Radical and Word Entanglements: Take "女" as an Example

Pu Meng   pmeng@gmu.edu   George Mason University

The relationship between the part and the whole has been a concern of people since the emergence of logic. In linguistics, the relationship between radical and the word has been vaguely summarized as follows: as a part of the word, radical endows and affects semantics. To give a more scientific and rational explanation to the long-standing problem, this study collected 477 words containing radical "女" from the latest edition of Xinhua Dictionary. The clustering data is divided into 6 classes by meaning and corresponding extracted contextual word embeddings from a Chinese BERT model. This unsupervised machine learning observes the relationship between classifiers in terms of distribution, joint probability, and usage. In addition, the number distribution of 6 semantic classes, the position of radical, and the number of strokes in the word are also analyzes to help prove the research results: words with radical "女" can be divided into classes of "female", "Quality", "Movement", "Name", "Emotion", and "Phenomenon", and the word distribution should be affected by the frequency of use. The study also has broader implications for the language type distribution for computational research.