MEETING READERS' EXPECTATIONS

Presenting context and content effectively

Introduction

- Dr Simon Milligan, University of Bern – acalanser.ch
- Dr Réka Mihálka, ETH Zurich – manuscriptmanufacture.com

Overview

- Context and content: A cognitive view
- Application
- Q&A

CONTEXT AND CONTENT

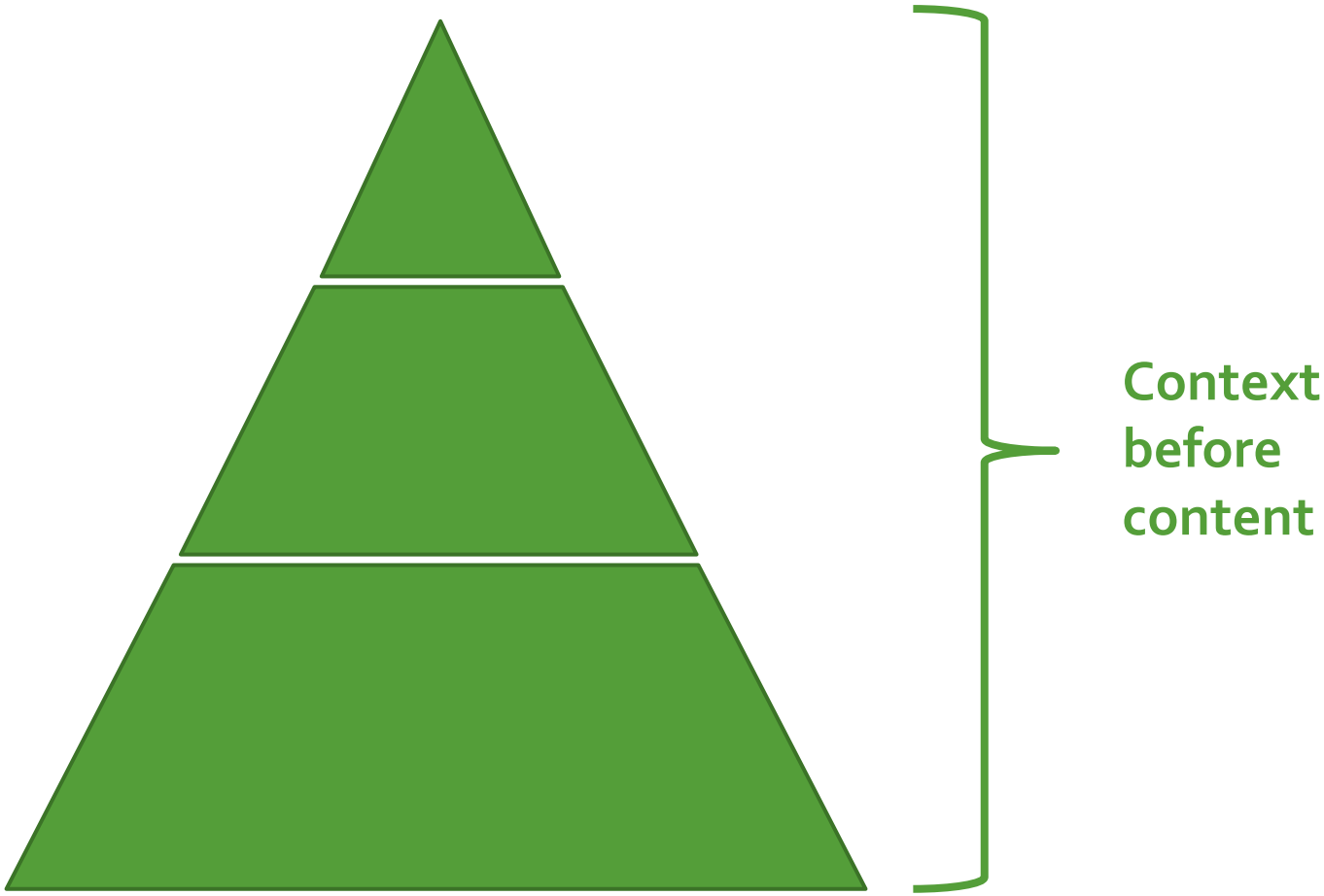
A cognitive view

The ebb and flow of understanding

- Familiar information (context) before new information (content)
- The alternating dynamic of context and content
  - Activates reader's background knowledge
  - Focuses the reader's attention
  - Creates expectations for the new information
  - Frames the new information
  - Helps the reader process information
  - Orients the reader within the text
  - Provides navigation aid

