How to clock a research study and article in one go

Written By: Terence van Zyl
Website: https://tvanzyl.github.io/

What is the thinking

Contrary to common conceptions, research is not activities such as browsing the internet, reading books or textbooks, reviewing others' research, or summarizing existing studies (commonly referred to as a literature survey). The primary objective of research is a methodical and innovative endeavour aimed at expanding or validating humanity's knowledge. This undertaking involves the following key activities:

1. **Selecting a research topic:** *This marks the initial step where a researcher identifies an area of interest for exploration.*
2. **Finding an open research problem within the topic:** *The researcher identifies a gap or problem within the chosen topic that warrants investigation.*
3. **Proposing a hypothetical solution to the research problem:** *A researcher formulates a preliminary solution or hypothesis to address the identified problem.*
4. **Coming up with research questions/objectives:** *Research questions and objectives are established to provide a framework for investigating and potentially falsifying the hypothesis.*
5. **Formulating the experimental design:** *The researcher plans the methodology and experimental design to test the hypothesis effectively.*
6. **Implementing the experimental design:** *The planned experiments are carried out, generating data for analysis.*
7. **Analysing the results:** *The collected data is analysed to draw conclusions and insights relevant to the research questions.*
8. **Writing up the study:** *The final step involves documenting the entire research process in a comprehensive paper.*

Essential to this intricate process is effective management, and incorporating writing seamlessly can prove to be a successful strategy. Embedding discovery writing into the research process provides a dynamic and adaptable approach, enabling the exploration of ideas, surmounting obstacles, pinpointing research gaps, and sustaining motivation. These benefits in more detail follow:

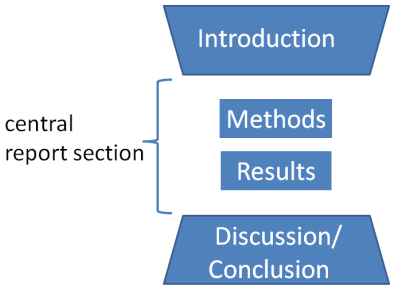
Generating Ideas: *Discovery writing in academia allows for the exploration of thoughts and initial ideas without the constraints of a predetermined structure.*

Overcoming Writer's Block: *Initiating the writing process without detailed planning can help overcome writer's block, allowing for a more natural flow of ideas.*

Exploring Research Gaps: *Writing actively may reveal knowledge gaps, guiding subsequent research efforts and formulating new and refined research questions.*

Motivating the Writing Process: *Engaging with the material through active writing can motivate, maintain enthusiasm for the project and foster a sense of ownership over evolving ideas.*

Structure of a research paper



Research articles in machine learning follow the IMRaD organisational structure. The Introduction usually consists of between one and three sections, depending on the page limits set by the conference or journal. These are followed by a section describing the experimental setup and results. The final sections usually include a discussion and conclusion. Some of these headings may have slight name variations, some are collapsed into a single section with subsections, and some are left out depending on the type of research article. For instance, a more complete structure would follow something like this:

```
%---Introduction---
\section{Introduction}
\section{Related Work}
\section{Name Of Contribution}

%-----Methods-----
\section{Experimental Design}
   \subsection{Datasets}
   \subsection{Experimental Setup}
   \subsection{Analaysis}
   \subsection{Tools and Libraries}

%-----Results-----
\section{Experimental Results}

%---Discussion----
\section{Discussion}
\section{Conclusion}
```

Alternatively later on if you are limited for space, don't have a proposed new algorithm or model or have less to write you might collapse into the following headings:

```
Introduction
Experimental Setup
Experimental Results
Discussion and Conclusion
```

