



Authorship attribution in multi-author settings

Tim Althoff, Denny Britz, Zifei Shan

Authorship attribution

- **Definition:** Identifying the rightful author of a document
- **Our setting:** Authorship attribution in *multi-authored* scientific publications
- **Applications:** E.g. Double-blind review process

How to Write a Scientific Paper

Vasil Hnatyshin and Andrea Lobo
Computer Science Department
Rowan University
{hnatyshin, lobo}@rowan.edu



Abstract

This document describes a general outline for writing a scientific research paper. Even though this document does not follow any specific format, it provides general guidelines to writing a paper or technical document. Usually the publisher will provide a set of specific instructions for the paper formatting (e.g. text and heading fonts, single or double column, references style, etc.) or will state which publication style to follow (e.g. American Psychological Association (APA), Modern Language Association (MLA), etc.). Most scientific papers include the following sections: Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. Some of these sections are optional and can be omitted or combined. In this paper we provide description of each of these sections give some examples.

Keywords: scientific paper, research paper, document structure, writing style

1. Introduction

The ability to write and to express thoughts in a clear manner is a vital step on the path to becoming a well-rounded and educated person. However, often students who study scientific disciplines such as computer science, engineering, physics, and chemistry discount this very important skill. This becomes clear when one examines the technical documents written by undergraduate students. This paper provides some general guidelines to assist undergraduate students in writing better technical documents and scientific / research papers.

Why should you care about this? Well, presumably you are writing the paper to sell your work in some way. In the context of a class, you are trying to convince your Professor that you have done a good job. In the context of a paper you publish at a conference or in a journal, you want to convince others of exactly the same thing. Write a poor paper and people may dismiss your work out of hand. Write a good paper and you may succeed in publicizing your work to a large audience.

Most scientific / research papers are divided into several sections, such as Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. While the actual section names may differ from those mentioned above, the content of the individual sections should generally match this outline. For example, the Introduction will introduce the reader to the general issues addressed in the paper, while the Set-Up / Configuration of the Study describes how the study of the proposed solution was conducted. To simplify this document, we will use the section names mentioned above.

Previous approaches

- Different types of **features**: content-based, network-based and style-based
- Different types of **models**: Similarity-based (clustering), supervised (LR, SVM)
- No focus on multi-authored documents

Our approach

How to Write a Scientific Paper

Vasil Hnatyshin and Andrea Lobo
Computer Science Department
Rowan University
{hnatyshin, lobo}@rowan.edu

Abstract

This document describes a general outline for writing a scientific research paper. Even though this document does not follow any specific format, it provides general guidelines to writing a paper or technical document. Usually the publisher will provide a set of specific instructions for the paper formatting (e.g. text and heading fonts, single or double column, references style, etc.) or will state which publication style to follow (e.g. American Psychological Association (APA), Modern Language Association (MLA), etc.). Most scientific papers include the following sections: Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. Some of these sections are optional and can be omitted or combined. In this paper we provide description of each of these sections give some examples.

Keywords: scientific paper, research paper, document structure, writing style

1. Introduction

The ability to write and to express thoughts in a clear manner is a vital step on the path to becoming a well-rounded and educated person. However, often students who study scientific disciplines such as computer science, engineering, physics, and chemistry discount this very important skill. This becomes clear when one examines the technical documents written by undergraduate students. This paper provides some general guidelines to assist undergraduate students in writing better technical documents and scientific / research papers.

Why should you care about this? Well, presumably you are writing the paper to sell your work in some way. In the context of a class, you are trying to convince your Professor that you have done a good job. In the context of a paper you publish at a conference or in a journal, you want to convince others of exactly the same thing. Write a poor paper and people may dismiss your work out of hand. Write a good paper and you may succeed in publicizing your work to a large audience.

Most scientific / research papers are divided into several sections, such as Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. While the actual section names may differ from those mentioned above, the content of the individual sections should generally match this outline. For example, the Introduction will introduce the reader to the general issues addressed in the paper, while the Set-Up / Configuration of the Study describes how the study of the proposed solution was conducted. To simplify this document, we will use the section names mentioned above.

A

B

Our approach

A

How to Write a Scientific Paper

Vasil Hnatyshin and Andrea Lobo
Computer Science Department
Rowan University
{hnatyshin, lobo}@rowan.edu

Abstract

This document describes a general outline for writing a scientific research paper. Even though this document does not follow any specific format, it provides general guidelines to writing a paper or technical document. Usually the publisher will provide a set of specific instructions for the paper formatting (e.g. text and heading fonts, single or double column, references style, etc.) or will state which publication style to follow (e.g. American Psychological Association (APA), Modern Language Association (MLA), etc.). Most scientific papers include the following sections: Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. Some of these sections are optional and can be omitted or combined. In this paper we provide description of each of these sections give some examples.