Applicability

- Macrostructure
- Paragraph structure
- Sentence structure



SENTENCE STRUCTURE

Exercise 1. Compare the two paragraphs below and decide which one has better flow.

VERSION 1. Human pose recovery has garnered a lot of research interest in the vision community. A wide range of applications such as augmented reality, virtual shopping, human–robot interaction, etc. use it. The availability of 3D pose supervision from large-scale datasets [19, 42, 57] has made recent advances in human pose recovery possible. Although in-studio benchmarks enable superior performance, the models usually suffer from poor cross-dataset performance. Domain bias [36, 27] is often induced by training on synthetic and in-studio datasets, which lack diversity in subject appearance, lighting, background variation, among others, thereby restricting generalizability. The question arises from this: across diverse data domains, how can we bridge performance gaps?

VERSION 2. Human pose recovery has garnered a lot of research interest in the vision community. It is extensively used in a wide range of applications such as augmented reality, virtual shopping, human–robot interaction, etc. Recent advances in human pose recovery is largely attributed to the availability of 3D pose supervision from large-scale datasets [19, 42, 57]. Although the models achieve superior performance on in-studio benchmarks, they usually suffer from poor cross-dataset performance. Synthetic and in-studio datasets lack diversity in subject appearance, lighting, background variation, among others, which is why training on these datasets can easily induce a domain bias [36, 27] and restrict generalizability. This poses the question: how can we bridge performance gaps across diverse data domains?

Adapted from: Ramesha Rakesh Mugaludi, Jogendra Nath Kundu, Varun Jampani, Venkatesh Babu R. "Aligning Silhouette Topology for Self-Adaptive 3D Human Pose Recovery" Advances in Neural Information Processing Systems 34 (NeurIPS 2021)

Exercise 1. Compare the two paragraphs below and decide which one has better flow.

VERSION 1. Human pose recovery has garnered a lot of research interest in the vision community. A wide range of applications such as augmented reality, virtual shopping, human–robot interaction, etc. use it. The availability of 3D pose supervision from large-scale datasets [19, 42, 57] has made recent advances in human pose recovery possible. Although in-studio benchmarks enable superior performance, the models usually suffer from poor cross-dataset performance. Domain bias [36, 27] is often induced by training on synthetic and in-studio datasets, which lack diversity in subject appearance, lighting, background variation, among others, thereby restricting generalizability. The question arises from this: across diverse data domains, how can we bridge performance gaps?

VERSION 2. Human pose recovery has garnered a lot of research interest in the vision community. It is extensively used in a wide range of applications such as augmented reality, virtual shopping, human–robot interaction, etc. Recent advances in human pose recovery is largely attributed to the availability of 3D pose supervision from large-scale datasets [19, 42, 57]. Although the models achieve superior performance on in-studio benchmarks, they usually suffer from poor cross-dataset performance. Synthetic and in-studio datasets lack diversity in subject appearance, lighting, background variation, among others, which is why training on these datasets can easily induce a domain bias [36, 27] and restrict generalizability. This poses the question: how can we bridge performance gaps across diverse data domains?

Adapted from: Ramesha Rakesh Mugaludi, Jogendra Nath Kundu, Varun Jampani, Venkatesh Babu R. "Aligning Silhouette Topology for Self-Adaptive 3D Human Pose Recovery" Advances in Neural Information Processing Systems 34 (NeurIPS 2021)

Sequencing old and new information

Two basic approaches:

1. Old information → New information (Sentence 1)
   → New information (Sentence 2)
   → New information (Sentence 3)

2. Old information → New information (Sentence 1)
   Old information → New information (Sentence 2)
   Old information → New information (Sentence 3)

Exercise 2. Which structure is used in the paragraph below?