Further Reading

Big picture step-by-step guide

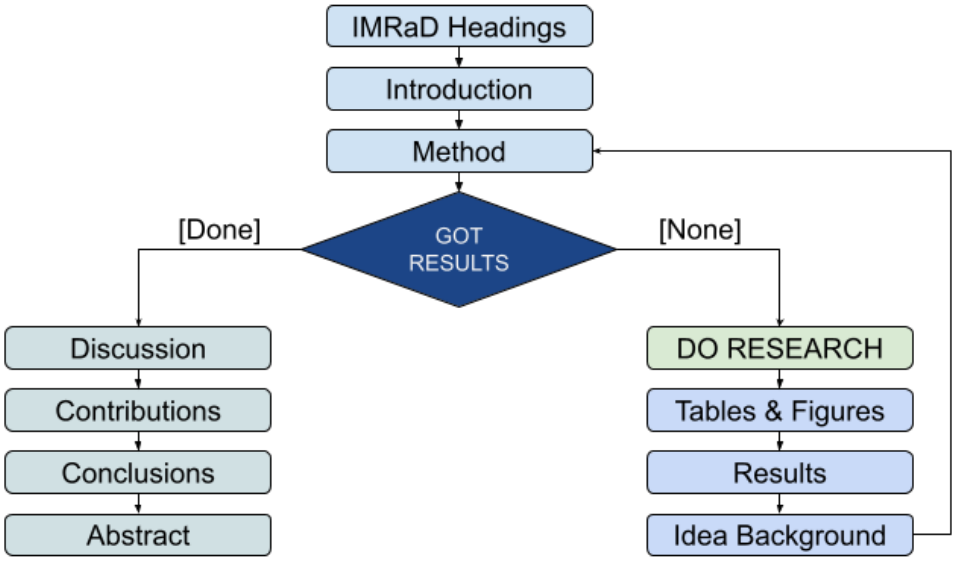
1. Use a latex template in Overleaf (IEEE or LNCS conference).
    - Use BibLaTeX for managing your references in a .bib file.
2. Put down the IMRaD section headings.
3. Write the initial introduction to Create a Research Space (CARS).
    - **Here is an interesting problem.** *(The Territory)* `- Paragraphs 2`
    - **It is unsolved!** *(The Niche)* `- Paragraphs 1`
    - **Here is an idea on how to solve it.** *(Occupy The Niche)* `- Paragraphs 1-2`
4. Write up the intended experimental design/research method
    - **How to falsify the idea** `- Paragraphs 5-6`

> **Do the research**: Analyse the results as you are going.

> **When you are done:** You should know the study's 1-3 main contributions.

5. Add a single figure that expresses the main idea of the study.
    - This figure is the **bootstrap** for an exceptional paper.
6. Fill in the rest of the tables and figures for the results.
    - Each contribution should be backed by a single table or figure.
    - These are the **kernels** of the paper, all linked hierarchically back to the **bootstrap**
7. Complete results `(- One/two pages)` with description of the results.
8. Write up your idea. `- One/two pages`
9. Write up the related work. `- One/two pages`
10. Refresh experimental design with details (e.g. hyperparameters used).
11. Write up the discussion. `- Half a page`
12. Expand the Introduction `(- One page)` and include contributions.
13. Add conclusion. `- Half a page`
14. Add abstract. `- 200/250 words`
15. Finalise the title. `- Ten words or less`

If you change your mind, go back to the start.



Small picture step-by-step guide (IMRaD)

1. Introduction (the moves)

1.1 Establishing the territory `Here is an interesting problem.`

Describe the current state of research on the topic. From phrase bank start with two paragraphs with a sentence from each of:

1. Establishing the importance of the topic [...]
2. Referring to previous work to establish what is already known

Eventually these paragraphs should have references to all the key studies that lead to identifying your niche.

