How to Write an Independent Work Paper

Xiaoyan Li, IW Coordinator

April 11, 2023

Adapted from slides by Prof. David Dobkin (Fall 2021)

Overview

- Setting the paper in context
- Tools for preparing the paper
- What the paper should look like
- Some words of advice

Opening question…

- What is the difference between the talk you are preparing now and the paper we are about to discuss?

Opening question…

- What is the difference between the talk you are preparing now and the paper we are about to discuss?
  - 9-minute talk vs 25-page paper
  - Due dates: 4/16 vs 5/1
    - Can carry more experiments after 4/16
  - Different skills
  - Different audiences
    - Your paper can be more technical and have more details.
    - Others can reproduce your results.

Before we dig into details…

- When to start writing?
  - NOW! (actually, last month, or February)
- How to view the writing-research process?
  - Iterative.
  - Describing intermediate results yields ideas about other experiments to run, and other data to collect.
  - Advisor feedback can iteratively improve both the report and the research itself.

How about LaTeX?

- Don’t need to use it, but…
- Easiest way to include figures, equations, make citations, cross references...
- Written by programmers for technical writing
- Runs on your computer or web
  - Check out: Overleaf.com
- Try it, you’ll like it!
  - Templates on iw.cs.princeton.edu at
    - https://www.cs.princeton.edu/sites/default/files/uploads/template-20160413.pdf
    - Files to do this at https://www.cs.princeton.edu/courses/archive/www-coursefiles/iw/IWreport.zip
    - Thesis Templates https://static.us.edusercontent.com/files/MR65koh4cnT1RK4PMbf3Ihj9
  - Example at https://www.overleaf.com/read/vwbpbhswvntr

Outline on Web

- Introduction: Motivation and Goals
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions

Enhanced Outline

- Abstract
- Introduction: Motivation and Goals
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Abstract

- Key idea
- Succinct! (About 3-5 sentences)
  - Problem
  - Method
  - Implications/Major findings/achievements
- Not
  - Notation
  - Background

Abstract Example 1

“This paper details the design, development, and evaluation of a Kinect-powered application to facilitate the instruction of ballroom dance. The application uses the Kinect camera’s skeletal tracking capabilities to teach and evaluate users through various ballroom dance positions and concepts, taking the form of a number of training modules that end with a game-like assessment portion. This application aims to fill a hole in ballroom dance instruction, providing the ease of access of self-study materials alongside the quality of instruction of live coaching.”

“Using Kinect To Learn How To Ballroom Dance”
Michael Li, Fall 2015

Abstract Example 2

"The Spotify Million Playlist Dataset is an important resource for researchers studying music recommendations, listening habits, Automatic Playlist Continuation and more. In this paper, I detail findings from an analysis of a 57,000 playlist subset of this dataset using unsupervised learning techniques to offer researchers better context as they use this dataset in their own work. Six clusters were discovered within the subset that can be characterized by different aggregate audio attributes, providing important context to the results of the ACM RecSys Challenge."

"Exploring the Million Playlist Dataset: k-means Clustering of Spotify Playlists Using Aggregate Song Attributes "
Adam Ziff, Spring 2022

Enhanced Outline

- Abstract
- Introduction
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Introduction

- Motivate the work
  - Why do we care?
    - give examples of real-world problems
      - concrete, specific examples are strongest motivators
    - statistics on prevalence of problem helps
- Define the problem you solve clearly
- Be explicit about your contribution!
- Get to the point!
  - state your goals early!
- Could include the organization of the paper

Enhanced Outline

- Abstract
- Introduction
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Background and Related Work

- Context for your work
- What is known?
- What is similar?
- What is different about your project?
- May list several works at once
  - "Several others have proposed approximations [a,b,c]."
- Summarize closest, most important to your work

- You have space (pages)!

Related Work: *Before or After?*

- Pros of discussing related work at beginning
  - Give fuller context for your work
  - Answer the questions of knowledgeable readers
  - Give better understanding of novelty of the work
- Pros of discussing related work at end
  - Readers now know your work and can more easily understand the differences with existing work

Enhanced Outline

- Abstract
- Introduction
- Problem Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Approach

- Big picture driving details to follow
- What is big idea of your solution?
  - Design?
  - Experimental approach?
  - Theoretical approach?
  - New domain?
  - Evaluation metrics?
- What makes it different from previous?
- What makes you choose this approach?