Keywords: scientific paper, research paper, document structure, writing style

1. Introduction

The ability to write and to express thoughts in a clear manner is a vital step on the path to becoming a well-rounded and educated person. However, often students who study scientific disciplines such as computer science, engineering, physics, and chemistry discount this very important skill. This becomes clear when one examines the technical documents written by undergraduate students. This paper provides some general guidelines to assist undergraduate students in writing better technical documents and scientific / research papers.

Why should you care about this? Well, presumably you are writing the paper to sell your work in some way. In the context of a class, you are trying to convince your Professor that you have done a good job. In the context of a paper you publish at a conference or in a journal, you want to convince others of exactly the same thing. Write a poor paper and people may dismiss your work out of hand. Write a good paper and you may succeed in publicizing your work to a large audience.

B

Most scientific / research papers are divided into several sections, such as Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. While the actual section names may differ from those mentioned above, the content of the individual sections should generally match this outline. For example, the Introduction will introduce the reader to the general issues addressed in the paper, while the Set-Up / Configuration of the Study describes how the study of the proposed solution was conducted. To simplify this document, we will use the section names mentioned above.

Paper-level

$$P(A \mid \text{Paper}) = 0.7$$

$$P(B \mid \text{Paper}) = 0.5$$

$$P(C \mid \text{Paper}) = 0.6$$

Our approach

A

How to Write a Scientific Paper

Vasil Hnatyshin and Andrea Lobo
Computer Science Department
Rowan University
{hnatyshin, lobo}@rowan.edu

Abstract

This document describes a general outline for writing a scientific research paper. Even though this document does not follow any specific format, it provides general guidelines to writing a paper or technical document. Usually the publisher will provide a set of specific instructions for the paper formatting (e.g. text and heading fonts, single or double column, references style, etc.) or will state which publication style to follow (e.g. American Psychological Association (APA), Modern Language Association (MLA), etc.). Most scientific papers include the following sections: Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. Some of these sections are optional and can be omitted or combined. In this paper we provide description of each of these sections give some examples.

Keywords: scientific paper, research paper, document structure, writing style

1. Introduction

The ability to write and to express thoughts in a clear manner is a vital step on the path to becoming a well-rounded and educated person. However, often students who study scientific disciplines such as computer science, engineering, physics, and chemistry discount this very important skill. This becomes clear when one examines the technical documents written by undergraduate students. This paper provides some general guidelines to assist undergraduate students in writing better technical documents and scientific / research papers.

Why should you care about this? Well, presumably you are writing the paper to sell your work in some way. In the context of a class, you are trying to convince your Professor that you have done a good job. In the context of a paper you publish at a conference or in a journal, you want to convince others of exactly the same thing. Write a poor paper and people may dismiss your work out of hand. Write a good paper and you may succeed in publicizing your work to a large audience.

B

Most scientific / research papers are divided into several sections, such as Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. While the actual section names may differ from those mentioned above, the content of the individual sections should generally match this outline. For example, the Introduction will introduce the reader to the general issues addressed in the paper, while the Set-Up / Configuration of the Study describes how the study of the proposed solution was conducted. To simplify this document, we will use the section names mentioned above.

Paper-level

$$P(A \mid \text{paper}) = 0.7$$

$$P(B \mid \text{paper}) = 0.5$$

$$P(C \mid \text{paper}) = 0.6$$

Sentence-level

$$P(A \mid \text{sentence}) = 0.9$$

$$P(B \mid \text{sentence}) = 0.7$$

$$P(C \mid \text{sentence}) = 0$$

Features

- **Content:** Word n-grams

“distant supervision” “POMDP”

“ , ; , ”

- **Style:** E.g. Word length, punctuation, function word frequencies

“from there”

Preliminary Results

arXiv data: 126 documents, 36 authors.
3 authors per document.

Paper-level precision@1

	Content	Style
Random	0.083	
Paper-Level LR	0.27	0.25
Sentence LR	0.90	0.90
Sentence EM	0.92	0.94

Sentence-level precision@1: ~0.15

Summary & Ongoing Work

- First attempt at disentangling authorship in multi-authored documents
- Sentence-level aggregation leads to better results
- Experimented with real-world dataset
- Including higher-level knowledge (e.g. CRF) promising

Dataset generation

How to Write a Scientific Paper

Vasil Hnatyshin and Andrea Lobo
Computer Science Department
Rowan University
{hnatyshin, lobo}@rowan.edu

Abstract

This document describes a general outline for writing a scientific research paper. Even though this document does not follow any specific format, it provides general guidelines to writing a paper or technical document. Usually the publisher will provide a set of specific instructions for the paper formatting (e.g. text and heading fonts, single or double column, references style, etc.) or will state which publication style to follow (e.g. American Psychological Association (APA), Modern Language Association (MLA), etc.). Most scientific papers include the following sections: Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. Some of these sections are optional and can be omitted or combined. In this paper we provide description of each of these sections give some examples.

