HAL
open science

Unsupervised Aspect-Based Neural Abstractive Summarization of User-Generated Reviews

Maximin Coavoux, Hady Elsahar, Matthias Gallé

To cite this version:

Maximin Coavoux, Hady Elsahar, Matthias Gallé. Unsupervised Aspect-Based Neural Abstractive Summarization of User-Generated Reviews. EurNLP, Oct 2019, London, United Kingdom. hal-02874104

HAL Id: hal-02874104

https://hal.science/hal-02874104v1

Submitted on 29 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised Aspect-Based Neural Abstractive Summarization of User-Generated Reviews

Maximin Coavoux[1→2] & Hady Elsahar[1] & Matthias Gallé[1]

[1]Naver Labs Europe

[2]LIG, CNRS, Université Grenoble Alpes

maximin.coavoux@univ-grenoble-alpes.fr,
{matthias.galle,hady.elsahar}@naverlabs.com

Opinion Summarization

Input = large set of reviews (typically ≈ 300 sentences) for a place-of-interest or a product:

```
I bought this for my son as a Christmas gift .  After less than 3 weeks of ordinary use , the
strap is ripping away from the case .  I guess it 's okay for storage or infrequent use , but
it 's too flimsy for daily use .
...
I have a Dell 1150 14 '' screen and I had to return the case because it was n't big enough .
Also , the hole in the back pocket (for sliding on the luggage carrier) seems a little big --
do n't know if I would trust papers not to fall out of there .
```

Output = summary of the reviews

```
The quality of the bag is outstanding.  The design is perfect to fit the laptop, batteries,
pens, paperwork, etc..  The padding inside really keeps the laptop secure.  Nice looking, and
well built.  There is NO way this thin!  thin!  thin!  case would protect a pencil let alone
a heavy laptop.  but will brake due to poor quality.
```

[Extract from Oposum dataset, Angelidis and Lapata, 2018]

Opinion summarization is a type of multidocument summarization task. Some challenges for opinion summarization:

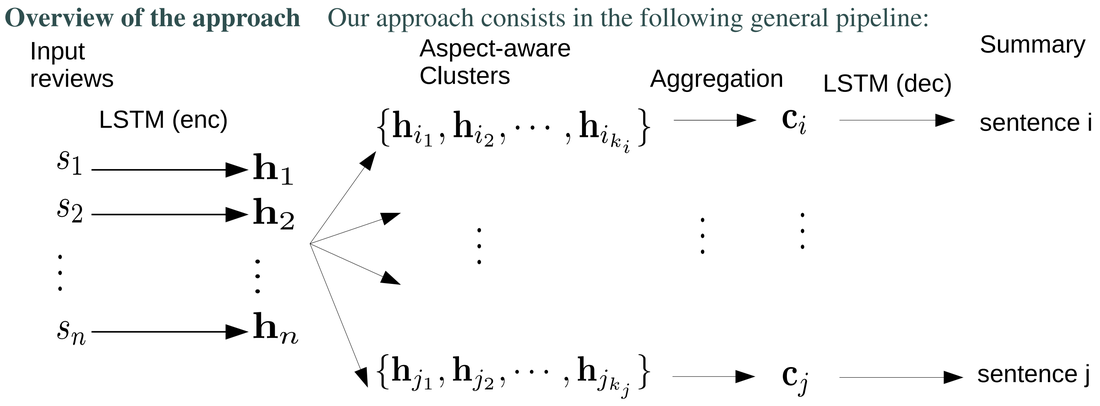
- Contradictory opinions
- Non-relevant information
- Multiple aspects (ex: about a restaurant, quality of food, of drinks, price, location, atmosphere)

Prior Work

- Most prior work on unsupervised (or weakly supervised) opinion summarization is **extractive**, i.e. select a subset of sentences supposed to be representative of the reviews [Radev et al., 2004, Mihalcea and Tarau, 2004, Angelidis and Lapata, 2018].
- **Abstractive** unsupervised methods are either based on word graphs [Ganesan et al., 2010] or more recently on the seq2seq paradigm [Chu and Liu, 2019].
- Our proposal is based on the approach by Chu and Liu [2019, **MeanSum**], who leverages a neural seq2seq sentence auto-encoder to perform summarization. In MeanSumn, the model computes the **average** representation of all available reviews and uses it to seed the decoder of the auto-encoder. In our work, we explore other ways to aggregate representations and construct a summary and argue that the mean of reviews' representations is not a good summary of the representations.

Model

Overview of the approach Our approach consists in the following general pipeline:



1. **Encoding step**: sentence ⟶ vector (LSTM)
2. **Clustering step**: cluster sentence representations into meaningful groups (i.e. cluster together sentences that are about the same aspect). We experiment with:
   - Aspect classification
   - Kmeans
3. **Aggregation step**: compute a single vector representation for each cluster, from the representations of sentences in the cluster. We first **prune** the cluster to the 16 most salient elements (see next section) and compute their average. This is a form of **hard attention** over sentence representations.
4. **Generation step**: generate a sentence for each cluster, using the same LSTM as in the encoding step.