Deep Neural Networks (DNNs) have achieved unprecedented success in a wide range of applications due to their remarkably high accuracy [15]. However, this high performance stems from significant growth in DNN model size; i.e., massive overparameterization. Furthermore, these highly overparameterized models are known to be susceptible to the out-of-distribution (OOD) shifts encountered during their deployment in the wild [5]. This resource-inefficiency and OOD brittleness of state-of-the-art (SOTA) DNNs severely limits the potential applications DL can make an impact on.

Source: James Diffenderfer, Brian Bartoldson, Shreya Chaganti, Jize Zhang, Bhavya Kailkhura. "A Winning Hand: Compressing Deep Networks Can Improve Out-of-Distribution Robustness" Advances in Neural Information Processing Systems 34 (NeurIPS 2021)

Exercise 2. Which structure is used in the paragraph below?

Deep Neural Networks (DNNs) have achieved unprecedented success in a wide range of applications due to their remarkably high accuracy [15]. However, this high performance stems from significant growth in DNN model size; i.e., massive overparameterization. Furthermore, these highly overparameterized models are known to be susceptible to the out-of-distribution (OOD) shifts encountered during their deployment in the wild [5]. This resource-inefficiency and OOD brittleness of state-of-the-art (SOTA) DNNs severely limits the potential applications DL can make an impact on.

Old information	New information
Deep neural networks (DNNs)	have achieved unprecedented success…
this high performance	… massive over-parameterization
… these highly over-parameterized models…	… susceptible to OOD shifts…
This resource inefficiency and OOD brittleness	… limits the potential applications

Old (contextual) information in the sentence

- **Before or in the subject position**
  - Before: Graph neural networks (GNNs) are… **In GNNs,** the prediction target can be set…
  - In: Graph neural network (GNNs) are … **GNNs** have been applied to…
- **Common forms:**
  - Repetition (for technical terms): *model – model*
  - Synonym (for non-technical terms): *high accuracy – high performance*
  - Hypernym (group noun, umbrella term): *overparameterization – this resource inefficiency*
  - Metonym (attribute, part): *DNNs – the OOD brittleness of state-of-the-art DNNs*
  - Word transformation (verb -> noun): *overparameterization – overparameterized*
  - Demonstrative pronoun/adjective: *this high performance, these […] models*
  - Condensation of the noun phrase: *out-of-distribution – OOD shifts*

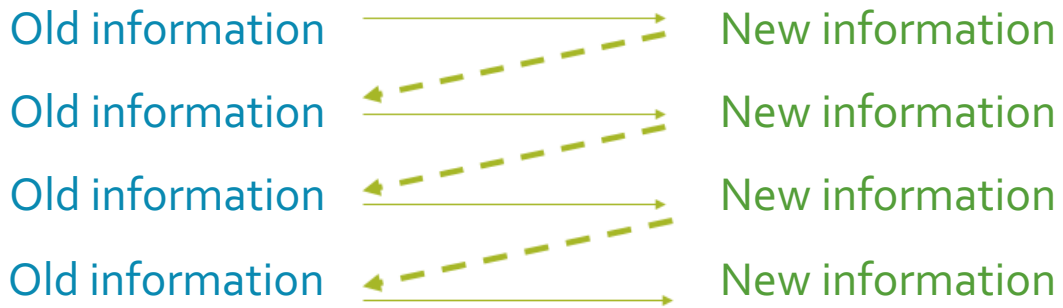
Exercise 3. Fill in the old information.

The worst-case optimal algorithms, however, tend to be too conservative in actual practice. (1)_____ **[Demonstrative pronoun]** is because true worst-case environments are quite rare in real-world applications. Rather, (2) _____ **[Repetition]** may have structures that are convenient for the learner, and it is desirable that the algorithm takes advantage of such structures to improve performance. To exploit such (3) _____ **[Repetition]**, two main categories of approaches have been studied: adapting to (nearly) stochastic environments and developing data-dependent regret bounds.

