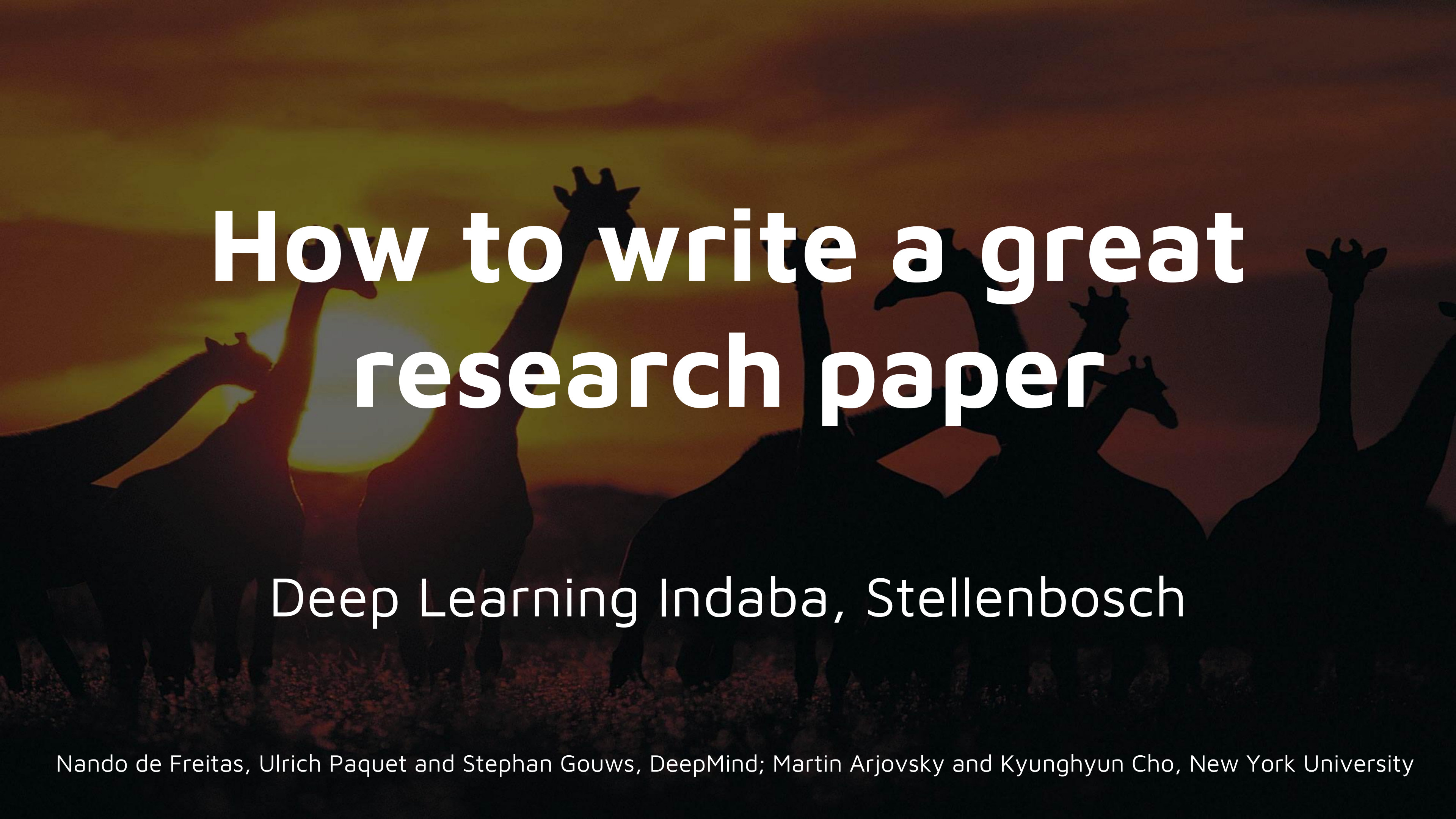


How to write a great research paper

Deep Learning Indaba, Stellenbosch

Nando de Freitas, Ulrich Paquet and Stephan Gouws, DeepMind; Martin Arjovsky and Kyunghyun Cho, New York University

How to write a ~~great~~ any research paper

Deep Learning Indaba, Stellenbosch

Nando de Freitas, Ulrich Paquet and Stephan Gouws, DeepMind; Martin Arjovsky and Kyunghyun Cho, New York University

Why this session?

