

Scientific writing

Lecture 3

23.10.26

The goal

Research paper: new result.

Problems with writing, preparation etc. can be fixed or ignored.

Thesis: proving competence.

It is the process that matters. You should do and document necessary steps properly.

Thesis vs research paper

The goal

Research paper: new result.

Problems with writing, preparation etc. can be fixed or ignored.

Thesis: proving competence.

It is the process that matters. You should do and document necessary steps properly.

Popular science.

Give some information; keeping attention is difficult. Not necessarily precise.

Pure theory research paper

Begin writing, when you have something to say.

Pure theory research paper

Begin writing, when you have something to say.

- Find a topic
- Review literature
- formulate problem
- prove
- write

Pure theory research paper

Begin writing, when you have something to say.

- Find a topic
- Review literature
- formulate problem
- prove
- write

Meticulous

- Lots of formalization and details.
- may be written at one go

Pure theory research paper

Begin writing, when you have something to say.

- Find a topic
- Review literature
- formulate problem
- prove
- write

Meticulous

- Lots of formalization and details.
- may be written at one go

A rough estimate: 1 page a day.

Experimental thesis work outline

Begin writing, when you have something to say.

Not entirely true for experimental work

Experimental thesis work outline

Begin writing, when you have something to say.

Not entirely true for experimental work

- Formulate topic
- Formulate question, exact hypothesis
- hypotheses should be falsifiable
- Review literature
write a sketch, take notes, collect the papers!
- Plan experiments: what constitute a convincing answer? $\times$ “data was collected”
- Gather evidence
 - proofs
 - test on real data
 - test on model
- measures
- Construct argument
- Finalize Writing

Experimental thesis work outline

Begin writing, when you have something to say.

Not entirely true for experimental work

- Formulate topic
- Formulate question, exact hypothesis
- hypotheses should be falsifiable
- Review literature
write a sketch, take notes, collect the papers!
- Plan experiments: what constitute a convincing answer? × “data was collected”
- Gather evidence
 - proofs
 - test on real data
 - test on model
- measures
- Construct argument
- Finalize Writing

It is good to write something at each stage

Devising paper plan

What do you want to say

- The main thing (the message)
- what is the argument for that (theoretical, experimental, what types of experiments)?
- counterarguments, alternatives explanations — include answers to this
except for pure math, no answer is perfect
- what are the definitions/(experimental) setting needed to understand the answers?

There will be hard decisions: linearising may be tricky.

Recall Writing scientific papers is (also) telling a story.

Experimental novel

- The Gates of Paradise (*Bramy raju*)
- A Clockwork Orange
- Hopscotch (*Rayuela*)
- Conversation in The Cathedral (*Conversación en La Catedral*)
- Extinction (*Auslöschung*)
- A Song of Ice and Fire
- (spoiler)
- ...

Recall Writing scientific papers is (also) telling a story.

× Experimental novel

- The Gates of Paradise (*Bramy raju*)
- A Clockwork Orange
- Hopscotch (*Rayuela*)
- Conversation in The Cathedral (*Conversación en La Catedral*)
- Extinction (*Auslöschung*)
- A Song of Ice and Fire
- (spoiler)
- ...

Comic relief

Recall Writing scientific papers is (also) telling a story.

× Experimental novel

- The Gates of Paradise (*Bramy raju*)
- A Clockwork Orange
- Hopscotch (*Rayuela*)
- Conversation in The Cathedral (*Conversación en La Catedral*)
- Extinction (*Auslöschung*)
- A Song of Ice and Fire
- (spoiler)
- ...

Think of classic crime/mystery genre:

- choose the murder and setting (answer)
- develop backwards, try to keep attention
- make it logical

Make the plan clear

Structure in literature/movies

- Lalka
- The wire

Make the plan clear

Structure in literature/movies

- Lalka
- The wire

Bottomline: structure can be overlooked.

- announce the plan (introduction?)
- Recapitulate what was done
- explain how you will proceed
- connect past and future

Alternative approach: extended abstract way

- Conference publishing format: 6–12 pages + appendices.
 - The main message has to fit.
 - Appendices: proofs, data, ...

Alternative approach: extended abstract way

- Conference publishing format: 6–12 pages + appendices.
 - The main message has to fit.
 - Appendices: proofs, data, ...
- Advantage: aimed at easy reading.
- Advocates of similar writing in mathematics.

