Title	感情表出の顕現性と書き手特性を考慮した書き手の感情強度推定
Author(s)	岡留, 司真
Citation	
Issue Date	2026-03
Type	Thesis or Dissertation
Text version	author
URL	https://hdl.handle.net/10119/20551
Rights	
Description	Supervisor:白井　清昭, 先端科学技術研究科, 修士(情報科学)

Estimating Writer's Emotion Intensity in Text Considering Emotion Salience and Writer-Specific Characteristics

2210038 Kazuma Okadome

In recent years, emotion recognition in text has become an active area of research in the field of natural language processing. Among various tasks, emotion intensity estimation has attracted particular attention as an important and challenging problem. Emotion intensity estimation aims not only to classify the type of emotion expressed in a text, but also to estimate how strongly the writer feels that emotion. This task plays a crucial role in practical applications such as dialogue systems, sentiment analysis, and opinion mining, where understanding the degree of emotional involvement is essential. However, many existing studies assume that emotions are explicitly expressed in text. As a result, they often overlook cases in which a writer's emotions are not directly reflected in linguistic expressions, as well as discrepancies between the emotions actually felt by the writer and those interpreted by readers. Consequently, most conventional approaches rely heavily on surface-level cues, such as emotion words or intensifiers, to estimate emotion intensity. These methods tend to struggle with texts that contain implicit emotional expressions, making it difficult to accurately estimate the true emotion intensity of the writer.

This study aims to estimate the writer's emotion intensity more accurately by considering both the salience of emotion and the writer's characteristics. Salience of emotion refers to the degree to which a writer's emotion is explicitly expressed in the text. In this study, it is measured by the difference between the emotion intensity annotated by a writer and that annotated by a reader. When the discrepancy between the writer's and reader's emotion intensity scores is large, the emotion is considered to be weakly expressed, indicating low salience of emotion. Conversely, a small difference suggests that the emotion is clearly expressed in the text. In addition to salience of emotion, this study also focuses on the writer ' s characteristics. Some writers tend to express their emotions directly and explicitly, while others prefer indirect or implicit expressions. Such individual differences in emotional expression are important factors in estimating emotion intensity. By modeling these tendencies as writer's characteristics, we can capture writer-specific patterns of how emotions are expressed in text. This study proposes an approach that incorporates both salience of emotion and writer's characteristics into an emotion intensity estimation model.

To construct the emotion intensity estimation model, we use WRIME, a Japanese emotion dataset that includes annotations from both writers and multiple readers. A distinctive feature of WRIME is that it includes emotion intensity scores assigned by the writers themselves as well as by readers for the same text. This allows for a detailed analysis of discrepancies in emotion recognition between writers and readers. Furthermore, WRIME provides a writer ID for each text, enabling the modeling of writer-specific characteristics. This study adopts two experimental settings when splitting WRIME into training and test data: "Open Task" and "Closed Task." In the Open Task setting, texts written by the authors in the test data do not appear in the training data, whereas in the Closed Task setting, they do. By comparing these two settings, we examine how much writer's characteristics and salience of emotion contribute to improving emotion intensity estimation performance when past texts written by the target writer are available or unavailable during training.

This study proposes two models for emotion intensity estimation: a late fusion model and a multi-task learning model. The late fusion model estimates emotion intensity by integrating multiple types of information derived from text. Specifically, it extracts three kinds of feature vectors representing the semantic meaning of the text, the salience of emotion, and the writer's characteristics. These feature vectors are concatenated and fed into a Fully Connected Layer (FCL) to predict the emotion intensity. The semantic feature vector is obtained from a fine-tuned pre-trained language model, BERT, which captures contextual and semantic information from the input text. The feature vector representing salience of emotion is obtained from a model referred to as SE-BERT. In this study, the "salience of emotion estimation task" is formulated as a binary classification problem that determines whether there is a difference between the emotion intensity annotated by the writer and that annotated by the reader. A BERT-based model is fine-tuned to solve this task, and the resulting model is defined as SE-BERT. The feature vector representing the writer's characteristics is obtained from another BERT-based model, referred to as ID-BERT. In this case, the "writer estimation task" is defined as the task of predicting the writer ID from a given text. A BERT model is fine-tuned to solve this task, and the resulting model is used as ID-BERT. During the training of the late fusion model, the parameters of SE-BERT and ID-BERT are fixed and not updated. Only the parameters of the BERT model used to obtain the semantic feature vector and the parameters of the FCL for emotion intensity estimation are updated.

In the multi-task learning model, emotion intensity estimation is treated as the main task, while salience of emotion estimation task is treated as an auxiliary task. A pre-trained BERT model is used as a common base model for the two tasks, and FCL for obtaining the outputs of the main task and the auxiliary task are respectively placed on top of it. The parameters of the entire model are updated using a loss function that combines the losses of the two tasks.

Several experiments were conducted to evaluate the effectiveness of the proposed methods. Using the WRIME dataset, emotion intensity for each of the eight emotion classes are estimated. First, we discuss the results of the late fusion model. In the Open Task setting, incorporating the salience of emotion into the model led to improved performance in emotion intensity estimation. In contrast, the writer's characteristics did not show effectiveness under this setting. Furthermore, combining salience of emotion and writer's characteristics did not result in additional performance gains. This result indicates that, in situations where texts written by the same writer cannot be used during training, it is difficult to sufficiently capture writer-specific tendencies, and that information about the extent to which emotions are explicitly expressed in the text plays a major role in estimating the writer's emotion intensity.

On the other hand, in the Closed Task setting, it was found that writer's characteristics work effectively for emotion intensity estimation. Furthermore, combining salience of emotion with writer's characteristics resulted in higher performance than using either feature alone. This result demonstrates that, under conditions where the writer is known, the model is able to learn writer-specific emotional expression tendencies to some extent, and that writer's characteristics plays a complementary role with the salience of emotion in emotion intensity estimation.

For the multi-task learning model, no clear performance improvement was observed by incorporating salience of emotion in either the Open Task or Closed Task settings. This result suggests that although emotion intensity estimation and salience of emotion estimation are related tasks, the features required for each task may not be fully aligned. In other words, simultaneous optimization of these two tasks may interfere with the learning of task-specific representations, potentially hindering performance improvements for emotion intensity estimation.