Source: Shinji Ito "Hybrid Regret Bounds for Combinatorial Semi-Bandits and Adversarial Linear Bandits". Advances in Neural Information Processing Systems 34 (NeurIPS 2021).

Exercise 3. Fill in the old information.

The worst-case optimal algorithms, however, tend to be too conservative in actual practice. **This [subject position]** is because true worst-case environments are quite rare in real-world applications. Rather, **the environments [subject position]** may have structures that are convenient for the learner, and it is desirable that the algorithm takes advantage of such structures to improve performance. To exploit such **structures [before the subject]**, two main categories of approaches have been studied: adapting to (nearly) stochastic environments and developing data-dependent regret bounds.



Common pitfalls

- No context in the sentence, only content ⟶ Fragmentation, lack of flow
- Old information is at the end ⟶ Tiresome reading process (expectations frustrated)
- New information before the subject ⟶ Disruption of the flow (unless the new information reframes the old information in the subject position)

PARAGRAPH STRUCTURE

Establishing context with a topic sentence

- The first sentence in a paragraph, which
  - identifies the main topic,
  - creates expectations for the remainder of the paragraph,
  - enhances coherence and the flow, and
  - aids navigation within the text (similar to a subheading).

Reinforcing context with a preview sentence

- Is optional
- Creates more specific expectations about the sequence of information in the paragraph (‘table of contents’)
- Only used when several related points are mentioned in the paragraph
- Can be merged with the topic sentence

Paragraph structure

- Topic sentence (context)
- Preview sentence (context)
- Main point 1 (content)
  - Further details
- Main point 2 (content)
  - Further details

Establishing expectations

Meeting expectations

(Other paragraph patterns are also possible.)

Exercise 4. Can you identify the function of each sentence in this paragraph?

While we have shown benefits of our method, it has limitations and interesting future work. First, matching using the homography matrix calculated by RANSAC might not be the best option for our hypergraph propagation. It does not consider local features' contextual cues when applying matching, and it has the offline homography calculation overhead. Recent deep-learning-based matching techniques such as SuperGlue [29] show performance improvement in many tasks in terms of both speed and accuracy and can also be helpful for our hypergraph propagation. Second, to achieve a better result, both hypergraph propagation and community selection can be combined with existing query expansion, diffusion, and spatial verification methods. For example, it is possible to adapt region diffusion [10] to the hypergraph model or adapt community selection on ordinary diffusion methods.

Source: Guoyuan An, Yuchi Huo, Sung-eui Yoon, Hypergraph Propagation and Community Selection for Objects Retrieval, Advances in Neural Information Processing Systems 34 (NeurIPS 2021)

Exercise 4. Can you identify the function of each sentence in this paragraph?

While we have shown benefits of our method, it has limitations and interesting future work. [TOPIC AND PREVIEW SENTENCE] First, matching using the homography matrix calculated by RANSAC might not be the best option for our hypergraph propagation. [MAIN POINT 1] It does not consider local features' contextual cues when applying matching, and it has the offline homography calculation overhead. [FURTHER DETAILS] Recent deep-learning-based matching techniques such as SuperGlue [29] show performance improvement in many tasks in terms of both speed and accuracy and can also be helpful for our hypergraph propagation. [FURTHER DETAILS] Second, to achieve a better result, both hypergraph propagation and community selection can be combined with existing query expansion, diffusion, and spatial verification methods. [MAIN POINT 2] For example, it is possible to adapt region diffusion [10] to the hypergraph model or adapt community selection on ordinary diffusion methods. [FURTHER DETAILS]

Writing an effective topic sentence

- Clear
- Precise
- Short
- Tailored

Exercise 5. Write a topic sentence for the paragraph below.

