

Improvements in Similarity Measurement Method for Emotion Classification of Japanese Sentences

Daichi NAYA* and Takehiko YOSHIMI**

** PERSOL AVC TECHNOLOGY CO., LTD.,
Takatsuki, Osaka, 569-1194, Japan*

***Faculty of Advanced Science and Technology, Ryukoku University,
Otsu, Shiga, 520-2194, Japan*

Abstract

This study proposes the following three improvements to address the shortfalls of existing methods for emotion classification based on similarity measures between Japanese sentences. (1) The proposed method utilizes, as a similarity measure, an automatic evaluation metric for machine translation based on correlation considering global word order. Similarity measures used in existing methods are too strict or lax in their consideration of the order of words during the processing of Japanese sentences. They fail to incorporate the fact that Japanese exhibits a higher degree of freedom in word order compared to English. (2) Rather than word-level alignment, the proposed method aligns two sentences at the level of a base phrase constituted by one or more content words followed by zero or more function words. Word-level alignment, used by existing methods, may cause inappropriate matching, because it independently processes content words and function words which constitute a base phrase. (3) The proposed method integrates, using linear interpolation, similarity scores by the alignment of word-level and those of base phrase-level. Integration can avoid data sparsity, namely a problem that a similarity score between two sentences becomes zero, occurring when base phrase-level alignment is made. The experimental results found that (1) the proposed method demonstrated significantly higher classification accuracy than existing alternatives; (2) the base phrase-level alignment was effective because the alignment exhibited higher classification accuracy than the word-level alignment; and (3) integration of the word-level and base phrase-level alignments was valid because the integrated method outperformed the base phrase-level alignment.

Key Words: Emotion classification, Similarity measure, Japanese, Word order, Base phrase, Linear interpolation