To help you communicate your ideas in the best and clearest possible way

...from being paper reviewers and writers, being examiners, being on grants committees, being paper readers

Outline

- Panel discussion
    - Simon Peyton Jones’s 7 simple suggestions
- Some thoughts by Ulrich
- Some thoughts by Stephan
- Some thoughts by Nando
- Some thoughts by Martin
- Some thoughts by Kyunghyun

Simon Peyton Jones's 7 simple suggestions

1. Don't wait: write
2. Identify your key idea
3. Tell **one** story
4. Nail your contributions to the mast
5. Related work: later
6. Put your readers first
7. Listen to your readers

Link (videos, slides) here.


1. Don’t wait. Write

2. Identify your key idea

3. Tell one story

4. Nail your contributions to the mast

5. Related work: later


6. Put your readers first

7. Listen to your readers

A few thoughts from Ulrich

Framing research problems

Do you have a clear problem statement in the abstract?

It is clear from the paper?

Can you write a research statement for your paper in a single sentence?

If a reviewer cannot form such a sentence for your paper after reading just the abstract, then your paper is usually doomed.

Framing research problems

The “one thing” is stated in the first two lines of the abstract...

Abstract

> How can we efficiently propagate uncertainty in a latent state representation with recurrent neural networks? This paper introduces *stochastic recurrent neural networks* which glue a deterministic recurrent neural network and a state space model together to form a stochastic and sequential neural generative model.

The clear separation of deterministic and stochastic layers allows a structured variational inference network to track the factorization of the model’s posterior distribution. By retaining both the nonlinear recursive structure of a recurrent neural network and averaging over the uncertainty in a latent path, like a state space model, we improve the state of the art results on the Blizzard and TIMIT speech modeling data sets by a large margin, while achieving comparable performances to competing methods on polyphonic music modeling.

Example paper: Sequential Neural Models with Stochastic Layers (NIPS oral)

Clear framing helps readers

Reviewer 4

Summary

solution

problem

State space models (SSM) are widely used in statistics and machine learning to model time series data. The problem is that they cannot capture long-term dependencies. This paper provides a solution to this problem by parameterizing the state transitions of the SSM with a neural network that depends on the hidden state of a recurrent neural network - thus taking advantage of the memory of the rnn and introducing nonlinearity. The paper also presents a principled way of doing inference with this new model by building an inference network that captures the structure of the true posterior distribution of the latent states.

Qualitative Assessment

contributions clear and easy to enumerate

The paper is well written and adds significant improvement to previous work on adding stochasticity to recurrent neural nets. These improvements include: 1 - a new model for sequential data that can both capture long-term dependencies using recurrent neural nets and capture uncertainty well via stochastic latent states of state space models 2 - an inference network for doing posterior approximation with this new model...this inference method includes a solution to the vanishing KL term of the evidence lower bound... 3 - rich approximating variational distribution that encodes the same conditional independence relationship among the latent states as the true posterior distribution of these latent states. One avenue of improvement would be to include results of the VRNN model for the polyphonic music modeling experiment. The reason why this would be a great improvement is that the VRNN is the closest related work and the paper mentions the SRNN (new model as presented in the paper) yields better results because of the separation of the deterministic layer of the recurrent neural net and the stochastic layer of the state space model. Evidence of this is lacking for the

Framing research problems

Your research statement should be falsifiable.

A paper claims:

"To the best of our knowledge, this is most sophisticated neural network solution ever mentioned in the literature."

Reviewer: What problem does it solve? What is the benchmark? I can't measure "sophistication" :)

Keep your reasons real

Consider the opening sentence of this (fictional) introduction:

"<u>Machine learning has gathered a lot of interest recently</u>. Deep Learning is now a popular tool. We therefore use it to …"

Reviewer: This was your **one chance** to convince me of the **problem** you're working on. And now you told me you're working on it because it is popular…

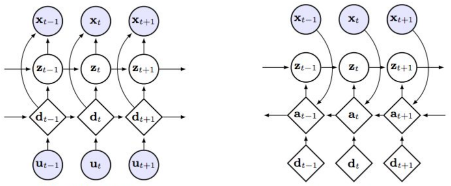
It is surprising how many papers use this kind of opening as reason.