[…] The authors conjecture that the sampling based optimization procedure of UMAP prevents the minimization of the supposed loss function, thus not reproducing the high-dimensional similarities in embedding space. They substantiate this hypothesis by qualitatively estimating the relative size of attractive and repulsive forces. In addition, they implement a BarnesHut approximation to the loss function (6) and find that it yields a diverged embedding. We analyze UMAP's sampling procedure in depth, compute UMAP's true loss function in closed form and contrast it against the supposed loss in Section 5. Based on this analytic effective loss function, we can further explain Böhm et al. [4]'s empirical finding that the specific high-dimensional similarities provide little gain over the binary weights of a shared kNN graph, see Section 6. Finally, our theoretical framework leads us to a new tentative explanation for UMAP's success in Section 7.

Source: Sebastian Damrich, Fred A. Hamprecht. "On UMAP's True Loss Function." Advances in Neural Information Processing Systems 34 (NeurIPS 2021).

Exercise 5. Write a topic sentence for the paragraph below.

Our work aligns with Böhm et al. [4]. The authors conjecture that the sampling based optimization procedure of UMAP prevents the minimization of the supposed loss function, thus not reproducing the high-dimensional similarities in embedding space. They substantiate this hypothesis by qualitatively estimating the relative size of attractive and repulsive forces. In addition, they implement a BarnesHut approximation to the loss function (6) and find that it yields a diverged embedding. We analyze UMAP's sampling procedure in depth, compute UMAP's true loss function in closed form and contrast it against the supposed loss in Section 5. Based on this analytic effective loss function, we can further explain Böhm et al. [4]'s empirical finding that the specific high-dimensional similarities provide little gain over the binary weights of a shared kNN graph, see Section 6. Finally, our theoretical framework leads us to a new tentative explanation for UMAP's success in Section 7.

Common pitfalls

- The topic sentence is not a good fit for the rest of the paragraph ⟹ False expectations
- The topic sentence is too long ⟹ Possibility of misinterpretation
- There is no topic sentence ⟹ Difficult reading experience

MACROSTRUCTURE

Sequencing context and content in sections

- Abstract: situation – problem – solution – evaluation
- Introduction: broad perspective – literature review – gap – contribution
- Main sections: Background – methods – results – discussion – conclusion

Context and content within a section

4 Optimization of Deep Neural Networks is Roughly Convex

Section 3 has shown that the extent to which gradient descent matches gradient flow depends on "how convex" the objective function is around the gradient flow trajectory. More precisely, the larger (less negative or more positive) the minimal eigenvalue of the Hessian is around this trajectory, the longer gradient descent (with given step size) is guaranteed to follow it. In this section we establish that over training losses of deep neural networks (fully connected as well as convolutional) with homogeneous activations (e.g. linear, rectified linear or leaky rectified linear), when emanating from near-zero initialization (as commonly employed in practice), trajectories of gradient flow are "roughly convex," in the sense that the minimal eigenvalue of the Hessian along them is far greater than in arbitrary points in space, particularly towards convergence. This finding suggests that when optimizing deep neural networks, gradient descent may closely resemble gradient flow. We demonstrate a formal application of the finding in Section 5, translating an analysis of gradient flow over deep linear neural networks into a guarantee of efficient convergence (to global minimum) for gradient descent, which applies almost surely with respect to a random near-zero initialization.

Source: Omer Elkabetz, Nadav Cohen, "Continuous vs. Discrete Optimization of Deep Neural Networks". Advances in Neural Information Processing Systems 34 (NeurIPS 2021)

Exercise 6: Food for thought

Which sections provide context to the content of which other sections? List a few pairs.

What adjustments do you need to make to an introduction if you don’t have a background section?

To what other aspects of the scientific process can you apply the principle of «context before content»?

Do’s and don’ts

Do’s

- Tailor the amount and detail of the background (context) to your target readership
- Separate your own work from the context (“Here, we…”, “In this study, …”)
- Remind the reader at the beginning of a new section of the main points established earlier

Don’ts

- Don’t start the introduction with “We study” or “In this paper, we …”: the lack of contextual information will challenge the reader
- Don’t assume the reader is as familiar with your context as you are (curse of knowledge)

Thank you for your attention.

Q&A:

Simon answers specific questions on Zoom.

Réka answers general-interest questions in front of the audience.

Please fill in this survey: https://forms.gle/RyKWxin5xFz1iBsMA