Alternative approach: extended abstract way

- Conference publishing format: 6–12 pages + appendices.
 - The main message has to fit.
 - Appendices: proofs, data, ...
- Advantage: aimed at easy reading.
- Advocates of similar writing in mathematics.
 - Introduction
 - definitions/setting etc.
 - sketch of main argument
 - all crucial notation, definitions, etc.
 - correct main statements
 - sketches of main arguments
 - main graphs and results
 - crucial info of experimental setting
 - main conclusions (if any)
 - the gory details

Common ways of telling a story

Main goal: logical and informative

- chain
- outline + several sections on details: when there are several components
- by example: first the approach is explained in (some) detail on an example, then it is explained in full
- simple to complex: first much simplified case, then generalize
usual RadixSort presentation is a good example

Some thoughts

Aim

- ✗ list the work undertaken
- ✓ evidence explaining an argument

Some thoughts

Aim

- ✗ list the work undertaken
- ✓ evidence explaining an argument

Pyramid

Summarize the argument and explain in iterations.

Like in a newspaper:

- title gives the gist
- first sentence gives some detail
- first paragraph gives important information

Some thoughts

Aim

- ✗ list the work undertaken
- ✓ evidence explaining an argument

Pyramid

Summarize the argument and explain in iterations.

Like in a newspaper:

- title gives the gist
- first sentence gives some detail
- first paragraph gives important information

Introduction-Related Work-Methods-Experiments-Analysis

Traditional way.

May be poor in more complex setting.

Predicting music from EEG signal

Research topic

Predicting music from EEG signal.

Predicting music from EEG signal

Research topic

Predicting music from EEG signal.

Literature review

something was done before (with mild results)

Ablation was poorly done

There are data sets; we will not make new; datasets are small, not unified

Predicting music from EEG signal

Research topic

Predicting music from EEG signal.

Literature review

something was done before (with mild results)

Ablation was poorly done

There are data sets; we will not make new; datasets are small, not unified

Research question

can we recreate

evaluate more (simple) models

can we make better ablation

Predicting music from EEG signal

Research topic

Predicting music from EEG signal.

Literature review

something was done before (with mild results)

Ablation was poorly done

There are data sets; we will not make new; datasets are small, not unified

Research question

can we recreate

evaluate more (simple) models

can we make better ablation

Context

what is EEG signal, how is it recorded

- it measures everything, lots of noise
- there is lots of radio noise anyway, also from equipment

1 intro

① intro

② EEG signal

① history

② characteristics

(non-stationary, artifacts, bands, noise, ...)

③ data collection (equipment and methodology)

④ processing

- ① intro
- ② EEG signal
 - ① history
 - ② characteristics
(non-stationary, artifacts, bands, noise, ...)
 - ③ data collection (equipment and methodology)
 - ④ processing
- ③ datasets and previous work
 - ① music processing
 - ② dataset + task + result (3 times)

- ① intro
- ② EEG signal
 - ① history
 - ② characteristics
(non-stationary, artifacts, bands, noise, ...)
 - ③ data collection (equipment and methodology)
 - ④ processing
- ③ datasets and previous work
 - ① music processing
 - ② dataset + task + result (3 times)
- ④ Experiments:
 - ① models used
 - ② results on tasks

- ➊ intro
- ➋ EEG signal
 - ➊ history
 - ➋ characteristics
(non-stationary, artifacts, bands, noise, ...)
 - ➌ data collection (equipment and methodology)
 - ➍ processing
- ➌ datasets and previous work
 - ➊ music processing
 - ➋ dataset + task + result (3 times)
- ➍ Experiments:
 - ➊ models used
 - ➋ results on tasks
- ➎ Ablation (separately for each dataset)

- ① intro
- ② EEG signal
 - ① history
 - ② characteristics
(non-stationary, artifacts, bands, noise, ...)
 - ③ data collection (equipment and methodology)
 - ④ processing
- ③ datasets and previous work
 - ① music processing
 - ② dataset + task + result (3 times)
- ④ Experiments:
 - ① models used
 - ② results on tasks
- ⑤ Ablation (separately for each dataset)
- ⑥ Evaluation (how good simpler explanations are)

- ① intro
- ② EEG signal
 - ① history
 - ② characteristics
(non-stationary, artifacts, bands, noise, ...)
 - ③ data collection (equipment and methodology)
 - ④ processing
- ③ datasets and previous work
 - ① music processing
 - ② dataset + task + result (3 times)
- ④ Experiments:
 - ① models used
 - ② results on tasks
- ⑤ Ablation (separately for each dataset)
- ⑥ Evaluation (how good simpler explanations are)
- ⑦ Conclusions

- 1 intro
- 2 EEG signal
 - 1 history
 - 2 characteristics
(non-stationary, artifacts, bands, noise, ...)
 - 3 data collection (equipment and methodology)
 - 4 processing
- 3 datasets and previous work
 - 1 music processing
 - 2 dataset + task + result (3 times)
- 4 Experiments:
 - 1 models used
 - 2 results on tasks
- 5 Ablation (separately for each dataset)
- 6 Evaluation (how good simpler explanations are)
- 7 Conclusions

- 1 intro
- 2 EEG signal (no processing of EEG signal)
- 3 Methods
 - 1 music processing
 - 2 EEG processing
 - 3 Models
- 4 datasets
- 5 previous results ("Related Work")
- 6 Experiments
- 7 Ablation
- 8 Conclusions

Abstract

Is the paper of relevance to the reader?