Injecting Knowledge with Multitask Learning

- We use **K-means** as the baseline clustering method.
- As a more involved strategy, we use an **aspect classifier**, trained jointly with the LSTM language model. → requires a small amount of annotated data.
  - The input to the aspect classifier is the sentence representation, as computed by the LSTM.
  - At test time, we generate 1 sentence per aspect.
- **Pruning** based on **polarity** predictions: we also use a polarity classifier (3 classes: positive, neutral, negative) to perform pruning in each cluster. For each cluster we keep the 16 sentences for which the polarity classifier decision is the most confident

Training Objective

We use a multitask objective:

- Language modelling ($\mathcal{L}_{\text{lm}}$)
- Aspect classification ($\mathcal{L}_{\text{aspect}}$)
- Polarity classification ($\mathcal{L}_{\text{polarity}}$)

$$\mathcal{L}_{\text{lm}} = \sum_{i=1}^{n} -\log P(w_i | w_0^{i-1}; \theta_{\text{LSTM}}),$$
$$\mathcal{L}_{\text{polarity}} = -\log P(y_p | w_0^n; \theta_{\text{LSTM}}, \theta_{\text{polarity}}),$$
$$\mathcal{L}_{\text{aspect}} = -\log P(y_a | w_0^n; \theta_{\text{LSTM}}, \theta_{\text{aspect}}),$$
$$\mathcal{L}_{\text{MTL}} = \mathcal{L}_{\text{lm}} + \mathcal{L}_{\text{polarity}} + \mathcal{L}_{\text{aspect}},$$

Experiments and results

- **Dataset**: **Oposum** [Angelidis and Lapata, 2018], reviews for 6 product types (each model is for a single product type)
- **Decoding**: for each cluster, we perform top-k sampling [Fan et al., 2018] 10 times and rerank sentences based on their similarity to the cluster centroid.

Model Variations

Setting	Clustering	Training objective
Mean	No clustering (like MeanSum)	$\mathcal{L}_{\text{lm}}$
Kmeans	K-means	$\mathcal{L}_{\text{lm}}$
Kmeans + MTL	K-means	$\mathcal{L}_{\text{MTL}}$
Aspect + MTL	supervised aspect classification	$\mathcal{L}_{\text{MTL}}$

ROUGE-L evaluation per product type

Model	Bags_and_cases	Bluetooth	Boots	Keyboards	TV	Vacuums
TextRank	0.35	0.28	0.31	0.30	0.30	0.30
Mean	0.18 ±0.03	0.15 ±0.02	0.16 ±0.02	0.17 ±0.02	0.16 ±0.03	0.15 ±0.02
Kmeans	0.38 ±0.02	0.37 ±0.01	0.37 ±0.01	0.37 ±0.01	0.35 ±0.01	0.38 ±0.02
Kmeans + MTL	0.38 ±0.01	0.36 ±0.01	**0.38** ±0.02	0.35 ±0.01	0.35 ±0.02	0.36 ±0.02
Aspect + MTL	**0.4** ±0.02	**0.38** ±0.01	**0.38** ±0.01	**0.38** ±0.01	**0.37** ±0.01	**0.39** ±0.01

ROUGE-{1, 2, L} metrics on the full dataset

Model	ROUGE-1	ROUGE-2	ROUGE-L
TextRank	0.27 ±0.02	0.03 ±0.0	0.31 ±0.02
Mean	0.12 ±0.02	0.01 ±0.01	0.16 ±0.03
Kmeans	0.32 ±0.02	0.05 ±0.01	0.37 ±0.02
Kmeans + MTL	0.31 ±0.02	0.05 ±0.01	0.36 ±0.02
Aspect + MTL	**0.33** ±0.02	0.05 ±0.01	**0.38** ±0.02
Angelidis and Lapata [2018]	**0.44**	**0.21**	**0.43**

Discussion

- **Clustering is crucial** to obtain good results.
- MTL training has no effect when using K-means clustering.
- Aspect classification leads to sightly better results than K-means, despite very small amount of aspect-annotated data for training the classifier (700 sentences).
- Our model still falls short of the extractive system of Angelidis and Lapata [2018].

References

Stefanos Angelidis and Mirella Lapata. Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3675–3686, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/D18-1403.

Eric Chu and Peter Liu. MeanSum: a neural model for unsupervised multi-document abstractive summarization. pages 1223–1232, 2019.

Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 889–898, Melbourne, Australia, July 2018. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/P18-1082.

Kavita Ganesan, ChengXiang Zhai, and Jiawei Han. Opinosis: A graph based approach to abstractive summarization of highly redundant opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 340–348, Beijing, China, August 2010. Coling 2010 Organizing Committee. URL https://www.aclweb.org/anthology/C10-1039.

Rada Mihalcea and Paul Tarau. TextRank: Bringing order into text. In *Proceedings of EMNLP 2004*, pages 404–411, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/W04-3252.

Dragomir R. Radev, Hongyan Jing, Małgorzata Styś, and Daniel Tam. Centroid-based summarization of multiple documents. *Information Processing and Management*, 40(6):919 – 938, 2004. ISSN 0306-4573. doi: https://doi.org/10.1016/j.ipm.2003.10.006. URL http://www.sciencedirect.com/science/article/pii/S0306457303000955.