Write to be understood

For **every line** in your paper, ask questions about your reader's **mental model**:

1. What does my reader understand up to this point?
2. What is my reader thinking at this point?
3. How will my next narrative change that?

I'm reading here. I understand some things. **One thing is currently "top of mind"**. Some things already worry me.



(a) Generative model p_θ (b) Inference network q_ϕ

Figure 2: A SRNN as a generative model p_θ for a sequence $\mathbf{x}_{1:T}$. Posterior inference of $\mathbf{z}_{1:T}$ and $\mathbf{d}_{1:T}$ is done through an inference network q_ϕ, which uses a backward-recurrent state $\mathbf{a}_t$ to approximate the nonlinear dependence of $\mathbf{z}_t$ on future observations $\mathbf{x}_{t:T}$ and states $\mathbf{d}_{t:T}$; see Equation (7).

$$p_\theta(\mathbf{x}_{1:T}, \mathbf{z}_{1:T}, \mathbf{d}_{1:T} | \mathbf{u}_{1:T}, \mathbf{z}_0, \mathbf{d}_0) = p_{\theta_x}(\mathbf{x}_{1:T} | \mathbf{z}_{1:T}, \mathbf{d}_{1:T})\, p_{\theta_z}(\mathbf{z}_{1:T} | \mathbf{d}_{1:T}, \mathbf{z}_0)\, p_{\theta_d}(\mathbf{d}_{1:T} | \mathbf{u}_{1:T}, \mathbf{d}_0)$$
$$= \prod_{t=1}^{T} p_{\theta_x}(\mathbf{x}_t | \mathbf{z}_t, \mathbf{d}_t)\, p_{\theta_z}(\mathbf{z}_t | \mathbf{z}_{t-1}, \mathbf{d}_t)\, p_{\theta_d}(\mathbf{d}_t | \mathbf{d}_{t-1}, \mathbf{u}_t) \,. \quad (1)$$

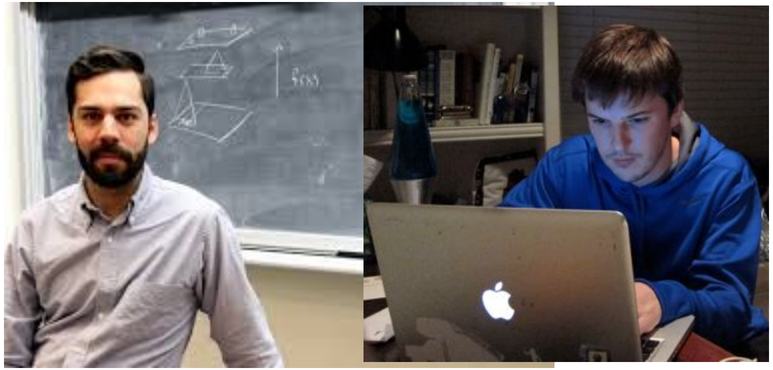
The SSM and RNN are further tied with skip-connections from $\mathbf{d}_t$ to $\mathbf{x}_t$. The joint density in (1) is parameterized by $\theta = \{\theta_x, \theta_z, \theta_d\}$, which will be adapted together with parameters ϕ of a so-called "inference network" q_ϕ to best model N independently observed data sequences $\{\mathbf{x}^i_{1:T_i}\}_{i=1}^N$ that are described by the log marginal likelihood or evidence

$$\mathcal{L}(\theta) = \log p_\theta\left(\{\mathbf{x}^i_{1:T_i}\} \,|\, \{\mathbf{u}^i_{1:T_i}, \mathbf{z}^i_0, \mathbf{d}^i_0\}_{i=1}^N\right) = \sum_i \log p_\theta(\mathbf{x}^i_{1:T_i} | \mathbf{u}^i_{1:T_i}, \mathbf{z}^i_0, \mathbf{d}^i_0) = \sum_i \mathcal{L}_i(\theta) \,. \quad (2)$$

