Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction

and They Are BothWeakly Supervised

1 Introduction

The majority of work on opinion summarization is entity-centric, aiming to create summaries from text collections that are relevant to a particular entity of interest, e.g., product, person, company, and so on.

A popular decomposition of the problem involves three subtasks:

(1) aspect extraction which aims to find specific features pertaining to the entity of interest (e.g., battery life, sound quality, ease of use) and identify expressions that discuss them

(2) sentiment prediction which determines the sentiment orientation (positive or negative) on the aspects found in the first step

(3) summary generation which presents the identified opinions to the user

In this paper, we present a neural framework for opinion extraction from product reviews. Central to our system is the ability to accurately identify aspect specific opinions by using different sources of information freely available with product reviews (product domain labels, user ratings) and minimal domain knowledge (essentially a few aspect denoting keywords).

2 Related Work

Aspect Extraction

Sentiment Prediction

Multi-document Summarization

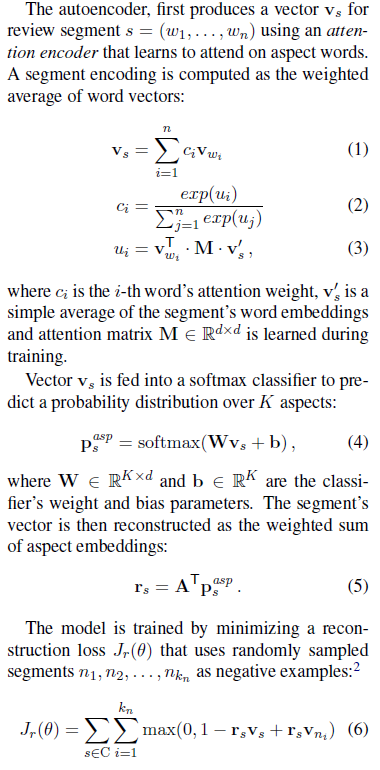
3 Problem Formulation

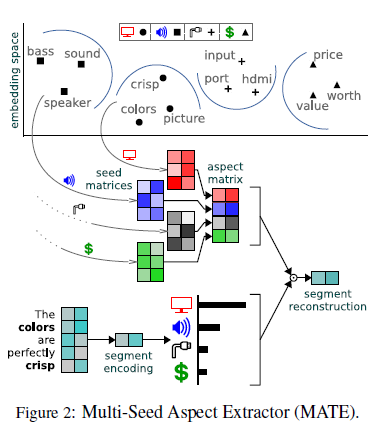
4 Aspect Extraction

Multi- Seed Aspect Extractor (MATE) model

4.1 Aspect-Based Autoencoder

The model learns a segment-level aspect predictor without supervision by attempting to reconstruct the input segment’s encoding as a linear combination of aspect embeddings.





4.2 Multi-Seed Aspect Extractor

4.3 Multi-Task Objective

MATE (and ABAE) relies on the attention encoder to identify and attend to each segment’s aspect-signalling words. The reconstruction objective only provides a weak training signal, so we devise a multi-task extension to enhance the encoder’s effectiveness without additional annotations. It is important not to use the out-of-domain segments for segment reconstruction, as they will confuse the aspect extractor due to the aspect mismatch between different domains.

5 Opinion Summarization

Opinion Polarity

Opinion Ranking

Opinion Selection

6 The OPOSUM Dataset

Aspects

Opinion Summaries

7 Experiments

Implementation Details: Reviews were lemmatized and stop words were removed.

Aspect Extraction

Opinion Salience

Opinion Summaries

8 Conclusions

We presented a weakly supervised neural framework for aspect-based opinion summarization. Our method combined a seeded aspect extractor that is trained under a multi-task objective without direct supervision, and a multiple instance learning sentiment predictor, to identify and extract useful comments in product reviews. We evaluated our weakly supervised models on a new opinion summarization corpus across three subtasks, namely aspect identification, salient opinion extraction, and summary generation. Our approach delivered significant improvements over strong baselines in each of the subtasks, while a large-scale judgment elicitation study showed that crowd workers favor our summarizer over competitive extractive and abstractive systems.