Doing Resarch in Natural Language Processing

Session 13: Scientific Writing: Abstracts

Frank Keller

1 November 2023

School of Informatics
University of Edinburgh
keller@inf.ed.ac.uk

Overview

Importance

Content

Organization

Reading: Alley (2018), pp. 139–143, Zobel (2014), p. 57.

Please also look at Alley's web site, which has a lot of videos and additional materials: https://www.craftofscientificwriting.com/

Recap

In previous sessions on scientific writing, we talked about:

- audience, purpose, occasion
- precision and clarity
- structure

In this session, we will apply what we've learned to one of the most important parts of a paper: **the abstract.**

Importance

Why is the Abstract Important?

For readers:

- find out what the paper is about (beyond the information in the title)
- get a summary of the most important findings
- decide whether they want to read the paper or not

For reviewers:

- find out whether the paper is within their area of expertise
- decide whether they want to review the paper or not
- form a first opinion about the quality of the paper

Many readers will **only read the title and the abstract.** So this is the one chance to get your message across!

Why is the Abstract Important?

The abstract can also be a tool for the writer:

- helps you decide what your most important points are
- helps you clarify the overall argumentation of the paper
- provides a way of repeating important information
- allows you to influence who will read (and review!) the paper

Content

Content of the Abstract

According to Alley (who calls it summary), the abstract should:

- contain the most important points of the paper
- contain **only** the important points
- only include material that occurs elsewhere in the paper (verbatim or paraphrased)
- be self-contained, i.e., the reader should be able to understand the abstract without having to read anything else
- this means unusual terms, techniques, etc. need to be explained in the abstract
- don't assume all readers will be specialists in the topic of the paper; assume a broad readership.

Content of the Abstract

Alley distinguishes:

- **informative summary:** describes the most important results of a paper;
- **descriptive summary:** states what kind of information the paper provides (like a table of contents), but doesn't give the actual results.

The abstract of a conference or journal paper is a mixture of both: it provides signposting (which information to expect in the paper), but also summarizes the results.

Content of the Abstract

Zobel offers the following practical advice:

- the abstract is typically a single paragraph of about 50–200 words
- it presents a summary of the paper's aims, scope, and conclusions
- do not use acronyms, mathematics, abbreviations, citations (the abstract should be self-contained!)
- be as specific as possible (instead of *we improve the state of the art*, write things like *we improve the state of the art by 3.5%*)
- but only include **important** details.

Organization

Organization of the Abstract

Zobel suggests to start by writing one sentence on each of the following:

1. A general statement introducing the broad research area.
2. An explanation of the specific problem to be solved.
3. A review of existing solutions and their limitations.
4. An outline of the proposed new solution.
5. A summary of how the solution was evaluated and the result of the evaluation.

So you start with five sentences, but then you can add additional sentences, re-write the ones you have, merge them, etc.

Organization of the Abstract

My own experience shows:

- longer documents may require longer abstracts: the abstract of a journal paper is somewhat longer than that of a conference paper
- the abstract of a PhD thesis is typically a whole page; it should summarize each (content) chapter
- abstracts can contain sentences extracted from the main body of the text (you may need to edit them for coherence)
- but: it is sometimes a good strategy to write the abstract **before** writing the paper – helps planning the overall argumentation, deciding what to focus on
- and then once the paper is finished, you need to completely re-write the abstract!

Over to You

Exercise 1: Evaluating an Abstract

Look at the abstracts on the next pages. Investigate the following questions:

1. Does the abstract follow Zobel's structure?
2. Is it targeted at a general reader? Is it self-contained?
3. Does it avoid acronyms, mathematics, abbreviations, citations?
4. Does it contain (only) the most important points of the paper?
5. Is enough detail provided, and is all the detail important?

All abstracts are from papers that appeared at ACL 2020.

Exercise 1: Evaluating an Abstract

Abstract of Li et al. (2020):

Recently many efforts have been devoted to interpreting the black-box NMT models, but little progress has been made on metrics to evaluate explanation methods. Word Alignment Error Rate can be used as such a metric that matches human understanding, however, it can not measure explanation methods on those target words that are not aligned to any source word. This paper thereby makes an initial attempt to evaluate explanation methods from an alternative viewpoint. To this end, it proposes a principled metric based on fidelity in regard to the predictive behavior of the NMT model. As the exact computation for this metric is intractable, we employ an efficient approach as its approximation. On six standard translation tasks, we quantitatively evaluate several explanation methods in terms of the proposed metric and we reveal some valuable findings for these explanation methods in our experiments.

Exercise 1: Evaluating an Abstract

Abstract of Yang et al. (2020):

We present easy-to-use retrieval focused multilingual sentence embedding models, made available on TensorFlow Hub. The models embed text from 16 languages into a shared semantic space using a multi-task trained dual-encoder that learns tied cross-lingual representations via translation bridge tasks (Chidambaram et al., 2018). The models achieve a new state-of-the-art in performance on monolingual and cross-lingual semantic retrieval (SR). Competitive performance is obtained on the related tasks of translation pair bitext retrieval (BR) and retrieval question answering (ReQA). On transfer learning tasks, our multilingual embeddings approach, and in some cases exceed, the performance of English only sentence embeddings.

This was a system demonstration paper. How does that affect the writing?