1.2 Identifying a niche - `It is unsolved!`

Provides a problem statement for the study that can be addressed in a paragraph. From phrase bank select one or two of these as sentences to start:

1. Identifying a knowledge gap in the field of study
2. Claim something is wrong with the previous research
3. Explaining the inadequacies of previous studies
4. Identifying the scarcity or lack of previous research

1.3 Occupying the niche - `Here is an idea on how to solve it.`

Describes how the research contributes to solving the problem statement. From phrase bank start by selecting one sentence from each of:

1. Stating the focus, aim, or argument of a short paper
2. Stating the purpose of the current research
3. Explaining the significance of the current study

1.3 Stake your claim - `Here is how I solve it.`

After you have completed your study revisit the introduction, and add paragraph:

1. Describing the research design and the methods used
2. Share the contributions

~~

2. Related Work (Optional)

If you have the space you would expand on the related work that was briefly introduced in the introduction to create a research space. This section is typically more fully developed in journal articles than in conference articles and would capture the full literature survey.

~~

3. Idea Background (Optional)

`The idea.`

Suppose you have devised a novel idea and compared it to several other baselines and state-of-the-art methods. Then you would describe the idea and how it links to the other methods here.

1. Describe the idea, algorithm, mathematics, etc.
2. Describe other background needed to understand the idea

~~

4. Method (the moves)

`How to falsifying the idea.`

From phrase bank contextualizing why the study method is used:

1. Restate research purpose, hypothesis or gaps in research
2. Provide relevant theoretical or empirical information to show connections to the field
3. Substantiate your methodology in relation to previous studies with referencing

e.g.

> (1) To test the primary hypothesis that A outperforms B, we conducted an empirical evaluation of the two models. (2) Traditionally, forecast accuracy has been assessed by measuring the MSE. (3) To this end, our experiment follows the procedure used by Joe Soap et al. (2001).

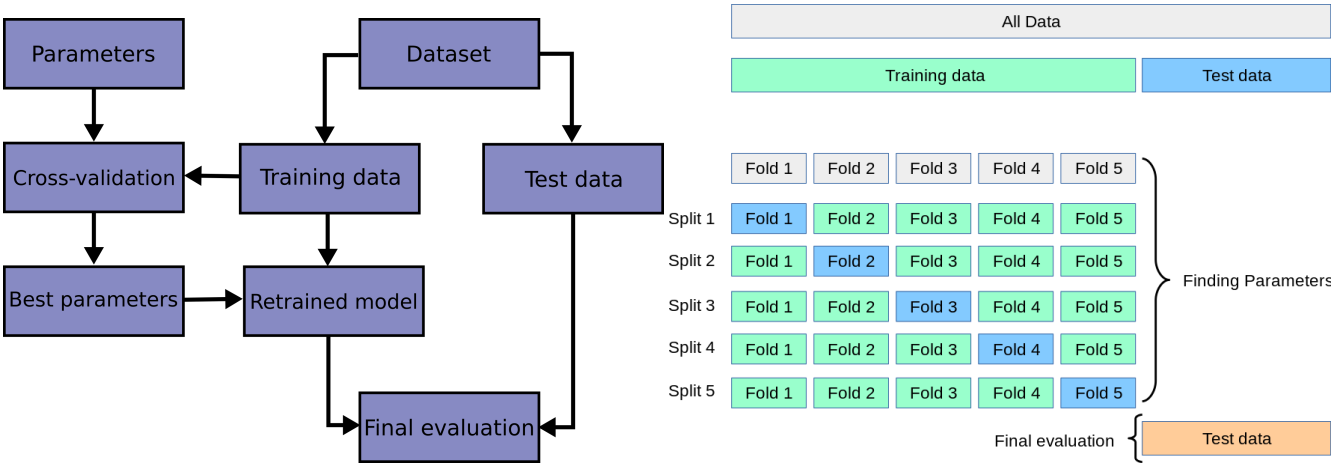
4.1 Datasets

1. Describe the dataset(s) used
2. Describe any preporocessing on the dataset(s) used

4.2 Experimental Setup

See model Selection and Evaluation.

1. Describe the step-by-step study procedures in a paragraph or two.



Figures taken from SciKit Learn.

