MRCBert: A MRC Approach for Unsupervised Summarization

Saurabh Jain[†], Lim Sze Chi, [‡] and Guokai Tang[‡]

[†] Department of Computer Science, National University of Singapore

Motivation

When doing online shopping, it becomes important for the customer to read the product reviews efficiently and make a decision quickly. However, reviews can be lengthy, contain repeated or irrelevant information that does not help in decision making. Hence, the focus of our work demonstrates aspect-wise reviews summarisation to provide only key information about a product in an automatic and unsupervised manner.

Contributions

Our key contributions are:

- A novel unsupervised approach and accompanied framework of using machine reading comprehension (MRC) and deep-learning to extract relevant rating-wise and aspect-wise opinions from reviews.
- Transfer learning using existing pre-trained MRC and summarisation deep-learning models to obtain reasonable performance with minimal training, which is useful for learning under limited/low resource scenarios.
- Our methodology does not require domain-specific data set for training and requires only a set of curated aspects to achieve the desired outputs.
- Introduced a unique sentiment accuracy metric to perform unsupervised evaluation.

Extracting Opinions

Input to the model:

$$x = ([CLS], q1, \ldots, qm, [SEP], d1, \ldots, dn, [SEP])$$

where $q = (q1, \ldots, qm)$ is a question asking opinion on an aspect; $d = (d1, \ldots, dn)$ is a review and $[CLS]$ and $[SEP]$ are a dummy token and a token to separate q and d respectively.

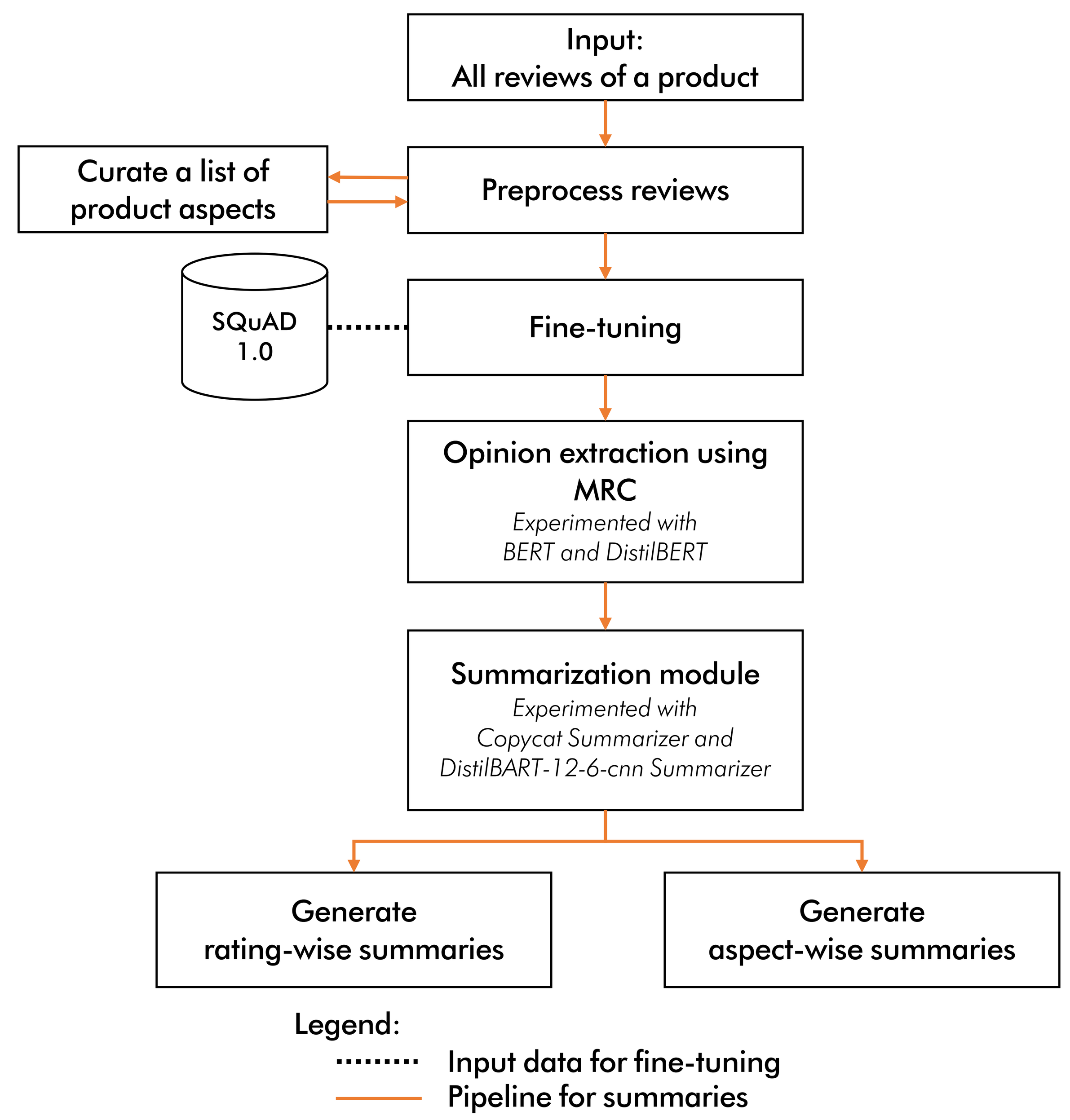
$$h = BERT(x) R^{\text{rh} * |\text{x}|}$$

