

Title	書と属性に対する感情分析のマルチタスク学習に関する研究
Author(s)	小泉, さやか
Citation	
Issue Date	2025-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/20050
Rights	
Description	Supervisor: 白井 清昭, 先端科学技術研究科, 修士 (情報科学)

A Study on Multi-Task Learning of Document-Level and Aspect-Based Sentiment Analysis

2330007 Sayaka Koizumi

In recent years, with the proliferation of e-commerce sites and social media, users have more opportunities to express their opinions about products and services. For companies providing products and services, as well as for people considering the purchase of those products or utilization of those services, reviews that include users' impressions and opinions serve as extremely valuable sources of information. Given this background, sentiment analysis, which is a task to classify subjective evaluations within text, has attracted significant attention across a wide range of fields, including marketing, recommendation systems, and customer satisfaction analysis. Among the various subtasks of sentiment analysis, Aspect-Based Sentiment Analysis (ABSA), which determines the polarity (positive, negative, or neutral) for specific aspects such as "price," "customer service," or "cleanliness," enables more fine-grained analysis of user opinions. However, to train models for ABSA, training data is required in which a polarity label is assigned to each aspect, and constructing such data demands substantial effort. In particular, for the Japanese language, the development of datasets for ABSA remains insufficient, and the lack of training data poses a significant barrier.

The goal of this research is to explore methods for ABSA with high accuracy in scenarios where only a limited amount of training data is available. Specifically, the study addresses the problem of data sparseness by leveraging a dataset labeled with polarity at the document level as auxiliary data. On many websites that facilitate user reviews, users can assign evaluation scores for target products or services. It is relatively easy to assign polarity labels to entire reviews based on users' scores; thus, a large-scale dataset can be relatively easily constructed. Regarding ABSA as the main task and document-level sentiment analysis as the auxiliary task, this research employs multi-task learning to improve the performance of ABSA by utilizing the information obtained from the auxiliary task's labeled data as external knowledge.

This study performs multi-task learning of aspect-based and document-level sentiment analysis. In our multi-task learning framework, BERT is used as an encoder shared across these two tasks. Furthermore, two models for multi-task learning are proposed. The first is a basic multi-task learning model, referred to as MTL-Basic. An input text is encoded by BERT, and the encoded output is then fed into two distinct classification layers corresponding to each task. Sharing the encoder enables learning of the classification

model from diverse information contained in multiple datasets, which is expected to offer benefits such as suppression of overfitting and the enhancement of the generalizability of abstract representations of review texts. The second model is a multi-task learning model that uses a shared intermediate layer, referred to as MTL-Shared. To further enhance information sharing between the two tasks, the intermediate layer shared between tasks is added after the BERT encoder. The integration of features common to both document-level and aspect-level sentiment analysis within this shared layer is expected to enhance the performance of both the main and auxiliary tasks.

In addition to the above two multi-task learning models, a filtering method is proposed to enhance the quality of the dataset used for document-level sentiment analysis, which serves as the auxiliary task. Only one polarity label is assigned to each document in the dataset of document-level sentiment analysis. It causes inconsistency between the content of a review and its polarity label. For example, even when both positive and negative opinions are included in a review, either a “positive” or “negative” label is assigned. Even the “neutral” label is assigned to such a review, as the reviewer may give a neutral score due to holding both positive and negative opinions across different aspects. To address this issue, reviews that contain both positive and negative expressions are removed from the dataset. Specifically, a sentiment lexicon is employed for the filtering. A review is discarded when both positive and negative words in the sentiment lexicon appear in it. After applying this filtering process, a polarity classification model is trained using MTL-Basic. This model is referred to as MTL-Refined.

Two datasets were used in the experiments to evaluate the effectiveness of the proposed method. For ABSA, the “Rakuten Travel Review: Aspect and Sentiment-tagged corpus” (“ABSA corpus” in short) was used. This dataset contains polarity labels for seven types of aspects, including “location,” “room,” “service,” and so on. For document-level sentiment analysis, a collection of reviews posted on Rakuten Travel, which is a part of the Rakuten dataset, was used. Polarity labels were automatically assigned based on the 5-point rating scores provided by users at the time of review submission. For both datasets, balanced datasets are constructed by randomly selecting an equal number of positive, neutral, and negative samples. The balanced datasets consist of 10,000 samples in total. The balanced dataset of the ABSA corpus is not perfectly balanced as a number of samples of the minor class is less than one-third of 10,000. To conduct experiments, the datasets are subdivided into 70% training data, 20% development data, and 10% test data. In addition, to verify how the size of the training data influences the model's performance, the number of samples in the ABSA dataset was diminished. Specifically, we prepared datasets of four different sizes, i.e., 100%, 50%, 25%, and 10% of the

original dataset. While the size of the training and development data is reduced, the test data remains constant to ensure a fair comparison of the models trained from the different-sized training data. Besides, the size of the dataset for document-level sentiment analysis is not decreased, considering that it can be relatively easily constructed. The performance of ABSA models was evaluated using the accuracy and F1 score. A single-task learning model was used as the baseline and compared with the three proposed multi-task learning models.

Experimental results showed that there was no significant difference in performance between the multi-task learning models and the single-task learning model when the training data was sufficiently large. On the other hand, when the size of the training data was limited to 10%, the multi-task learning models outperformed the single-task model. In particular, for polarity labels that are difficult to classify, such as “negative” and “neutral”, MTL-Basic showed improvements of up to 3 points in both the accuracy and F1 score compared to the single-task learning model. Moreover, multi-task learning models demonstrated smoother transitions in the accuracy and F1 score during training, with a tendency to attain high performance even in the early stages of training (with fewer epochs). MTL-Shared was effective for some aspects, but showed a strong bias toward the “positive” label, i.e., MTL-Shared classified many test samples as positive. MTL-Refined improved the accuracy for the “neutral” label, although improvements for the other labels were limited. A case study analyzing correctly and wrongly predicted samples revealed that the multi-task learning models could correctly predict polarity even for short reviews or reviews including euphemistic expressions.

Future work will focus on further improving the flexibility and performance of multi-task learning models through dynamic adjustment of task loss weights, refinement of parameter-sharing strategies between tasks, and incorporation of attention mechanisms. Furthermore, efforts will be made to enhance the interpretability of ABSA models and to explore their practical applications.