Throughout the paper, we omit superscript i when only one sequence is referred to, or when it is clear from the context. In each log likelihood term $\mathcal{L}_i(\theta)$ in (2), the latent states $\mathbf{z}_{1:T}$ and $\mathbf{d}_{1:T}$ were averaged out of (1). Integrating out $\mathbf{d}_{1:T}$ is done by simply substituting its deterministically obtained value, but $\mathbf{z}_{1:T}$ requires more care, and we return to it in Section 3.2. Following Figure 2a, the states $\mathbf{d}_{1:T}$ are determined from $\mathbf{d}_0$ and $\mathbf{u}_{1:T}$ through the recursion $\mathbf{d}_t = f_{\theta_d}(\mathbf{d}_{t-1}, \mathbf{u}_t)$. In our implementation f_{θ_d} is a GRU network with parameters θ_d. For later convenience we denote the value of $\mathbf{d}_{1:T}$, as computed by application of f_{θ_d}, by $\widetilde{\mathbf{d}}_{1:T}$. Therefore $p_{\theta_d}(\mathbf{d}_t | \mathbf{d}_{t-1}, \mathbf{u}_t) = \delta(\mathbf{d}_t - \widetilde{\mathbf{d}}_t)$, i.e. ... tribution centered at $\widetilde{\mathbf{d}}_{1:T}$.

Blank slate

Write to be understood



For **every line** in your paper, ask questions about your reader's **mental model**:

1. What does my reader understand up to this point?
2. What is my reader thinking at this point?
3. How will my next narrative change that?

* These ideas were picked up from Chris Maddison & Danny Tarlow, when writing "A* sampling" (NIPS best paper 2014)

“Top of mind” example

We had to introduce a function and notation here. It’s important, but not directly applicable to what we want to explain right now.

also noted in [20]. Instead of factorizing q_ϕ as a m… p
the structured form of the posterior factors, includ… $\mathbf{z}_t$’s dependence on $\mathbf{z}_t$… variational
approximation

$$q_\phi(\mathbf{z}_{1:T}|\mathbf{d}_{1:T}, \mathbf{x}_{1:T}, \mathbf{z}_0) = \prod_t q_\phi(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{d}_{t:T}, \mathbf{x}_{t:T}) = \prod_t q_{\phi_\mathrm{z}}(\mathbf{z}_t|\mathbf{z}_{t-1}, \mathbf{a}_t = \boxed{g_{\phi_\mathrm{a}}}(\mathbf{a}_{t+1}, [\mathbf{d}_t, \mathbf{x}_t])) \,, \tag{7}$$

where $[\mathbf{d}_t, \mathbf{x}_t]$ is the concatenation of the vectors $\mathbf{d}_t$ and $\mathbf{x}_t$. The graphical model for the inference network is shown in Figure 2b. Apart from the direct dependence of the posterior approximation at time t on both $\mathbf{d}_{t:T}$ and $\mathbf{x}_{t:T}$, the distribution also depends on $\mathbf{d}_{1:t-1}$ and $\mathbf{x}_{1:t-1}$ through $\mathbf{z}_{t-1}$. We mimic each posterior factor’s nonlinear long-term dependence on $\mathbf{d}_{t:T}$ and $\mathbf{x}_{t:T}$ through a backward-recurrent function g_{ϕ_a}, shown in (7), which we will return to in greater detail in Section 3.3. The

We want to keep a **flowing story line**. A reader would be worried that the notation and function is not explained. So, we describe what it is, where it is introduced, and tell the reader *not to worry*: we’ll explain it soon :)

Keeping a flowing story line

Qualitative Assessment

I really enjoyed the paper and the idea. It's written very clearly. And the experiment and its analysis are thorough. Also, the improvement is significant.

Don’t confuse or frustrate your readers, by…

- Switching context “mid way” / “mid flight”
- Using undefined notation
- Changing notation

They want to enjoy reading your paper!

Notation, notation

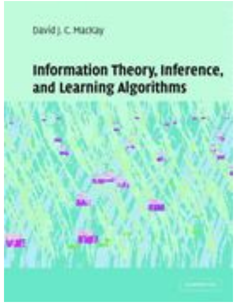
When a concept requires specific notation, I like to introduce the notation *with* the concept. As early as possible!

