DSC Capstone Sequence

Lecture 04
Scientific Writing

Announcements / Explanations

- The standards for effective (written) communication vary by Domain.
- Your Domain should have a paper to use as an exemplar.
  - If you are replicating a paper, it's that one!
  - If you are unclear about the paper to emulate, ask your mentor:
    - "When writing our Q1 survey paper and the Q2 project report, we are supposed to follow the style/example of a paper in our area. What paper should we use?"
  - Some of the details in this lecture may not match the standard in your domain. The concepts discuss, however, *do apply*.
- **As you go through this lecture, map the content to your paper.**
- You will be writing the introduction for your survey paper soon. You will be graded on the principles discussed in this lecture (along w/correctness)

Lecture Outline

- The purpose of scientific writing
- The anatomy of a scientific paper

Scientific Writing Overview

- Scientific papers typically have a set structure:
  - “Context - Content - Conclusion”
- Set structure has a number of advantages:
  - Helps different audiences with different goals to get information.
  - Clarifies the process involved in the project
- (Multiple) sections correspond to project checkpoints.

Paper Sections
Abstract
Introduction
Methods
Results
Discussion
References
Appendix

Scientific Writing Overview

- Figures (tables and graphs) should be self-contained and complete:
  - **The reader should understand your paper *only* by glancing at the figures!**
  - Every figure should be titled and have understandable labels.
- Your paper is a narrative essay that answers the questions laid out in assignment checkpoint prompts and these slides.
  - Do **not** write you paper as short-answer responses to given questions!
- Cite others work (and do so a lot!)
  - Learning from and using others research is something to be proud of!
  - However, only discuss and cite work that's *relevant* to your project.
- Only include statements if they are correct.

Scientific Writing Style

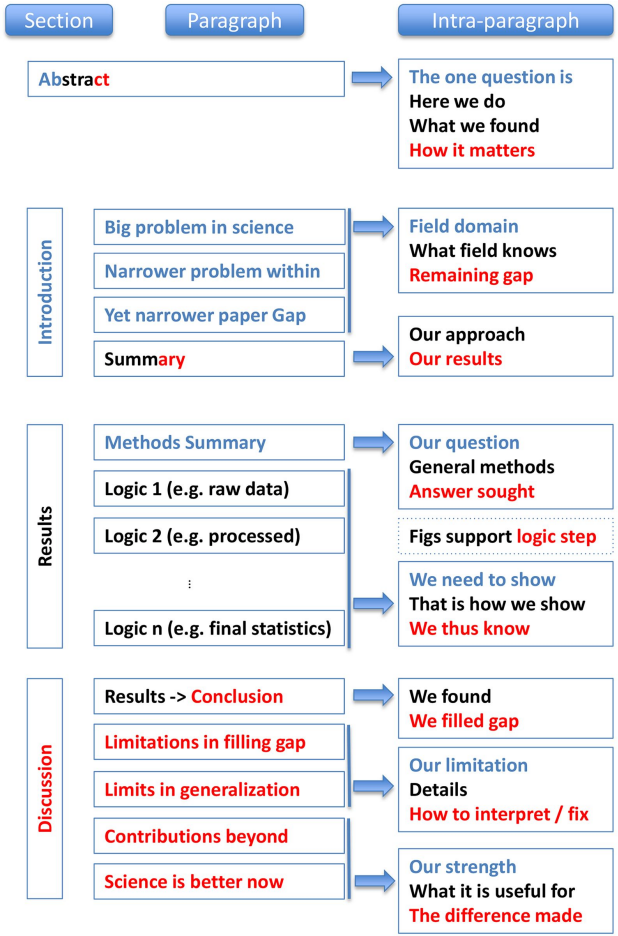
- Your paper is a narrative essay that describes a problem, methods to solve the problem, and the results of the approach.
    - It should not be a chronology of your project experience. Every project produces work irrelevant to the results; they should be omitted.
- Use active voice with first person plural (“we”) to refer to both yourselves, the author(s), as well as the reader.
- Use plain-spoken language without jargon; always define acronyms and specialized terms.
- Avoid flowery adjectives and adverbs (e.g. “very”, “large”).
- Always be as specific as possible. Vague statements contribute nothing.

Anatomy of a Scientific Paper

Paper Sections	Purpose of Section
Abstract	*What did I do?*
Introduction	*What is the problem?*
Methods	*How did I approach it?*
Results	*What did I find out?*
Discussion	*What did it mean?*
References	*Who did I cite?*
Appendix	*Details!*

Anatomy of a Scientific Paper

- This figure describes how each section of a paper breaks up into paragraphs, alongside the general content of each paragraph.
- The content varies slightly from the standard we'll use, but it's still a useful reference!



Abstract and Title

- Both should describe the contents of the paper:
  - Abstract: in one paragraph (~1 sentence/section)
  - TItle: in one phrase
- Write minimal content in plain-spoken language.
- Both should be immediately understandable.
  - People browse abstracts to decide if a paper is worth reading.
- Do this portion *last!*
  - Useful for developing your report, but will be rewritten at the end

The Anatomy of an Abstract

- Contextualize the problem.
- Review gaps in prior approaches.
- Clearly state your project's contribution to the problem.
- Specify relevant details to your methods

Can you take apart each of these components in the abstract at right?