Keywords: scientific paper, research paper, document structure, writing style

1. Introduction

The ability to write and to express thoughts in a clear manner is a vital step on the path to becoming a well-rounded and educated person. However, often students who study scientific disciplines such as computer science, engineering, physics, and chemistry discount this very important skill. This becomes clear when one examines the technical documents written by undergraduate students. This paper provides some general guidelines to assist undergraduate students in writing better technical documents and scientific / research papers.

Why should you care about this? Well, presumably you are writing the paper to sell your work in some way. In the context of a class, you are trying to convince your Professor that you have done a good job. In the context of a paper you publish at a conference or in a journal, you want to convince others of exactly the same thing. Write a poor paper and people may dismiss your work out of hand. Write a good paper and you may succeed in publicizing your work to a large audience.

Most scientific / research papers are divided into several sections, such as Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. While the actual section names may differ from those mentioned above, the content of the individual sections should generally match this outline. For example, the Introduction will introduce the reader to the general issues addressed in the paper, while the Set-Up / Configuration of the Study describes how the study of the proposed solution was conducted. To simplify this document, we will use the section names mentioned above.

How to Write a Scientific Paper

Vasil Hnatyshin and Andrea Lobo
Computer Science Department
Rowan University
{hnatyshin, lobo}@rowan.edu

Abstract

This document does not follow any specific format, it provides general guidelines to writing a paper or technical document. Usually the publisher will provide a set of specific instructions for the paper formatting (e.g. text and heading fonts, single or double column, references style, etc.) or will state which publication style to follow (e.g. American Psychological Association (APA), Modern Language Association (MLA), etc.). Most scientific papers include the following sections: Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. Some of these sections are optional and can be omitted or combined. In this paper we provide description of each of these sections give some examples.

Keywords: scientific paper, research paper, document structure, writing style

1. Introduction

The ability to write and to express thoughts in a clear manner is a vital step on the path to becoming a well-rounded and educated person. However, often students who study scientific disciplines such as computer science, engineering, physics, and chemistry discount this very important skill. This becomes clear when one examines the technical documents written by undergraduate students. This paper provides some general guidelines to assist undergraduate students in writing better technical documents and scientific / research papers.

Why should you care about this? Well, presumably you are writing the paper to sell your work in some way. In the context of a class, you are trying to convince your Professor that you have done a good job. In the context of a paper you publish at a conference or in a journal, you want to convince others of exactly the same thing. Write a poor paper and people may dismiss your work out of hand. Write a good paper and you may succeed in publicizing your work to a large audience.

Most scientific / research papers are divided into several sections, such as Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. While the actual section names may differ from those mentioned above, the content of the individual sections should generally match this outline. For example, the Introduction will introduce the reader to the general issues addressed in the paper, while the Set-Up / Configuration of the Study describes how the study of the proposed solution was conducted. To simplify this document, we will use the section names mentioned above.



How to Write a Scientific Paper

Vasil Hnatyshin and Andrea Lobo
Computer Science Department
Rowan University
{hnatyshin, lobo}@rowan.edu

Abstract

This document describes a general outline for writing a scientific research paper. Even though this document does not follow any specific format, it provides general guidelines to writing a paper or technical document. Usually the publisher will provide a set of specific instructions for the paper formatting (e.g. text and heading fonts, single or double column, references style, etc.) or will state which publication style to follow (e.g. American Psychological Association (APA), Modern Language Association (MLA), etc.). Most scientific papers include the following sections: Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. Some of these sections are optional and can be omitted or combined. In this paper we provide description of each of these sections give some examples.

Keywords: scientific paper, research paper, document structure, writing style

1. Introduction

The ability to write and to express thoughts in a clear manner is a vital step on the path to becoming a well-rounded and educated person. However, often students who study scientific disciplines such as computer science, engineering, physics, and chemistry discount this very important skill. This becomes clear when one examines the technical documents written by undergraduate students. This paper provides some general guidelines to assist undergraduate students in writing better technical documents and scientific / research papers.

Why should you care about this? Well, presumably you are writing the paper to sell your work in some way. In the context of a class, you are trying to convince your Professor that you have done a good job. In the context of a paper you publish at a conference or in a journal, you want to convince others of exactly the same thing. Write a poor paper and people may dismiss your work out of hand. Write a good paper and you may succeed in publicizing your work to a large audience.

Most scientific / research papers are divided into several sections, such as Abstract, Keywords, Introduction, Related Work, Problem Statement, Solution Approach, Set-Up / Configuration of the Study or Simulation, Analysis of Results, Conclusions, Future Work, Acknowledgement, and References. While the actual section names may differ from those mentioned above, the content of the individual sections should generally match this outline. For example, the Introduction will introduce the reader to the general issues addressed in the paper, while the Set-Up / Configuration of the Study describes how the study of the proposed solution was conducted. To simplify this document, we will use the section names mentioned above.