

Kombinasi fitur semantik, fitur linguistik, dan fitur leksikal pada klasifikasi emoji di dataset twitter = Combining semantic feature, linguistic feature, and lexicon feature for emoji classification in twitter dataset / Rinda Wahyuni

Rinda Wahyuni, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20476304&lokasi=lokal>

Abstrak

Emoji merupakan karakter gambar yang digunakan dalam komunikasi informal khususnya pada media sosial. Emoji digunakan oleh penulis pesan untuk mengekspresikan emosi sebuah pesan. Besarnya penggunaan emoji, membuat emoji sangat berpengaruh terhadap komunikasi di media sosial. Emoji digunakan sebagai salah satu fitur untuk analisis sentimen dan ekstraksi emosi dalam penelitian Natural Language Processing dan Information Retrieval, namun masih sedikit penelitian yang fokus menentukan emoji dari sebuah teks. Banyaknya emoji dan kemiripan makna antar emoji membuat klasifikasi emoji menjadi lebih kompleks jika dibandingkan dengan analisis sentimen atau klasifikasi teks pada umumnya. Penelitian ini menggunakan fitur leksikal, fitur semantik, dan fitur linguistik pada permasalahan klasifikasi emoji untuk mengetahui pengaruh setiap fitur pada performa klasifikasi emoji dan mengetahui kombinasi fitur terbaik dalam klasifikasi emoji. Hasil eksperimen menunjukkan fitur semantik memiliki performa terbaik saat digunakan secara individu. Sedangkan fitur leksikal memiliki pengaruh besar terhadap kenaikan performa klasifikasi emoji saat dikombinasikan dengan fitur baseline. Hasil uji statistik paired t-test menunjukkan kombinasi tiga fitur dan kombinasi empat fitur menaikkan akurasi baseline secara signifikan. Kombinasi terbaik didapatkan ketika mengkombinasikan baseline, fitur linguistik, fitur leksikal, dan fitur semantik dengan peningkatan akurasi 12.19 dan f1-score sebesar 12 jika dibandingkan dengan hanya menggunakan fitur baseline.

<hr />

Emoji is a picture character used in informal communication especially in social media. Emoji used by message writer to express emotion of a text. The massive use of emoji make emoji have a great influence on social media communication. Emoji used as one of the features for sentiment analysis and mood extraction In Natural Language Processing and Information Retrieval Researches, yet there is still researches that focus to predict emoji from a text. Due to diversity of emoji and the similarity meaning between emoji, emoji classification task is more relative complex than common text classification task. This researched used semantic feature, linguistic feature, and lexicon feature used to know the influence of each feature on emoji classification task and the best combinaton feature in emoji classification performan. The experiment showed that semantic feature has the best performance in emoji classification when it used individually. Whereas lexicon feature has the greatest positive influence in baseline feature. The analysis using paired t test showed that combination of two features and three features increase baseline performance significantly. The best combination achieved when combined baseline feature, semantic feature, linguistik feature, and lexicon feature with accuration excalation about 12.19 and f1 score of 12 from baseline.