It helps shape your reader’s mental model, and minimizes later context switching

2 Recurrent Neural Networks and State Space Models

Recurrent neural networks and *state space models* are widely used to model temporal sequences of vectors $\mathbf{x}_{1:T} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_T)$ that possibly depend on inputs $\mathbf{u}_{1:T} = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_T)$. Both

Notation, notation



What does $\prod_{n=1}^{N}$ mean? This is like the summation $\sum_{n=1}^{N}$ but it denotes a product. It's pronounced 'product over n from 1 to N'. So, for example,

$$\prod_{n=1}^{N} n = 1 \times 2 \times 3 \times \cdots \times N = N! = \exp\left[\sum_{n=1}^{N} \ln n\right]. \tag{A.1}$$

I like to choose the name of the free variable in a sum or a product – here, n – to be the lower case version of the range of the sum. So n usually runs from 1 to N, and m usually runs from 1 to M. This is a habit I learnt from Yaser Abu-Mostafa, and I think it makes formulae easier to understand.

Notation, notation

Subscripts, consistency, etc

Huh??

$$f(\mathbf{x}_i) \cdots f(\mathbf{x}_i) \cdots f(x_y)$$

Do everything you can to not mess with the mind of your reader

Be academically honest. Don’t oversell

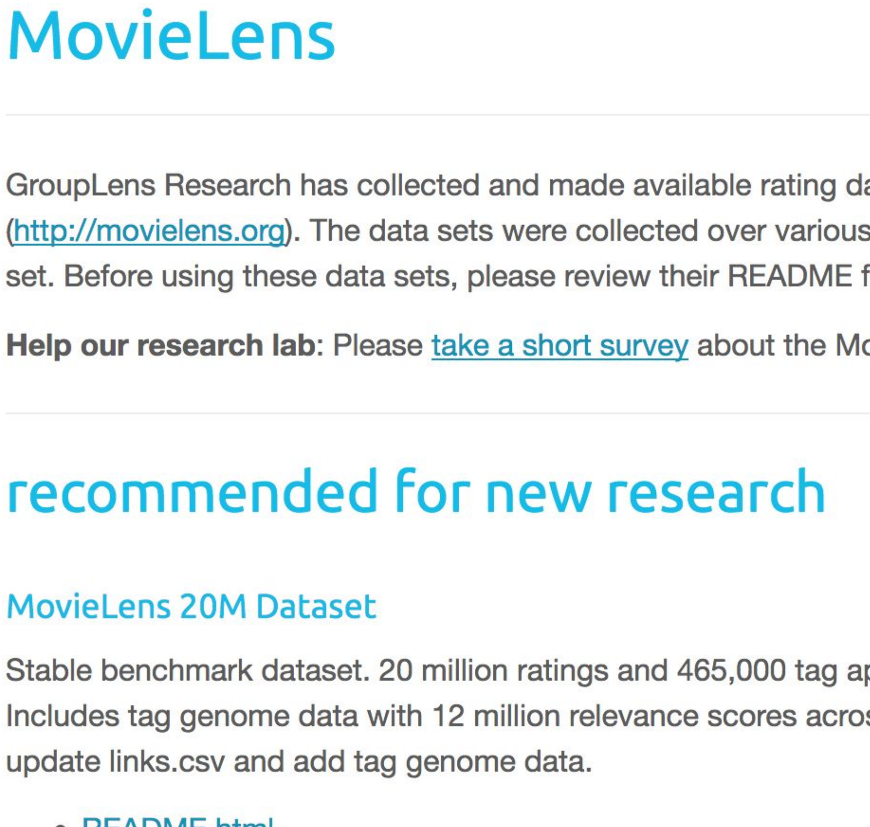
Consider this abstract:

“Internet scale recommender problems … big data …”

And the small print in the experimental results section:

“We test on MovieLens 100K dataset”

Reviewer: Uhm, internet scale? Why didn’t you download the 20M version?



Be academically honest. Don’t oversell

Consider this abstract:

“We outperform the state of the art”

And the small print in the experimental results section:

“We have one result where we beat the state-of-the-art by 0.1%”

Reviewer: I started reading your paper, expecting a method that outperforms everything I’ve ever seen before. And now I’m let down. I feel you weren’t honest with me from the beginning.


