CKIP at IJCNLP-2017 Task 2: Neural Valence-Arousal Prediction for Phrases

Peng-Hsuan Li
National Taiwan University

Wei-Yun Ma
Academia Sinica

Hsin-Yang Wang
Academia Sinica

Task

To predict VA for Chinese phrases, functions mapping Chinese character sequences to real numbers are built by regression techniques.

Main Idea

Utilize a range of regression models to capture the relationship between the modifier words and the head word of each Chinese phrase.

Major Results

The predictor took the second place out of 13 teams on phrase VA prediction, with 0.444 MAE and 0.935 PCC on valence, and 0.395 MAE and 0.904 PCC on arousal.

Chinese Phrase Segmentation ← (可能不同意)

With dictionaries of known head words and modifier words, for each phrase, we perform

1. Suffix longest matching of a head word, and then
2. (Iteratively) Prefix longest matching of modifier words.

Phrase Features ← (可能, 不, 同意)

The CKIP phrase VA predictor depends on knowledge of modifier words and head words. Including

- The types of known modifier words,
- VA of head words, and
- Distributional semantics of both these words, trained from ASBC and CNA corpora.

Regression Model ← $(\boldsymbol{Emb}_{可能}, \boldsymbol{Emb}_{不}, \boldsymbol{Emb}_{同意}, \boldsymbol{VA}_{同意}, \boldsymbol{Mod}_{(可能,不)})$

From the most generalizable to the most powerful: Ridge, SVR-RBF, Feed-Forward NN. Best configuration: 4 layer NN with shape (750, 600, 600, 450).

Discovery

Our approach achieves the second place out of 13 teams on phrase VA prediction.

Model	Valence MAE	Valence PCC	Arousal MAE	Arousal PCC
Official Linear Baseline	1.051	0.610	0.607	0.730
CKIP-Run2	0.444	0.935	0.395	0.904
THU_NGN-Run1	0.349	0.960	0.389	0.909
THU_NGN-Run2	**0.345**	**0.961**	**0.385**	**0.911**

Feature analyses show the effectiveness of capturing the non-linear relationship between head word VA and modifier existence.

	Valence MAE	Valence PCC	Arousal MAE	Arousal PCC
h, m, va, mod	0.41	0.95	0.37	**0.92**
m, va, mod	**0.36**	**0.96**	**0.36**	**0.92**
va, mod	0.45	0.93	0.40	0.90
va	1.34	0.44	0.73	0.67
mod	1.31	0.35	0.71	0.66