Enhanced Outline

- Abstract
- Introduction
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Implementation

- Give details important to
  - achieving your goals
  - proving your claims
- Can someone reproduce your work from your description of it?
  - Link to your code/dataset in GitHub
  - List important code in appendices
- Why as well as how
  - Many options, compare then decide

Implementation - *Advice*

- This is not a diary!
  - stream-of-consciousness writing does not highlight key ideas
- However, sometimes failed or discarded attempts worth mentioning
  - Do so when provides insight for the “why”
  - Example: “I chose clustering algorithm A among several tested because algorithm B turned out to be too slow, algorithm C didn’t work in this case, …”
  - A lesson for others as well, they do not repeat those failed attempts…

Enhanced Outline

- Abstract
- Introduction
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Evaluation

- How successful is your project?
- What are the criteria for success?
  - Should state these earlier (in Approach)
  - Can be more precise here
- Experiments to show success?
  - Performance evaluation
  - Quantified user studies
  - Comparison to other approaches
- Quantitative measures of success
  - Statistical significance of results
  - Comparison to “gold standards”

Evaluation

- Use figures and tables
  - Clear, readable, highlight best performance
- Summarize the results

Which Table Is Easier to Read?

Logistic Regression Classifier with SFS feature selection:

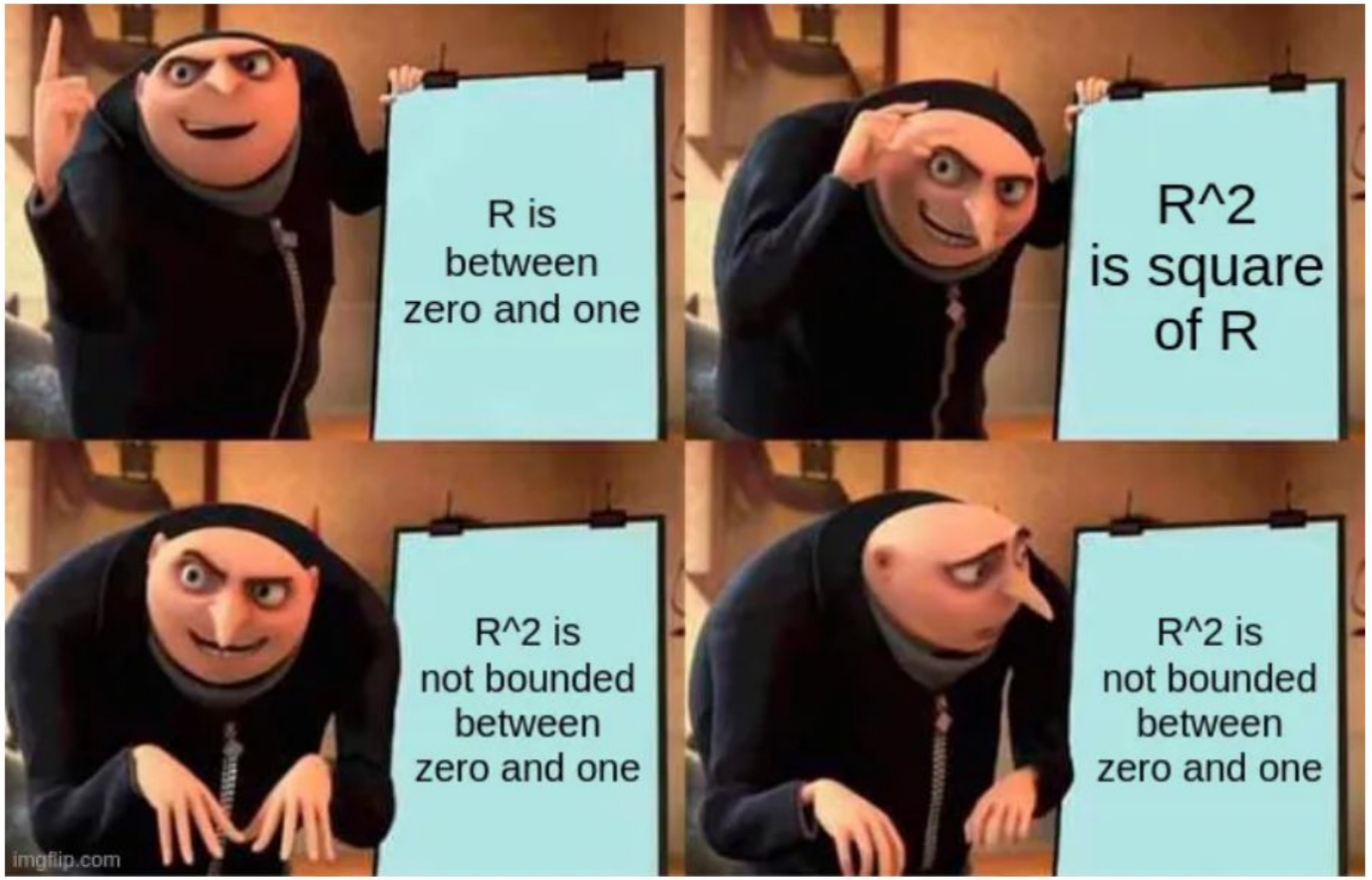
Game Lag	Accuracy	Precision	Recall	F1 Score
1	0.588	0.727	0.667	0.696
2	0.875	0.833	1	0.909
3	0.75	0.8	0.8	0.8
4	0.625	0.8	0.667	0.727
5	0.75	0.818	0.818	0.818
6	0.625	0.692	0.818	0.75