1. Briefly reiterate models to be compared to yours:
    1. baseline models
    2. state-of-the-art models
2. Should cover and substantiate choices for:
    1. how many experiment repeats (e.g. 10),
    2. train test split (e.g. 80-20),
    3. K-fold cross validation (e.g. 5-fold),
    4. model initialisation,
    5. hyperparameter selection (e.g. grid-search), and
    6. parameter optimisation (e.g. Adam).

4.3 Analysis

1. Describe the analysis and metrics used for selection and evaluation
2. Describe the final evaluation procedure and significance test on results (p-values)

4.4 Tools and Libraries

1. Describe tools and libraries used to do the study

e.g.

> All experiments were run on an Intel 80286 running MS-Dos 6.22. The algorithms were implemented using Microsoft BASIC. All codes, datasets and the results from these experiments can be found on [GitHub].

~~

5. Results (the moves)

`What did you find?`

The result's section identifies the tables or figures and indicates their content, and highlights, and point out and describe the relevant or significant outcomes.

For each result (figure or table), from phrase bank, take a sentence from each of:

1. Referring back to the research aims or procedures
2. Referring to data in a table or chart
3. Highlighting significant data in a table or chart

Then possibly a sentence or two for:

1. Stating a positive or negative result
2. Highlighting exciting or surprising results

~~

6. Discussion (the moves)

`What does it mean?`

1. First sentence - What is 'the answer' to the problem
2. Second sentence and on - justify your answer.

Then **with reference to previous literature (Literature Survey)** from phrase bank for each significant result, do one or more of:

1. Providing background information
    1. reference to the question
    2. reference to the literature
2. Restating the result or one of several results
3. Indicating an unexpected outcome
4. Comparing the result
    1. supporting previous findings
    2. contradicting previous findings
5. Offering an explanation for the findings

Finally:

1. Advising cautious interpretation of findings
2. Suggesting general hypothesis
3. Noting implications of the findings

~~

7. Conclusion (the moves)

`Bring together the main ideas and speculate on the implications.`

Give a final judgement on the study and comment on suggestions for improvement and speculate on future.

From phrase bank:

1. Restating the aims of the study

2. Summarising main research findings
3. Explaining the significance of the findings or contribution of the study
4. Recognising the limitations of the current study
5. Making recommendations for further research work

~~

8. Abstract (the moves)

`Summary of the entire study.`

1. **Locate** the paper to the larger field to give perspective,
2. **Focus** on the questions/issues/problems to be explored/examined,
3. **Anchor** the argument by outlining research, samples and analysis,
4. **Report** on major findings relevant to the argument, and
5. **Argue** out the argument and close with this article's perspective.

e.g.

> [LOCATE] Adaptive bandwidth kernel density estimators (AB-KDEs) have received attention from the academic community due to an analytical promise of increased performance over classical estimators. [FOCUS] However, the field is fragmented and there exists no comprehensive comparison of the existing state-of-the-art AB-KDEs. [ANCHOR] We provide a comparison of some state-of-the-art and classical AB-KDE methods as well a computational framework along with a novel implementation of a full principal axes rotation hyper-ellipsoid variant of the k-Nearest Neighbours algorithm. [REPORT] The extensive experimental results show that the fixed bandwidth rule-of-thumb methods achieve satisfactory results. Further, the balloon estimators are shown to be superior in the higher dimensional spaces, with higher modes or with data on non-linear manifolds. The sample point estimators show additional utility when data are scarce in low dimensions. [ARGUE] The experimental results lead us to conclude that balloon estimators such as the full rotation hyper-ellipsoid estimator will have a significant impact on data analysis algorithms which depend upon an underlying density estimates with larger volumes of higher dimensional data.

~~

9. References

Must include at a minimum:

1. The name of the author or authors
2. The full title of the source material
3. The name of the publisher
4. The date of publication

Bonus:

5. The page number of the source material