Approaches:
 - Multiclass or Multidimensional Regression
 - Choose of sentiment dimensions
 - Modify the error function
- Resulting Demo Application
 - Searching and using for existing Chatbot/Chat/Dialog User Interface

Prospects and further ideas (2)

- Multi Language Support
 - currently english
 - later german...
- Handling Emoji Modifiers
 - hair/skin colour
 - sex
- Consider the contexts of emoji usage
 - individual
 - dialog (“insider”)
 - overall

References

References

Hallsmar, F. and Palm, J.: *Multi-class Sentiment Classification on Twitter using an Emoji Training Heuristic*, KTH Royal Institute of Technology, Stockholm 2016.

Novak, P.K.; Smailovic, J.; Sluban, B. and Mozetic, I.: *Sentiment of Emojis* (Journal Paper), PLoS ONE 2015.

Zhao, L. and Zeng, C.: *Using Neural Networks to Predict Emoji Usage from Twitter Data*, Stanford University, 2017.

Thank you for your attention!