Benchmark	*Sentiment Analysis*	*Technical Indicators*	*Historical Prices*	*Min Volatility*	*Equal Weighted*	*Baseline RL*
Sharpe Ratio	**1.22**	0.92	0.97	1.00	1.03	0.97
Annualized Return (%)	**30.915%**	24.684%	25.084%	23.04%	24.458%	24.172%
Annualized Volatility	24.588%	28.427%	25.595%	**23.403%**	24.458%	25.738%
Final Portfolio Value (USD)	**170,840**	155,052	156,044	150,909	155,467	153,791

Evaluation

- Use figures and tables
  - Clear, readable, highlight best performance
- Summarize the results
- Investigate and Explain strange numbers and patterns
  - R^2: the coefficient of determination in linear regression
  - Usually between 0 and 1
  - Got a negative R^2???

Explaining negative R-squared



If you glossed over the math by instinct, this meme is for you.

https://towardsdatascience.com/explaining-negative-r-squared-17894ca26321

Enhanced Outline

- Abstract
- Introduction
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Conclusions and Future Work

- Can be one or two sections
- Summary of important contributions
  - Draw conclusions based on your experiments
  - Results/findings/lessons
  - No speculations!
- Any limitations of current work
  - Error analysis
- Discuss how you would go forward
- Discuss how others can go forward

Enhanced Outline

- Abstract
- Introduction
- Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Bibliography

- All papers, videos, … you mention in the text
- All tools you use (and so mention in the text)
- Other references you may have used but not cited in the text
  - e.g. background reading
- Relevant private communications
  - e.g. researcher sends you unpublished performance numbers that you use in text

    "Joe Smith, private communication, 2016"

Bibliographic Form

- Many acceptable forms
  - Different publishers, different forms
- Authors, paper title, publication title, publisher, date, pages. (online pointer)

[8] Jon Kleinberg. “Authoritative Sources in a Hyperlinked Environment.” In Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, pp. 668—677. New York: ACM Press, 1998.

Citations in text

- Use number in brackets to refer to biblio. entry:

  "The HITS algorithm[8] also computes a link-based …"

- Using a bibliographic tool makes things easier
  – e.g. bibtex for latex

- Footnotes for asides - <u>sparingly</u>

  "… traverse index in reverse chronological order[2]…"

"[2]Although this is not an absolute requirement …"

Citations in Text - *Style*

- Do not use citations as a noun.
  - You will see lots of published authors do this. It is bad style.
- You have used a citation correctly when it can be removed from the sentence, and it is still a sentence:
  - good: "The HITS algorithm[8] computes ..."
  - bad: "[8] defines the HITS algorithm to compute ..."