A few thoughts from Stephan

High-level thoughts

- Writing IS research. It is how we formulate, crystallize, and communicate our ideas. Make it a daily habit!

High-level thoughts

- Writing IS research. It is how we formulate, crystallize, and communicate our ideas. Make it a daily habit!

- Good papers are written, great papers are *re*-written, so get the first draft done asap. Just do it.

    - *"Finish the paper 2 weeks before actual deadline. Get feedback. Rewrite."* -- Slav Petrov (Best paper NAACL 2012)

High-level thoughts

- Writing IS research. It is how we formulate, crystallize, and communicate our ideas. Make it a daily habit!
- Good papers are written, great papers are *re*-written, so get the first draft done asap. Just do it.
    - *"Finish the paper 2 weeks before actual deadline. Get feedback. Rewrite."* -- Slav Petrov (Best paper NAACL 2012)
- Good papers leave the reader with one solution to solving a specific problem; great papers leave the reader with new ideas for their own problems.
    - Don't leave it up to your reader, always ask yourself "what have I learned" and make that explicit.

Write to...

- ...discover/understand (for **yourself**)
- ...get accepted (for the **reviewer**)
- ...enlighten (for the **reader**)

Write to discover/understand (for yourself)

Be precise in what you are trying to do. Use simple language. If you can't describe your idea in 2-3 simple sentences, maybe you don't understand it that well yourself. Work at it until you can. Read Strunk & White. It is a timeless classic.

Write to discover/understand (for yourself)

Be precise in what you are trying to do. Use simple language. If you can’t describe your idea in 2-3 simple sentences, maybe you don’t understand it that well yourself. Work at it until you can. Read Strunk & White. It is a timeless classic.

State your hypotheses. Make them falsifiable. It’s how science works.

Write to discover/understand (for yourself)

Be precise in what you are trying to do. Use simple language. If you can't describe your idea in 2-3 simple sentences, maybe you don't understand it that well yourself. Work at it until you can. Read Strunk & White. It is a timeless classic.

State your hypotheses. Make them falsifiable. It's how science works.

I like to do a "blank slate" experiment once I have a first draft written (of a paragraph, a section, or a paper): I start reading what I wrote as if I was a new reader reading this for the first time, and I ask "why?" or "why should I care?" after every sentence. It forces you to be more explicit and get to the point.

Write to discover/understand (for yourself)

Be precise in what you are trying to do. Use simple language. If you can't describe your idea in 2-3 simple sentences, maybe you don't understand it that well yourself. Work at it until you can. Read Strunk & White. It is a timeless classic.

State your hypotheses. Make them falsifiable. It's how science works.

I like to do a "blank slate" experiment once I have a first draft written (of a paragraph, a section, or a paper): I start reading what I wrote as if I was a new reader reading this for the first time, and I ask "why?" or "why should I care?" after every sentence. It forces you to be more explicit and get to the point.

Keep a daily "snippets" Google doc where you continually summarize your work, results, and ideas. It helps refine your thoughts.

Write to get accepted

Reviewers are the unpaid, overworked, gate-keepers of science. Don’t waste their time.

Write to get accepted

Reviewers are the unpaid, overworked, gate-keepers of science. Don’t waste their time.

Not all reviewers will be familiar with your work. It is up to you to bring your message across in the clearest way possible! Still, it’s a very noisy process.

Write to get accepted

Reviewers are the unpaid, overworked, gate-keepers of science. Don’t waste their time.

Not all reviewers will be familiar with your work. It is up to you to bring your message across in the clearest way possible! Still, it’s a very noisy process.

Reviewers usually have less than 1h per paper, sometimes only 30min. They are basically trying to answer the question “How can I justify rejecting this?” Get to the point!

Write to get accepted

Reviewers are the unpaid, overworked, gate-keepers of science. Don't waste their time.

Not all reviewers will be familiar with your work. It is up to you to bring your message across in the clearest way possible! Still, it's a very noisy process.

Reviewers usually have less than 1h per paper, sometimes only 30min. They are basically trying to answer the question "How can I justify rejecting this?" Get to the point!

Do not write your personal journey. Science is a random walk, but we tell it like a shortest path.