Unsupervised Opinion Summarization as Copycat-Review Generation

Arthur Bražinskas[1], **Mirella Lapata**[1] and **Ivan Titov**[1,2]
[1] ILCC, University of Edinburgh
[2] ILLC, University of Amsterdam
abrazinskas@ed.ac.uk, {mlap,ititov}@inf.ed.ac.uk

arXiv:1911.02247v2 [cs.CL] 19 Apr 2020

Abstract

Opinion summarization is the task of automatically creating summaries that reflect subjective information expressed in multiple documents, such as product reviews. While the majority of previous work has focused on the extractive setting, i.e., selecting fragments from input reviews to produce a summary, we let the model generate novel sentences and hence produce abstractive summaries. Recent progress in summarization has seen the development of supervised models which rely on large quantities of document-summary pairs. Since such training data is expensive to acquire, we instead consider the unsupervised setting, in other words, we do not use any summaries in training. We define a generative model for a review collection which capitalizes on the intuition that when generating a new review given a set of other reviews of a product, we should be able to control the "amount of novelty" going into the new review or, equivalently, vary the extent to which it deviates from the input. At test time, when generating summaries, we force the novelty to be minimal, and produce a text reflecting consensus opinions. We capture this intuition by defining a hierarchical variational autoencoder model. Both individual reviews and the products they correspond to are associated with stochastic latent codes, and the review generator ("decoder") has direct access to the text of input reviews through the pointer-generator mechanism. Experiments on Amazon and Yelp datasets, show that setting at test time the review's latent code to its mean, allows the model to produce fluent and coherent summaries reflecting common opinions.

1 Introduction

Summarization of user opinions expressed in online resources, such as blogs, reviews, social media, or internet forums, has drawn much attention due

Summary	This restaurant is a hidden gem in Toronto. The food is delicious, and the service is impeccable. Highly recommend for anyone who likes French bistro.
Reviews	We got the steak frites and the chicken frites both of which were very good ... Great service ... || I really love this place ... Côte de Boeuf ... A Jewel in the big city ... || French jewel of Spadina and Adelaide , Jules ... They are super accommodating ... moules and frites are delicious ... || Food came with tons of greens and fries along with my main course , thumbs uppp ... || Chef has a very cool and fun attitude ... || Great little French Bistro spot ... Go if you want French bistro food classics ... || Great place ... the steak frites and it was amazing ... Best Steak Frites ... in Downtown Toronto ... || Favourite french spot in the city ... crème brule for dessert

Table 1: A summary produced by our model; colors encode its alignment to the input reviews. The reviews are truncated, and delimited with the symbol '||'.

generation (Hu and Liu, 2004; Angelidis and Lapata, 2018; Medhat et al., 2014). Although there has been significant progress recently in summarizing non-subjective context (Rush et al., 2015; Nallapati et al., 2016; Paulus et al., 2017; See et al., 2017; Liu et al., 2018), modern deep learning methods rely on large amounts of annotated data that are not readily available in the opinion-summarization domain and expensive to produce. Moreover, annotation efforts would have to be undertaken for multiple domains as online reviews are inherently multi-domain (Blitzer et al., 2007) and summarization systems highly domain-sensitive (Isonuma et al., 2017). Thus, perhaps unsurprisingly, there is a long history of applying unsupervised and weakly supervised methods to opinion summariza-

Exercise 2: Shortening an Abstract

The abstract of Bražinskas et al. (2020) takes up almost half a page. Can you shorten it to 125 words?

Opinion summarization is the task of automatically creating summaries that reflect subjective information expressed in multiple documents, such as product reviews. While the majority of previous work has focused on the extractive setting, i.e., selecting fragments from input reviews to produce a summary, we let the model generate novel sentences and hence produce abstractive summaries. Recent progress in summarization has seen the development of supervised models which rely on large quantities of document-summary pairs. Since such training data is expensive to acquire, we instead consider the unsupervised setting, in other words, we do not use any summaries in training. We define a generative model for a review collection which capitalizes on the intuition that when generating a new review given a set of other reviews of a product, we should be able to control the "amount of novelty" going into the new review or, equivalently, vary the extent to which it deviates from the input. At test time, when generating summaries, we force the novelty to be minimal, and produce a text reflecting consensus opinions. We capture this intuition by defining a hierarchical variational autoencoder model. Both individual reviews and the products they correspond to are associated with stochastic latent codes, and the review generator ("decoder") has direct access to the text of input reviews through the pointer-generator mechanism. Experiments on Amazon and Yelp datasets, show that setting at test time the review's latent code to its mean, allows the model to produce fluent and coherent summaries reflecting common opinions.

Exercise 3: Writing Your Own Abstract

Write a draft of the abstract of your replication report. Use Zobel's guidelines and write one sentence each for:

1. A general statement introducing the broad research area.
2. An explanation of the specific problem to be solved.
3. A review of existing solutions and their limitations.
4. An outline of the proposed new solution.
5. A summary of how the solution was evaluated and the result of the evaluation.

References

ACL. 2020. *Proceedings of the 58rd Annual Meeting of the Association for Computational Linguistics*.

Alley, Michael. 2018. *The Craft of Scientific Writing*. Springer, New York, NY, 4 edition.

Bražinskas, Arthur, Mirella Lapata, and Ivan Titov. 2020. Unsupervised opinion summarization as copycat-review generation. In ACL, pages 5151–5169.

Li, Jierui, Lemao Liu, Huayang Li, Guanlin Li, Guoping Huang, and Shuming Shi. 2020. Evaluating explanation methods for neural machine translation. In ACL, pages 365–375.

Yang, Yinfei, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-hsuan Sung, Brian Strope, and Ray Kurzweil. 2020. Multilingual universal sentence encoder for semantic retrieval. In *Proceedings of the 58rd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. pages 87–94.

Zobel, Justin. 2014. *Writing for Computer Science*. Springer, New York, NY, 3 edition.