Enhanced Outline

- Abstract
- Introduction
- Problem Background and Related Work
- Approach
- Implementation
- Evaluation
- Conclusions and Future Work
- Bibliography
- Appendices

Appendices

- Optional
- Do not expect reader to even peruse them
- Uses
  - Data tables summarized in paper
  - Details of long proof
  - Details interesting to only those very involved
- A luxury of a thesis or “mini thesis”

FAQ: How many pages?

Averages for a small sample of A-level papers

Section	1 sem. proj.	thesis
Introduction	avg. 1.5	2-8, avg. 3.5
Related Work	avg. 4.5	avg. 7.5
Approach	1 – 8, avg. 3	1-8, avg. 3.5
Implementation	avg. 10	avg. 13.5
Evaluation	avg. 5.5	avg. 11.5
Conclusions	avg. 1.25	1-7, avg. 3.5

Table of Contents – *Senior Thesis*

A senior thesis may have ~50 page … A very long document!

- Overall structure
- Easy to navigate

Table of Contents – *Example 1*

- Overall structure
- Easy to navigate

Contents

Table of Contents – *Example 2*

- Overall structure
- Easy to navigate

CONTENTS

Writing Advice – *High Level*

- Write for a general technical audience
  - e.g. all your COS classmates
  - Not for your adviser!
- Don't blur your contributions with those of others.
  - "We know that …" Your result? Someone else's?
- Get feedback on drafts
  - Classmates, friends in CS, advisor, …

Writing Advice - *Approach*

- Put yourself in the place of the reader
  - What does my reader know so far?
  - Am I saying something my reader can't understand given what they know so far?
  - What do they need to know next?
- Eliminate redundancy
  - "What is the information content of this word or sentence?"
  - remove redundant words, phrases, sentences, paragraphs

Writing Advice - *Details*

- Avoid unnecessary complexity and jargon
  - From George Orwell's essay on *Politics and the English language*
    - Never use a metaphor, simile, or other figure of speech which you are used to seeing in print.
    - Never use a long word where a short one will do.
    - If it is possible to cut a word out, always cut it out.
    - Never use the passive where you can use the active.
    - Never use a foreign phrase, a scientific word, or a jargon word if you can think of an everyday English equivalent.
    - Break any of these rules sooner than say anything outright barbarous.
- Define technical terms, jargon and notation clearly
  - Before using!
  - write out domain-specific abbreviations first time used, e.g. "The Domain Name System (DNS) becomes a bottleneck."
- Proofread! Spell-check!

Writing Advice – *Graphics*

- Use figures to help clarity
  - data interpretation
  - architectures
  - interfaces
- Do not overuse figures
  - What does reader gain by seeing this figure?
- Figure sizes
  - Large enough to easily read
  - Don’t pad paper with unnecessarily large figures

Writing Advice - *Form*

- 12pt Times-Roman font
- 1-inch margins
- double-spaced
- Latex template files posted

Writing advice - *Procedure*

- Start with extended outline
- Don't try to write it all at once
- Write something, even a few lines, every day
- Use headers and sub-headers
  - helps illuminate logical flow of paper
- "Don't fall in love with your prose. Writing and rewriting is what every author does to create papers that are both convincing and clear"

[Dr. Rob Fish]

Look at some examples

- (for IW papers) https://www.cs.princeton.edu/ugrad/independent-work/guidelines-and-useful-information#Example_Single-Semester_Projects_from_Previous_Years

- (for senior theses) https://dataspace.princeton.edu/handle/88435/dsp01mp48sc83w/simple-search?query=2020

Summary

- Follow outline but don’t be shackled by it
- Don’t lose the big picture for the details
- Will this be clear to others?
- Allow time to develop the paper day by day
- See posted examples

Acknowledgments

- Thanks to IW coordinators (past and present) for slides and advice:
  - David Dobkin
  - Kyle Jamieson
  - Robert Fish

- Thanks to IW administrator and staff:
  - Mikki Hornstein
  - Kobi Kaplan

Questions?