- 1 general statement introducing the area
- 2 the specific problem to be solved
- 3 review of existing or standard solutions
- 4 an outline of the proposed new solution.
- 5 a summary of how the solution was evaluated and the results

Introduction

- Long abstract
- Important statement near beginning.
- detailed discussion of the results: not advised
 - ✗ “the analysis, as we show, led to surprising insights”
 - ✗ “as discussed later, this decision had unanticipated benefits”

Traditional content/Organization

Background (Literature review)

- general context, previous results: intro
- detailed discussion, comparison: later (not necessarily a chapter)

Methods–Results–Discussion

This has to appear.
But not necessarily in this order.
Section names should be meaningful.

Conclusions

Summary.
Often not needed in theoretical work.

Drafting

Strict mathematical content should be correct and precise from the beginning.
Hypotheses and problem as well.

Introduction, descriptions, methods, etc. can be drafted and reworked later.
It is difficult to write well in the first attempt.

In-memory sorting algorithms require random access to records. For large files stored on disk, random access is impractically slow. These files must be sorted in blocks. Each block is loaded into memory and sorted in turn. Sorted blocks are written to temporary files. These temporary files are then merged. There may be many files but in practice the merge can be completed in one pass. Thus each record is read twice and written twice. Temporary space is required for a complete copy of the original file.

Imitate: look at well written papers.

Getting it wrong: incomprehensible

All quotes from Zobel.

In thesis you have a captive audience
Some things make people stop reading.

Getting it wrong: incomprehensible

All quotes from Zobel.

In thesis you have a captive audience
Some things make people stop reading.

Internet supports all type of the forms of information that are digital, such as the pages of the Web everywhere and also libraries and email, so it is a language of all our information sources in a world repository that is our knowledge.

With the exploding of documents on the internet, systems of finding documents that are the answers of users their queries have become important in the recent years.

We have new networks of wireless like wired networks that our method makes use of in computers connected to each other but heterogeneously and distributed.

An ideal vector space is the base of IR research, so the basic problem of IR is to set up a suitable vector space, in this suitable vector space, query and document can be represented well by vectors.

Getting it wrong: Ignorance

Ignorance is not a simple lack of knowledge but an active aversion to knowledge, the refusal to know [...] Karl Popper

Explaining trivias

Misunderstanding literature

- errors in citing results
- lack of easy to find papers
- omitting developments of many years
- anachronism: applying old results to current setting

General misunderstanding of the world around

“with the growth of data online, users have given up trying to use Web search engines”

“the development of large RAM means that there is no longer a need for offline storage”

“increases in the efficiency of processors are encouraging people to develop their own database utilities, based on open-source toolkits”

“recent advances in machine learning mean that soon it may no longer be necessary to write software”

“As the scale of data on the Web grows to billions of pages, searchers can no longer find answers to queries.”

Paper looks bad

- formatting/layout
- figures
- misprints
- broken lines

Suggests that you do not care.

Getting it wrong: The Upper Hand (swaggering)

- analysis of this method is of course a straightforward application of tensor calculus
- we leave the details of this proof for the mentally infirmed
- the argument proceeds on Voltarian principles

Getting it wrong: The Upper Hand (swaggering)

- analysis of this method is of course a straightforward application of tensor calculus
- we leave the details of this proof for the mentally infirmed
- the argument proceeds on Voltarian principles

Do not force the reader to look up words in the dictionary.

Getting it wrong: Obfuscation

- ✗ Experiments, with the improved version of the algorithm as we have described, are the step that confirms our speculation that performance would improve. The previous version of the algorithm is rather slow on our test data and improvements lead to better performance.
- ✗ The status of the system is such that a number of components are now able to be operated.
 - ✓ Several of the system's components are working.
- ✗ In respect to the relative costs, the features of memory mean that with regard to systems today disk has greater associated expense for the elapsed time requirements of tasks involving access to stored data.
 - ✓ Memory can be accessed more quickly than disk.

Getting it wrong: Straw Men

Exaggerated hypothesis stated only in intent to criticize it/base your argument on.