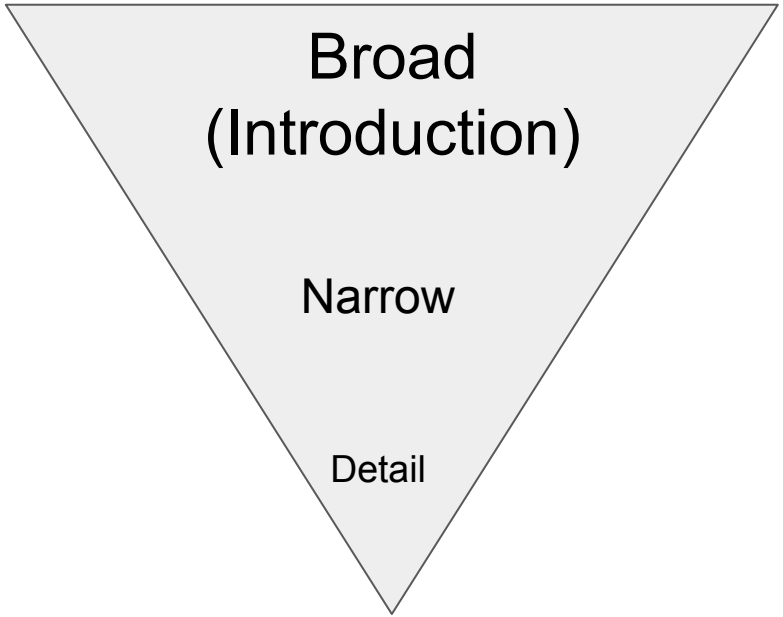
Robust Learning from Discriminative Feature Feedback

Authors: Sanjoy Dasgupta, Sivan Sabato

Abstract: Recent work introduced the model of learning from discriminative feature feedback, in which a human annotator not only provides labels of instances, but also identifies discriminative features that highlight important differences between pairs of instances. It was shown that such feedback can be conducive to learning, and makes it possible to efficiently learn some concept classes that would otherwise be intractable. However, these results all relied upon perfect annotator feedback. In this paper, we introduce a more realistic, robust version of the framework, in which the annotator is allowed to make mistakes. We show how such errors can be handled algorithmically, in both an adversarial and a stochastic setting. In particular, we derive regret bounds in both settings that, as in the case of a perfect annotator, are independent of the number of features. We show that this result cannot be obtained by a naive reduction from the robust setting to the non-robust setting

The Anatomy of an Introduction

- The introduction sets the context of the paper with an introductory paragraph:
  - Setting up the problem (context)
  - Problem statement
  - Description of your approach (and rationale for why it's reasonable).
- The introduction contains subsections that describe the problem's context.



Introduction: Supplying Context

The introduction contains subsections that supply context for the approach

- Section 1: The *literature review and prior work* describes how the problem has been approached in the past.
    - Reviewing past work sets up the context in which your problem is being approached.
- Section 2: The *data description* connects how the data in the project is used to approach the problem at hand.
    - This section qualitatively addresses the relationship between the population of study, data generating process, and the observed data. (i.e. is the data appropriate for the problem?)
    - This section places your work in the context of previous approaches.
    - (Checkpoint 1 -- report)

Methods

- Often goes by different monikers (e.g. “experimental design”)
- Describes the logic that lead to the results
  - Contains detail, but not so much to cloud the logic.
  - Consider leaving distracting details to the appendix.
- If your project involves developing multiple models (that qualitatively differ in method), each should have its own section. Examples include:
  - Preprocessing models on different data sources
  - Models used to label data vs model to classify an objective
  - Two approaches to a problem that result in a more robust conclusion.

Results

The results section:

- Presents data clearly without interpretation.
- Consists of self contained figures and a narrative explanation of the *content* of those figures.
- The row/columns of tables (e.g. of standard errors or classifier-performance results) should be described factually.
- The relevant/significant results should be called out in writing; the entries of the respective tables should be highlighted.

Discussion

The Discussion section:

- Interprets the results of the project
- Describes how the results compare with prior work on the problem
- Discusses the impact and applicability of the results
- Discusses limitations of the approach and why the results are stronger (or more broadly applicable).
- Possibilities for future work

Exercise 1: communicate through specificity

In an introduction, you will give context for *why* a problem is interesting. You will also *clearly* state the problem and scope in this context.

In (one of) the papers in your Domain,

- Specify the problem in concrete, testable terms and hypotheses
- Discuss your model's internal consistency (correctness) *and* broader applicability.

Exercise 2: start at the end!

Starting with the problem and testable hypothesis from exercise 1:

- Identify the most important figure/result in the results section, given the introduction. Why does this result summarize why the approach to the problem in the paper is impactful?