

---

# Extractive and Abstractive Sentence Labelling of Sentiment-bearing Topics

Mohamad Hardyman BARAWI<sup>1</sup>, Chenghua LIN<sup>2</sup>, Advait SIDDHARTHAN<sup>3</sup>, Yinbin LIU<sup>4</sup>

<sup>1</sup> Computing Science, University of Aberdeen, UK

<sup>2</sup> Department of Computer Science, University of Sheffield, UK

<sup>3</sup> Knowledge Media Institute, The Open University, UK

<sup>4</sup> School of Management, Shanghai University, China

*Front. Comput. Sci.*, **Just Accepted Manuscript** • 10.1007/s11704-019-8228-8

<http://journal.hep.com.cn> on April 22, 2019

© Higher Education Press and Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Just Accepted

This is a “Just Accepted” manuscript, which has been examined by the peer-review process and has been accepted for publication. A “Just Accepted” manuscript is published online shortly after its acceptance, which is prior to technical editing and formatting and author proofing. Higher Education Press (HEP) provides “Just Accepted” as an optional and free service which allows authors to make their results available to the research community as soon as possible after acceptance. After a manuscript has been technically edited and formatted, it will be removed from the “Just Accepted” Web site and published as an Online First article. Please note that technical editing may introduce minor changes to the manuscript text and/or graphics which may affect the content, and all legal disclaimers that apply to the journal pertain. In no event shall HEP be held responsible for errors or consequences arising from the use of any information contained in these “Just Accepted” manuscripts. To cite this manuscript please use its Digital Object Identifier (DOI(r)), which is identical for all formats of publication.”

# Extractive and Abstractive Sentence Labelling of Sentiment-bearing Topics

Mohamad Hardyman BARAWI<sup>1</sup>, Chenghua LIN<sup>2</sup>, Advait SIDDHARTHAN<sup>3</sup>, Yinbin LIU(✉)<sup>4</sup>

<sup>1</sup> Computing Science, University of Aberdeen, UK

<sup>2</sup> Department of Computer Science, University of Sheffield, UK

<sup>3</sup> Knowledge Media Institute, The Open University, UK

<sup>4</sup> School of Management, Shanghai University, China

© Higher Education Press and Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

This paper tackles the problem of automatically labelling sentiment-bearing topics with descriptive sentence labels. We propose two approaches to the problem, one extractive and the other abstractive. Both approaches rely on a novel mechanism to automatically learn the relevance of each sentence in a corpus to sentiment-bearing topics extracted from that corpus. The extractive approach uses a sentence ranking algorithm for label selection which for the first time jointly optimises topic–sentence relevance as well as aspect–sentiment co-coverage. The abstractive approach instead addresses aspect–sentiment co-coverage by using sentence fusion to generate a sentential label that includes relevant content from multiple sentences. To our knowledge, we are the first to study the problem of labelling sentiment-bearing topics. Our experimental results on three real-world datasets show that both the extractive and abstractive approaches outperform four strong baselines in terms of facilitating topic understanding and interpretation. In addition, when comparing extractive and abstractive labels, our evaluation shows that our best performing abstractive method is able to provide more topic information coverage in fewer words, at the cost of generating less grammatical labels than the extractive method. We conclude that abstractive methods can effectively synthesise the rich information contained in sentiment-bearing topics.

**Keywords** Sentiment-topic models, automatic topic labelling

## 1 Introduction

Probabilistic topic models such as latent Dirichlet allocation (LDA) [1] capture the thematic properties of documents by modelling texts as a mixture of distributions over words, known as topics. The words under each topic tend to co-occur together and consequently are thematically related to one another. These topics can therefore be used as a lens for exploring and understanding large archives of unstructured text. Since the introduction of LDA, many extensions have been proposed, an important one being the Joint Sentiment-Topic (JST) model that aims to mine and uncover rich opinion structures from opinionated documents [2]. This work has spurred subsequent research in developing variants of sentiment topic models for a range of opinion mining tasks such as aspect-based sentiment analysis [3], contrastive opinion mining [4], and the analysis of sentiment and topic dynamics [5].

JST is a hierarchical topic model which can detect sentiment and topic simultaneously from opinionated documents. The hidden topics discovered, therefore, are essentially sentiment-bearing topics resembling opinions. This is a key difference compared to the standard topics extracted by LDA which only express thematic information. We exemplify this difference using two topic examples shown in