Write to enlighten

Academic writing is not like writing prose. There are no set ups, no surprises, and no punch lines. Doesn't mean it has to be dry, though!

Ed Hovy used to say: *Tell the reader…:*

1. *What you want to tell her;*
2. *Then tell her; and finally*
3. *Tell her what you told her*

This hour-glass structure (general to specific to general) works very well at the level of paragraphs, sections, and papers.

My (ideal) process

1. Write a rough 2-4 sentence abstract first (what, why, how)
2. Write the Model description next. This is easy, it’s the idea you’re trying out.
3. Then write the Experimental section (ie get the results). Add your results tables, create your graphs.
4. Then write the Discussion & Conclusion sections (what did we learn from this?)
5. Finally write the Introduction (expand #1 by framing the research question, and introducing relevant background work)
6. Write the Abstract last.

Low-level ideas

- Learn Latex and use Sharelatex or some other collaborative editing platform with revision control
    - Split different sections into different files (easier to track, can export experiments directly to latex tables, etc)
- Download the conference style sheet and use it from the beginning (start early!)
- Add colourized TODO notes (different colour for each author) in the document using \newcommand. This way you can easily remove them to generate a draft for submission.

A few thoughts from Nando


A few thoughts from Martin

Claims

- Never make a claim that is not directly validated by a theorem, an experiment, or a reference.

Claims

- Never make a claim that is not directly validated by a theorem, an experiment, or a reference.
- E.g. If you’re claiming that a model has vanishing gradients, calculate the norm of the gradients!.

Claims (cont)

- Make claims that are useful to tell the story. More claims is not always better.

Claims (cont)

- Make claims that are useful to tell the story. More claims is not always better.
- It’s important to write the paper first to know what claims you have to make, and what experiments / theory are needed to validate your claims.

Highlight problems and negative results

- People will run into them. Better to prepare them, and propose potential solutions.

Highlight problems and negative results

- People will run into them. Better to prepare them, and propose potential solutions.
- These are avenues for further work. Other readers will often figure out how to solve them, and your algorithm will be even better later.

Flow of the paper

- You have to convince the reader to keep reading at every paragraph. Do not assume that the reader wants to read your paper, or that they will read all of it.

Flow of the paper

- You have to convince the reader to keep reading at every paragraph. Do not assume that the reader wants to read your paper, or that they will read all of it.
- E.g. Before switching sections, always have the last paragraph of the previous one introduce it. More importantly, explain why the next section is needed.

Flow of the paper (theory)

- Only have theorems that improve your story.

Flow of the paper (theory)

- Only have theorems that improve your story.
- E.g. You made a claim that your algorithm approximates some loss function -> prove a theorem quantifying this approximation.

Flow of the paper (theory)

- Only have theorems that improve your story.
- E.g. You made a claim that your algorithm approximates some loss function -> prove a theorem quantifying this approximation.
- Unless proofs are important for the story -> appendix.

Flow of the paper (theory)

- Only have theorems that improve your story.
- E.g. You made a claim that your algorithm approximates some loss function -> prove a theorem quantifying this approximation.
- Unless proofs are important for the story -> appendix.
- Do **not** say “Here are some guarantees from our algorithm”. Introduce and justify its existence first.

A last note on theory

- Have theorems that are meaningful.

A last note on theory

- **Have theorems that are meaningful.**
- E.g. If your algorithm is meant for high dimensions, **do not** have bounds that depend exponentially in the number of dimensions.

A last note on theory

- **Have theorems that are meaningful.**
- E.g. If your algorithm is meant for high dimensions, **do not** have bounds that depend exponentially in the number of dimensions.
- Be honest!

Final tips

- Make a bullet list with the core contributions at the end of your introduction!

Final tips

- Make a bullet list with the core contributions at the end of your introduction!
- It will tell readers and reviewers concisely what claims to expect (and an idea of the experiments / results).

Final tips

- Figures and their captions are the first thing the reader will see!

Final tips

- Figures and their captions are the first thing the reader will see!
- Make them self-contained, with extremely concise and clear captions, saying what they mean and their conclusion.

Final tips

- When there's a paper you like, take literally notes, and try to understand why you liked reading it!

A few thoughts from Kyunghyun