Master's Thesis

Softlabel Classification for Multimodal Sentiment Estimation using Multiple Third-party Annotations

ZHAO ZHEN

Supervisor SHOGO OKADA

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Information Science)

August 2024

Abstract

This thesis explores the challenges and proposes novel solutions in the field of multimodal sentiment analysis, with a particular focus on enhancing the accuracy of self-sentiment (SS) estimation in human-agent dialogue systems. The research addresses two critical issues in current sentiment analysis: the discrepancy between self-reported sentiments and third-party annotated sentiments, and the subjective nature of sentiment annotation leading to disagreements among annotators.

The study utilizes two shared multimodal dialogue datasets, Hazumi1902 and Hazumi1911, which contain rich multimodal data including linguistic, audio, and visual features. These datasets are unique in that they provide both self-reported sentiment labels (SS) and third-party annotated sentiment labels (TS), allowing for a comprehensive analysis of the differences between these two types of sentiment annotations.

A key contribution of this research is the development of a novel soft labeling approach for sentiment classification. This method addresses the inherent subjectivity in sentiment annotation by representing the sentiment as a probability distribution over possible classes, rather than as a single hard label. The soft labels are generated by considering the annotations from multiple third-party annotators, capturing the nuances and disagreements in their judgments.

The thesis presents a series of experiments that demonstrate the effectiveness of the proposed approach. Initially, a baseline model is established using a simple deep neural network (DNN) architecture that integrates audio, video, and text features. This baseline model is then compared with human-level performance, revealing a significant gap in accuracy.

To bridge this gap, the research explores several innovative strategies. First, it investigates the impact of using TS labels for samples where TS and SS labels are inconsistent. This approach shows a marked improvement over the baseline, highlighting the importance of addressing label inconsistencies. Building on this, the study introduces the soft label method, which proves to be even more effective. The soft label approach not only improves overall accuracy but also captures valuable information from minority annotators who may detect subtle emotional states that the majority miss. Furthermore, the research proposes a weighted loss function that assigns different importance to samples based on the consistency between their TS and SS labels. This technique further enhances the model's performance, bringing it closer to human-level accuracy.

This research explores approaches to address the challenge of estimating self-sentiment in dialogue systems, proposes methods for handling annotator disagreements, and investigates the potential of soft labels in representing the complex nature of human emotions. The findings of this study may have implications for the development of sentiment analysis systems, particularly in human-agent interaction contexts. By attempting to better capture the nuances of human emotion, these approaches could potentially contribute to improvements in human-computer interactions.