Emoji2Text: Translating Emoji Sequences into Natural Language Using Deep Learning

Nico Wedel (7366091), Laurenz Flender (7430935)
{ngrafvon, lflender}@smail.uni-koeln.de
Advanced Data Analytics for Business, University of Cologne

1. Problem Statement

Emojis are widely used in online communication. They often replace words or even full expressions. However, their meaning can change depending on context, culture, or usage. This makes it difficult to automatically interpret emoji sequences. In this project, we aim to build a model that translates emoji sequences into short natural-language phrases. For example, an emoji sequence like in Figure 1 may correspond to "I am dying of laughter." Our goal is not to create a perfect translation system. Instead, we want to explore whether modern deep learning methods can capture semantic patterns in emoji usage and produce meaningful textual interpretations.

😂☠️

Figure 1: Example of emoji sequence

2. Existing Methods

Neural sequence-to-sequence models are widely used for translation tasks. Sutskever et al. (2014) introduced encoder–decoder architectures based on recurrent networks, showing that neural models can learn to map one sequence to another. Bahdanau et al. (2015) extended this idea with attention mechanisms, enabling the model to focus on relevant parts of the input during decoding. Transformers (Vaswani et al., 2017) further improved translation performance by relying entirely on attention instead of recurrence, allowing efficient parallelization. Representation learning methods such as SimCLR (Chen et al., 2020) demonstrate how models can learn semantic embeddings without labeled data. Graph-based approaches like Graph Convolutional Networks (Kipf and Welling, 2017) capture structural relationships such as token co-occurrences.

These methods are well established for natural-language text, but emojis function differently: they are semantic units without grammatical rules and often appear in informal contexts. Applying translation and representation-learning techniques to emoji sequences is therefore a novel problem that has received limited attention.

3. Data Description

We plan to use the Twitter Emoji Prediction Dataset from Kaggle, which contains about 1.6 million tweets. Each tweet includes at least one emoji. This allows us to build realistic training pairs by extracting the emoji sequence and using the accompanying text as a natural-language description.

Main Dataset characteristics

Property	Description
Source	Twitter dataset on Kaggle
Size	∼1,600,000 samples
Inputs	Emoji sequences extracted from tweets
Outputs	Short text snippets from the tweet content
Supplementary semantic sources	Unicode descriptions; EmojiNet graphs

In addition to the this dataset, we may also use or even integrate information from Unicode emoji descriptions or EmojiNet. These provide semantic labels, definitions, and relations between the emojis. Using these resources can support cleaning, clustering, or enhancing training labels if we need it.

This combination could provide large-scale real-world usage data and also structured semantic information. This may improve our model performance.

4. Methodology

We will experiment with different deep learning methods to find a suitable approach. First, we may build a simple baseline model using a sequence-to-sequence LSTM encoder–decoder. Emojis will be the input tokens, and the decoder will generate text. Next, we may test a Transformer-based model, which is often more effective for translation tasks. Optionally, we may explore learning emoji embeddings using contrastive learning or modeling emoji co-occurrence using a graph-based method such as GCN. This depends on project time and feasibility. We will evaluate the models using qualitative examples and simple metrics such as BLEU score or cosine similarity between predicted and reference text. Our goal is to demonstrate whether the model can learn meaningful mappings from emoji patterns to natural-language interpretations.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations*, 2015.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *International Conference on Machine Learning*, 2020.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *International Conference on Learning Representations*, 2017.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27, 2014.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 2017.