“it can be argued that databases do not require indexes”

“Query languages have changed over the years. For the first database systems there were no query languages and records were retrieved with programs. Before then data was kept in filing cabinets and indexes were printed on paper. Records were retrieved by getting them from the cabinets and queries were verbal, which led to many mistakes being made. Such mistakes are impossible with new query languages like . . .”

“Since the invention of the internet, researchers have been using the Web to publish data.”

“Researchers working on computer image generation initially failed to consider the benefits of parallel processing.”

Getting it wrong: Rhetoric

Convince by arguments, evidence, data.

Calculemus!

Gottfried Wilhelm Leibniz

- ✗ We did not investigate partial interpretation because it is known to be ineffective.
- ✗ Execution was almost instantaneous.
- ✗ The Web is infinitely large.
- ✗ There is no limit to the possible efficiency gains.
- ✗ Most users prefer the graphical style of interface.
- ✓ We believe that most users prefer the graphical style of interface.
- ✗ Another possibility would be a disk-based method, but this approach is unlikely to be successful.
- ✓ Another possibility would be a disk-based method, but our experience suggests that this approach is unlikely to be successful.

- ✗ Writing a program is like building a model with connector blocks.
- ✗ One-sided protocols are like signals in football.

Contrasting look-ahead graph traversal with standard approaches, look-ahead uses a bird's-eye view of the local neighbourhood to avoid dead ends, but at significant cost: it is necessary to feed the bird and wait for it to return after each observation.

- ✗ Writing a program is like building a model with connector blocks.
- ✗ One-sided protocols are like signals in football.

Contrasting look-ahead graph traversal with standard approaches, look-ahead uses a bird's-eye view of the local neighbourhood to avoid dead ends, but at significant cost: it is necessary to feed the bird and wait for it to return after each observation.

Biblical analogies.

Quotations are very common in English-speaking world.
Especially in books, to less extent in theses.

I do not see why.

Intersection of analogy, humor and argument by authority.

Humor in conversation

- witty
- comic relief
- brings people together
- removes the awkward silence

Humor in conversation

- witty
- comic relief
- brings people together
- removes the awkward silence
- ego trip
- can be misunderstood
- requires context

Humor in conversation

- witty
- comic relief
- brings people together
- removes the awkward silence
- ego trip
- can be misunderstood
- requires context

Can also make serious statements difficult.
(Think of *Patrick Melrose*)

Humor in conversation

- witty
- comic relief
- brings people together
- removes the awkward silence
- ego trip
- can be misunderstood
- requires context

Can also make serious statements difficult.
(Think of *Patrick Melrose*)

What remains in writing?

- ages bad
- the object of humor cannot defend

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

French textbook in logic, with nude (women's) photos (somewhat connected to chapter titles).

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

French textbook in logic, with nude (women's) photos (somewhat connected to chapter titles).

$\emptyset$ is absorbing on the left and neutral on the right, like British toilet paper.

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

French textbook in logic, with nude (women's) photos (somewhat connected to chapter titles).

$\emptyset$ is absorbing on the left and neutral on the right, like British toilet paper.

The Toilet Paper Problem.

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

French textbook in logic, with nude (women's) photos (somewhat connected to chapter titles).

∅ is absorbing on the left and neutral on the right, like British toilet paper.

The Toilet Paper Problem.

Chapter names: An absorbing barrier; A process of elimination; Residues

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

French textbook in logic, with nude (women's) photos (somewhat connected to chapter titles).

∅ is absorbing on the left and neutral on the right, like British toilet paper.

The Toilet Paper Problem.

Chapter names: An absorbing barrier; A process of elimination; Residues

I want to thank the architect of the computer science building at Stanford University for implicitly suggesting this problem

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

French textbook in logic, with nude (women's) photos (somewhat connected to chapter titles).

∅ is absorbing on the left and neutral on the right, like British toilet paper.

The Toilet Paper Problem.

Chapter names: An absorbing barrier; A process of elimination; Residues

I want to thank the architect of the computer science building at Stanford University for implicitly suggesting this problem

Growing balls larger than (Name)

Humor examples

A note saying roughly

“I did prove a more general case, but the submission format was too short for me to write it down”.

The author received emails asking about this proof.

French textbook in logic, with nude (women’s) photos (somewhat connected to chapter titles).

∅ is absorbing on the left and neutral on the right, like British toilet paper.

The Toilet Paper Problem.

Chapter names: An absorbing barrier; A process of elimination; Residues

I want to thank the architect of the computer science building at Stanford University for implicitly suggesting this problem

Growing balls larger than (Name)

Footnotes with comments on songs, literature, assorted jokes.

Humor and alike

Data augmentation





Beware

Popular science articles are expected to be funny.

You need to keep Readers' attention.

Witty titles

“Traffic engineering is like an avocado”