where $BERT(.)/DistilBERT(.)$ is a fine-tuned $BERT$ and $DistilBERT$ model; $|x|$ is the length of a sequence and rh is the size of the hidden dimension.

$$Loss = -\frac{I(s)\sum \log L1 + I(e)\sum \log L2}{2}$$

where $I(s)$ and $I(e)$ are one-hot vectors representing ground truth values of spans, where s and e are start and end indices of span in the review.

Summarization Framework



Metrics

Sentiment Accuracy Metric

We trained a sentiment prediction model on Amazon Review Dataset [1] , the model predicts score bewteen [0, 5], We make the weights of both BERT layers trainable so that it can capture review-related linguistic knowledge during training. Our sentiment model is trained on domain specific data and therefore it provided better accuracy in determining the sentiment in the reviews domain.

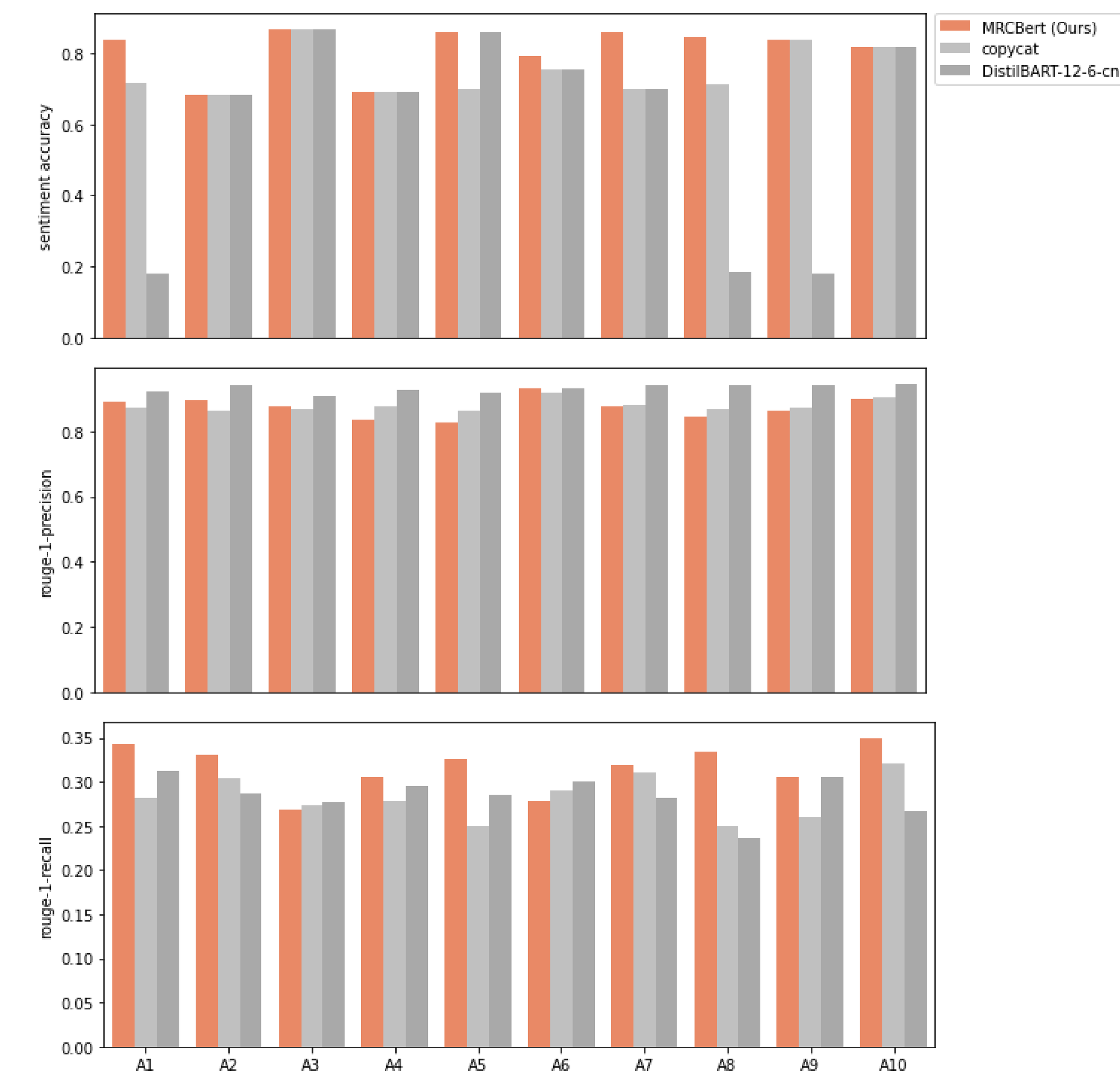
$$S_{sentiment} = 1 - \frac{1}{N}\sum_{i=1}^{N}\left[\log_6\left(\left|\frac{1}{k}\sum_{j=1}^{k} R(x_j) - S(s_i)\right| + 1\right)\right]$$

ROUGE Scores

To evaluate if our summary contains similar information that is found in the set of reviews that the summary comes from. We use ROUGE-1 score and ROUGE-2 to measure overlap of unigram (each word) and bigram (two words) between the reviews and summary respectively.

$$S_{Rouge} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{k}\sum_{j=1}^{k} ROUGE_N(s_i, x_j))\right]$$

Results



where A1-A10 are aspects on display, wireless connection, battery, OS, processor, brand, memory, sound, camera and quality of speaker respectively.

Our model captures sentiment better than copycat and DistillBART-12-6-cnn separately. The generated summaries also used relevant words from the original reviews as seen from the high precision. While recall is low, the low occurrence of 1-grams from the input reviews present in the generated summaries is intuitive as the input reviews are usually lengthy and may contain irrelevant information or incorrect use of words.
<I will add a table here tomorrow>

Sample summary about an iPad display: Product arrived on time and in good condition, works great, easy to use, good graphics. Screen is a bit darker than I expected it to be but it works great. Screen protector is not clear on the screen, but screen resolution is not as clear as the screen resolution. Sim card is a good size, but not compatible with sim card.

References

[1] Jianmo Ni, Jiacheng Li, and Julian McAuley. "Justifying Recommendations using Distantly-Labeled Reviews and Fine-Grained Aspects". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 188–197. DOI: 10.18653/v1/D19-1018. URL: https://www.aclweb.org/anthology/D19-1018.