Natural Language Processing 365

Data Science   Natural Language Processing   NLP Papers Summary

Day 152: NLP Papers Summary – OPINIONDIGEST: A Simple Framework For Opinion Summarization

By Ryan   31st May 2020   No Comments

Objective and Contribution

Presented OPINIONDIGEST, an abstractive opinion summarisation framework that uses Aspect-based Sentiment Analysis (ABSA) to extract opinion phrases from reviews and trains a Transformer to reconstruct the original reviews from these extracted opinion phrases. We merge opinion phrases from multiple reviews and select the most popular ones at summarisation time. OPINIONDIGEST can also generate summaries based on dif   user needs, by filtering selected opinion based on the user's aspect and / or sentiment. We show that

our model was able to outperformed other baseline models and produces informa ∧ ve customisable summaries.

OPINIONDIGEST Framework

For each entity, our goal is to generate abstractive summaries of the most salient opinions in the relevant reviews. OPINIONDIGEST has three main components:

1. Opinion Extraction
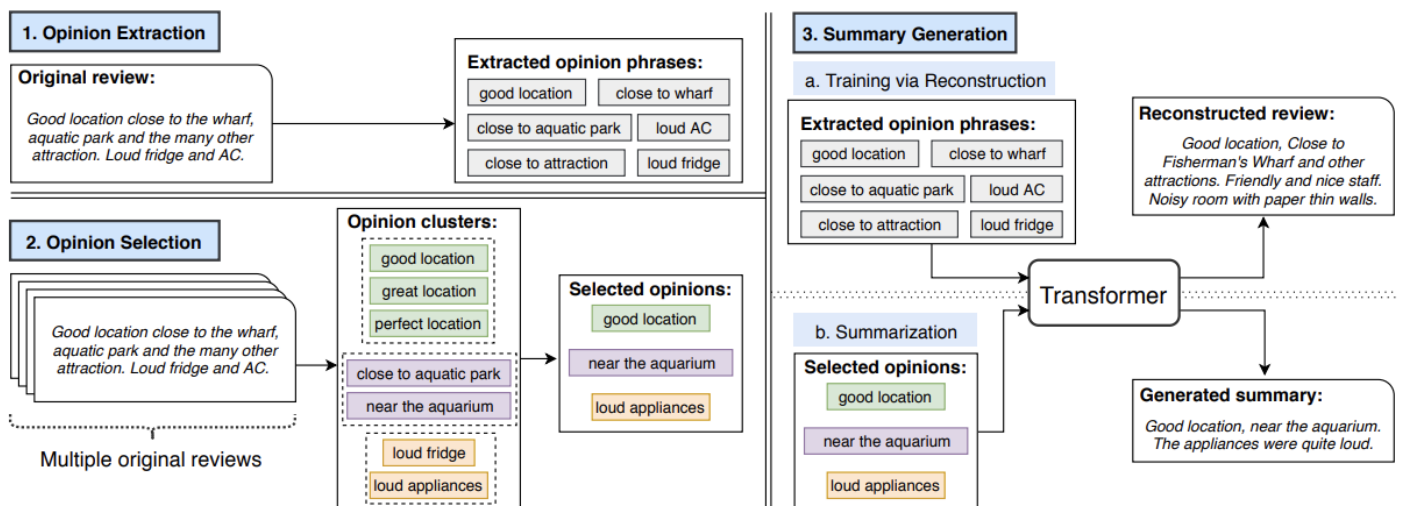2. Controllable Opinion Selection
3. Summary Generation

Figure 1: Overview of the OPINIONDIGEST framework.

OPINION EXTRACTION

Here, we used a pre-trained tagging model to obtain our opinion set that contains extracted opinion phrases and its respective polarity and aspect categories.

OPINION SELECTION

For each review that contains the entity, we want to extract the entity's opinion set. Summarising the opinion sets of each entity involves selecting the most salient opinions. Here, we have three operations: opinion merging, ranking, and filtering (optional). Opinion merging is where we used greedy algorithm and word embeddings to merge similar opinions (measured using cosine similarity) into different clusters. We would then rank these cluster   inion ranking) based on size where we assume the larger a cluster is, the more popular the opinions

are within the cluster. The last operation is opinion filtering (optional) where we contr ⌃ he selection process by filtering opinions based on aspect category or sentiment.

