

Title	クラスタリングに基づく弱教師あり学習によるレビューの暗黙的属性の分類
Author(s)	AYE AYE MAR
Citation	
Issue Date	2024-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/19067
Rights	
Description	Supervisor: 白井 清昭, 先端科学技術研究科, 博士

Implicit Aspect Classification of Online Reviews by Clustering-based Weak Supervision

1920405 Aye Aye Mar

Aspect Term Extraction (ATE) which is a process of extracting an aspect (also known as opinion target) from a customer review sentence plays a vital role in Aspect-Based Sentiment Analysis (ABSA). Many previous work of ATE focused on explicit aspect but only a few work considered to extract implicit aspects. However, customer reviews containing implicit aspects are widespread on the Web (such as Amazon.com) and these sentences are also important to fully understand the opinions and sentiments of the customers.

One of the bottleneck problem of implicit aspect extraction is lack of a large dataset of reviews annotated with implicit aspects. Although the corpus annotated with implicit aspects is required for every domain due to different types of aspects in different domains, constructing a corpus is labour intensive and time consuming. Therefore, a system to automatically construct a dataset annotated with implicit aspects is required. This study proposed a novel approach that automatically constructs a dataset annotated with implicit aspects using unlabelled Amazon reviews to address the challenge of implicit aspect extraction. To the best of our knowledge, no prior work has been performed on the automatic construction of such a dataset.

The goal of this study is to develop a system of ATE for implicit aspects. A dataset labeled with implicit aspects is automatically constructed by guessing implicit aspects in unlabeled review sentences. The proposed method involves clustering review sentences labeled with explicit aspects (which were extracted by CRF model trained on golden explicit review sentences) and unlabeled review sentences. In this study, using a K-means clustering approach with a relatively large number of clusters (10% of total review sentences) aims to generate many small but accurate clusters.

Cluster labels, considered as implicit aspects, are automatically assigned based on the assumption that sentences with similar context share a common aspect. When selecting the most relevant cluster label among the explicit aspects in the cluster, the frequency of the aspect in the list of aspects extracted by CRF and its occurrence in the review sentences within the cluster are considered to determine the relevance of the chosen cluster label. When there is more than one aspect that can be the cluster label, we did not consider such kind of cluster since the cluster label is not unique. Moreover, the reliability of the cluster label to be chosen was determined by the threshold value (T_r). Unlabeled sentences in clusters matching pre-defined implicit aspect categories are then obtained as implicit-aspect-labeled sentences. To

increase the number of clusters related with the implicit aspects, the aspect synonym list was identified.

The accuracy of the constructed corpus was evaluated by a human annotator by checking manually on 50 random sentences for each implicit aspect. The results showed that accuracy of the sentences in the constructed corpus was reasonably high, i.e., from 0.58 to 0.82. The study presents findings and observations regarding with constructing the corpus annotated with implicit aspects.

In this study, implicit aspect extraction problem is formulated as classification problem. Then, BERT model is fine-tuned for implicit aspect classification using the constructed dataset by investigating the best values of hyper-parameters. Experiments results of implicit aspect classification show that our method achieves 82% and 84% accuracy for the mobile phone and PC reviews respectively, which are 20 and 21 percentage points higher than the baseline.

Furthermore, the study explores the impact of explicit review sentences for implicit aspect classification by combining the explicit sentences and implicit sentences and then by training classification model on the combined dataset. The experimental results showed that it further boosts the performance of implicit aspect classification in both phone and PC domain.

Keywords: Aspect-based Sentiment Analysis, Aspect Extraction, Implicit Aspect, Weakly-supervised Learning, Online Review