



Open in app







Deep Learning and Time Embeddings To Predict Emojis

Overview

Emoji usage has become a new form of social communication, which is important because it can help to improve communication systems, such as chat applications. This paper investigates the *usage* and *semantics of emojis* over time to analyze *seasonal variation* of emoji usage. In addition, they develop an emoji prediction model based on the time information.

Contribution

Multiple emoji predictions studies have been carried out in the past (see notable work by <u>Felbo et al., 2017</u>), but non have considered *temporal information*. Temporal correlation between emojis and seasonal events are explored and used to disambiguate emoji meanings.

Example

Consider the leaf clover emoji (\(\frac{\psi}{\psi} \)), it is usually associated with good luck wishes all year round except in March, where it is mostly used to express event situations related to *parties* and *drinking* (due to St. Patrick day).

Challenges

- This study demonstrates that temporal information is useful for emoji prediction even for emojis that are not associated with time (and ♥).
- Emojis are innately subjective, which is why it is difficult to analyze their semantic meaning.

Dataset

Twitter is used to collect a 100 million US tweets corpus and organized as follows:

- Seasonal Emoji Dataset data is divided into four subsets by seasons: Spring, Summer, Autumn, and Winter (see figure below)
- *Emoji Prediction Dataset* data is reduced to tweets that only contain one frequent emoji (emoji must belong to the 300 frequent emojis)



Seasonal Emoji Semantic and Usage

Skip-gram word embedding models are trained using the four subsets of the seasonal datasets. These models provide information that basically helps to describe emojis in terms of their semantic similarity to each other. (See paper for more details)

By comparing the top 10 emojis associated to each emoji in the embedding space, it was discovered that emojis related to music,















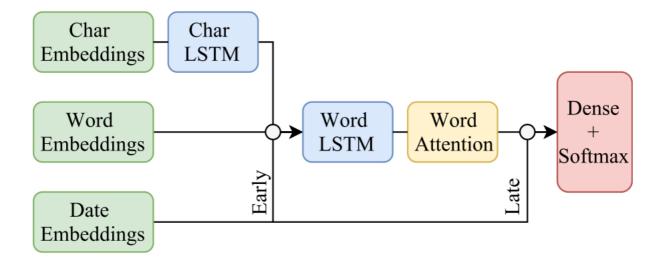


party emojis, and during Autumn it was associated with school-related emojis. Check out the top 10 associated emojis per season for the pine emoji (4) in the figure below — very season-dependent don't you think? Can you guess why? (hint: outdoors vs Christmas). (See paper for tons of interesting findings)



Emoji Prediction

The second dataset, which includes 300 emoji classes and 900,000 tweets total (3,000 tweets per class), is used for emoji prediction. The architecture of the emoji prediction model is as follows: character embeddings, word embeddings, and data embeddings are combined through both an early fusion approach and a late fusion approach. This produces two models (*Early* and *Late*). A third model is trained (*W/O*) which completely ignores the date embeddings. (See paper to find out how these embeddings are constructed)



Results

Precision, Recall, and F1 scores are reported for all models in the table below. We can observe that by combining time information using early fusion, the *Early* model outperforms the other models.

	P	R	F1	a@1	a@5	a@10	CE
W/O	21.97	23.22	21.89	23.13	38.22	45.70	44.29
Early	22.10	23.43	22.06	23.33	38.55	46.29	42.59
Late	21.83	23.00	21.63	22.91	37.85	45.62	43.91

The emojis that had higher gains in F1 score (without date vs. early date) are presented in the table below. You can definitely observe that many emojis are season-specific (e.g., \$, \blacksquare) and thus benefit from the date embeddings. Even emojis that are not associated to time (e.g., \P , \square) benefit from the temporal information.













Open in app

Upgrade

Early										
Emoji										
W/O	0.44	0.30	0.36	0.50	0.10	0.12	0.15	0.18	0.22	0.11
Early	0.49	0.34	0.40	0.54	0.14	0.16	0.19	0.22	0.26	0.15

Conclusion & Future Work

- A multimodal architecture was proposed to conduct emoji prediction based on deep neural networks.
- More analysis on the emoji semantics and usage over specific time of the day or week may be able to help improve the date embeddings and the overall predictive models.
- This work has a lot of room for improvement and it could be a very interesting topic to combine with emotion recognition, event detection, and computational health studies.

References

• Ref: https://arxiv.org/abs/1805.00731 — "Exploring Emoji Usage and Prediction Through a Temporal Variation Lens"