SUMMARY GENERATION

Here, our goal is to reconstruct reviews from our extracted opinion phrases and generate opinion summaries from our selected opinions. We train our transformer model by encoding extracted opinion phrases of a single review and learning to reconstruct the review's full text. We then used the trained transformer to generate summaries. During summarisation, we would use the selected opinions (concatenated together) as an input to our trained transformer model to generate the respective summary.

Experiments and Results

We have to evaluation datasets: YELP restaurant reviews and HOTEL reviews. The YELP dataset has 624K training reviews whereas the HOTEL has 688K reviews. Our evaluation metrics are ROUGE-1, -2, and -L. We have 3 baseline models: LexRank, MeanSum, and the Best/Worst Review (the highest/lowest average word overlap with input reviews).

RESULTS

Method	R1	R2	RL
Best Review	27.97	3.46	15.29
Worst Review	16.91	1.66	11.11
LexRank	24.62	3.03	14.43
MeanSum	27.86	3.95	16.56
OPINIONDIGEST	**29.30**	**5.77**	**18.56**

Table 1: Summarization results on YELP with ROUGE.

The results above on the YELP dataset shows that our framework outperformed all the baseline models. Our framework is not an unsupervised model but we only required labelled data for our opinion extractor model. To further evaluate our framework, we conducted human evaluation to evaluate the quality of our generated summaries. Judges are asked to evaluate the summary based on three criteria:

1. Informativeness (I-score)

2. Coherence (C-score)

3. Non-Redundancy (R-score)

The results are showcase in the table 2 below and OPINIONDIGEST's summaries scored the highest in both informativeness and coherence, however, the summaries still need to work on reduce redundancy. We also performed content support study where judges were given 8 input reviews from YELP, and are required to evaluate the extent to which each sentence is supported by the reviews. Table 3 below showcase the proportion of sentences in the summary that are fully, partially and not supported by the input reviews. We show that OPINIONDIGEST was able to generate summaries that have a higher proportion of fully and partially supported sentences.

Lastly, we want to measure how well our framework can generate summaries based on different aspects. We produced aspect-specific summaries and asked the participants to evaluate if the summaries includes the specified aspects exclusively, partially, or not   '. The

results are displayed below and we show that OPINIONDIGEST can generate summarie at are exclusively or partially related to specified aspects 89.7% of the time.

QUALITATIVE OUTPUT

Below are example summaries that showcase OPINIONDIGEST's ability to summaries more 100+ reviews and also how we can control the summaries using opinion filtering of aspects and sentiments. The first two examples showcase the ability of our model to generate summaries based on 8 and 128 reviews. Our framework doesn't aggregate review representations and so the scalability of our model is not tied down to the input size. Example 3 – 6 showcase how we can control the generated summaries using aspects and sentiments. Example 3 – 4 control the generated summaries using sentiments whereas Example 5 – 6 uses different aspects (staff vs food).

Source: https://arxiv.org/pdf/2005.01901.pdf

Ryan

Data Scientist

Next Post

Previous Post

Day 151: NLP Papers Summary - A Large-Scale Multi-Document Summarization Dataset from the Wikipedia Current Events Portal

Day 153: NLP Papers Summary - Span-ConveRT: Few-shot Span Extraction for Dialog with Pretrained Conversational Representations

Data Science　**Natural Language Processing**　**NLP Papers Summary**

Day 365: NLP Papers Summary – A Survey on Knowledge Graph Embedding

Ryan
30th December 2020

Data Science   **Implementation**   **Natural Language Processing**

Day 364: Ryan's PhD Journey – OpenKE-PyTorch Library Analysis + code snippets for 11 KE models

Ryan
29th December 2020

Data Science　Ryan's PhD Journey

Day 363: Ryan's PhD Journey – Literature Review – Knowledge Acquisition – 1st Passes XIV

Ryan
28th December 2020