Conference on Empirical Methods in Natural Language Processing

EMNLP 2018

Author Response

Title: ExtRA: Extracting Prominent Review Aspects from Customer Feedback

Authors: Zhiyi Luo, Shanshan Huang, Frank F. Xu, Bill Yuchen Lin, Hanyuan Shi and Kenny Zhu

Instructions

Dear authors,

The author response period for EMNLP 2018 has begun.
This page shows you the reviews for your submission, and allows you to enter comments that will be shared with the reviewers, area chairs and program chairs.
At the bottom of this page, you also have the option to enter confidential comments that will only be shared with the area and program chairs, but this should only be used if absolutely necessary.

You can enter your response until **Friday, July 13** (23:59pm Pacific Daylight Saving Time (UTC -7h)).

Guidelines

- For an explanation of our review form and the scores used for the overall recommendation, see http://emnlp2018.org/reviewform/
- After your reviews, you will first see one text box for each of your reviews and one box for general comments to the reviewers. Please use this first set of text boxes to respond to the points raised in the reviews, or to answer any questions the reviewers may have asked. This information will be shared with all reviewers.
- You can use the last text box for confidential comments that will not be shared with the reviewers.
- Please note: *you are not obligated to respond to the reviews*.
- You are also not allowed to present new results.
- The overall word limit (i.e. the total number of words, summed over all boxes) is 600.

Reviews

Review #1

What is this paper about, and what contributions does it make?

This paper addresses extracting prominent aspects for a given domain. The problem is different from sentence-level aspect extraction in the sense that extracted aspect words do not necessarily appear in a sentence, but should be readable and generally in the domain. The paper argues that traditional approaches do not understand that certain words discuss the same aspects. It proposes involving external knowledge and a PPR-based ranking approach. The proposed method outperforms others on specially prepared labels.

What strengths does this paper have?

1. The approach seems reasonable. The superiority is also demonstrated in experiments.

What weaknesses does this paper have?

1. I am rather uncertain about the necessity of the prominent aspect extraction problem or setting. Since the predefined domains are rather clear to ordinary people, and only five prominent aspects are needed, why not tell the domain name to an ordinary people and ask her to give the five aspects without reading the potentially large review corpus? I believe that the aspects are rather static over time, so direct human labors are worthwhile compared with automatic algorithms.
2. The authors claim that their model is superior in the sense that it is aware of different words expressing similar semantics. The model is also explicitly designed following this guide. However, it is not clear whether other baselines implicitly address this point. Only final results are observed in the experiment section.

Overall Recommendation: 3

Review #2

What is this paper about, and what contributions does it make?

The main contributions are author-identified and include a framework and evaluation dataset for prominent review aspect extraction, and experiments exhibiting how the developed unsupervised method outperforms the state of the art.

What strengths does this paper have?

The authors identify, describe, and approach an important commercial research gap—automatically identifying prominent review aspects.

The aspect taxonomy (graph) is a clever approach to avoid avoid from extracting semantically-overlapping aspects, and to support aspect ranking.

What weaknesses does this paper have?

-

The quantitative evaluation (Sec. 3.3.1) is complex. In addition to being difficult to interpret/intuit, no examples are provided. Thus, the paper would benefit significantly from detailed examples of how a set of prominent aspects is evaluated upon output.

- The paper's ordering is a bit strange (see presentation suggestions), and the conclusion is light. A bit more interpretation of the work and discussion of directions for improvement would better support the paper.

-

Overall Recommendation: 3.5

Questions for the Author(s)

none

Missing References

none

Presentation Improvements

-

Please make the Fig. 1 example user review larger (try to match the text size in the Fig. with the document).

- It could significantly help to orient readers if Sec. 4, Related Work, is placed at the beginning of the paper, ahead of Sec. 2. It's strange to read about related work

right before the end, specifically, all the way after the framework, dataset, and experiments are detailed. This also breaks the flow of the paper into the conclusion.

-

Please consider providing examples motivating each extraction rule. This will greatly support reader comprehension and intuition-building. It's okay if the examples come from Fig. 3, but it could help to have these presented separately, all in one place. Also, there's a typo on this example in the last paragraph of Sec. 2.2: "quite" should be "quiet".

-

Footnotes should be placed after punctuation (#1 should be after the comma).

-

Typos, Grammar, and Style

none

Review #3

What is this paper about, and what contributions does it make?

The paper is about extracting aspects for products on websites like Yelp and Amazon. The authors employ a simple but well-thought out unsupervised methodology for extracting these aspects.

What strengths does this paper have?

The paper is well written and easy to follow. The literature survey is strong. The data-set might be useful for the community.

What weaknesses does this paper have?

Topics are not meant to be aspects of products. So extracting K topics will of course result in "things" that do not correspond to aspects. A more fair comparison might have been the extraction of 2K or 3K topics and then hand-picking the topics that correspond to aspects.

Quantitatively and qualitatively, there doesn't seem to be a massive difference between AmodExt and ExtRA. AmodExt is rule based and uses past work AutoPhrase. Would've liked to see a comparison of AutoPhrase with ExtRA

Is the evaluation suggested in Liu et al 2017 not applicable here?

Would like to see if the results are statistically significantly different, comparing AmodExt and ExtRA. Absolute numbers don't fully give a picture here because the evaluation set is so small.

Suggestion: Determinental point processes for selecting the most diverse and high quality aspects.

Overall Recommendation: 3.5

Submit Response to Reviewers

Use the following boxes to enter your response to the reviews. Please limit the total amount of words in your comments to 600 words (longer responses will not be accepted by the system).

Response to Review #1:

Response to Review #2:

Response to Review #3:

General Response to Reviewers:

Response to Chairs

You can use this textbox to contact the area and program chairs directly.
The reviewers will not see what you write here.

Please use this option only when there are serious issues regarding the reviews.
Such issues can include reviewers who grossly misunderstood the submission, or have made unfair comparisons or requests in their reviews. Most submissions should not need to use this option.

Submit

START Conference Manager (V2.61.0 - Rev. 5257)