Research Fast and Slow

Min-Yen Kan
National University of Singapore

Fast and Slow

<table>
<thead>
<tr>
<th>System 1</th>
<th>System 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Automatic</td>
<td>Controlled</td>
</tr>
<tr>
<td>Intuitive</td>
<td>Analytical</td>
</tr>
<tr>
<td>Parallel</td>
<td>Serial</td>
</tr>
<tr>
<td>Associative</td>
<td>Logical</td>
</tr>
</tbody>
</table>

Daniel Kahneman


All tracings CC BY 4 by Min
The Neural Net – judgements in 1 second

To think about: what’s the loss function of research?
The Age of Accelerations

Thank You for Being Late

Kurzweil’s “Second half of the chessboard”

His three accelerations
- Moore’s law
- Globalization
- Mother Nature
Nearly instantaneous galaxy-wide communications accelerates exchange of information and ideas, greatly enhancing the research capabilities of scientists and spying power of agents (+100% Research)

"arXiv"

Image originally from Microprose software
A typical timeline of attention

- The first tweets appear
- News outlets pick up the research
- People start to bookmark and share it on other social networks
- People read, comment and blog about it
- Wikipedia articles are updated with references
- It gets featured as a research highlight
- Citations in other articles and policy documents appear

Volume of attention

Hours

Days

Months

Time since publication

Research published

24 Aug 2018

Research Faster

https://symplectic.co.uk/guest-blog/research-data-mechanics/
Research Too Fast?

We have our own age of accelerations for research

1. arXiv – 驚輸 “kiasu”

2. GitHub / Jupyter Notebook

3. Shared Tasks
   Overheard Yejin Choi: “Working on sQuaD makes people feel better”

“We shape our buildings, thereafter our buildings shapes us.” – Winston Churchill
Here is a neat summary of the current state of interpretations of skull comparisons in biological anthropology:

Reviewer 2 must be stopped

Reviewers also favor defensible positions.

Safer to use standard metrics for acceptance.

Understandable for journals but for conferences?
Loss function of research

Beam search analogy

Accelerations make the gradient steeper

Overload favors the convenient

But scholars can’t do random restart

Suboptimal local minima
Q: When your ML doesn’t outperform, what do you do?

Your turn: select an answer.

1. Sigh.

2. Gather more data. Maybe your model needs more data to find proper weights.

3. Simplify your model. Maybe it’s easier just to do more hyperparameter searching.

4. Read your arXiv feed.

5. Study your problem more. Let your text speak.
What are the fundamental differences between Natural Language Processing and Computational Linguistics?

The above answers are all good. I'd like to offer another perspective that I learned while teaching digital libraries that draws on the analogy used in biology:

**Computational biology** = the study of biology using computational techniques. The goal is to learn new biology, knowledge about living systems. It is about science.

**Bioinformatics** = the creation of tools (algorithms, databases) that solve problems. The goal is to build useful tools that work on biological data. It is about engineering.

To make the analogy for any field X, we thus have "Computational X" and "X-omics". In NLP/CL, NLP is the equivalent of "Linguamatics".

I don't really subscribe to the notion that CL encompasses NLP or vice versa. They both have a purpose. CL studies human language to computationally understand how we as humans have the capacity to produce and understand language. NLP takes a more pragmatic perspective and says that we wish to build systems that facilitate some language interface.

*Jason Eisner has a better answer on Quora*
Let your text speak…but ask the right questions

How would a NLP system help scholars with this?

Digital library systems today are still monolithic scalable CMS
• They serve authors: prestige, directed discovery (search)
• And provision appropriate access and provenance: OpenURL, DOI

What about the users? We can do better.
Let’s turn our science loose on our science.
Towards the introspective

Digital Library

Photo Credits: sizumaru @ Flickr
The (Slow) Process of Research

Enthusiasm
Discovering
Reading
Collecting Data
Sensemaking
Comparing
Publishing
Communicating
Maintaining

Title
Authors
Abstract
Introduction
Related Work
Method
Evaluation
References
Ancillary Artifacts
Discovering: Argumentative Zoning

Reading and understanding is hard.

Scientific discoveries are written for peer experts to understand, but not for novices to learn.

Let's use NLP to digest scholarly work to make them easier to understand for beginners.
Summarizing Scholarly Documents

Why not just use the abstract?

The *bā* construction is central to the study of Mandarin grammar. It has received many attempts at analysis and comes up frequently as a syntactic test in discussions of other phenomena. Yet, not even its part of speech has ever been convincingly established. This paper presents the case for treating *bā* as a verb, considering both language-internal arguments and arguments from universal properties of parts of speech. These arguments are intended to have cross-theoretic validity. On the basis of the conclusion that *bā* is a verb, an analysis is developed within the framework of Lexical Functional Grammar. On this analysis, *bā* selects for a subject, an object, and a complement clause, and further stipulates that its object controls the *TOPIC* function of its complement clause. This analysis is shown to account for both the core data and the data which is problematic for other analyses. Finally, the analysis is confirmed by evidence from telicity effects in the *bā* construction, universal properties of verbs and prepositions, and its compatibility with the known historical development of the construction.
Rhetorical Structure Theory

Let’s look at our example

It follows a specific presentation convention

There is a logical document structure

INTRODUCTION

...?

THE BASICS OF THE Bà CONSTRUCTION

...?

ANALYSIS

* Apologies to Mann and Thompson
Abstracting abstracts

These are the “ground truth” for a summary of a single paper

Want to select sentences that overlap significantly with the abstract

But even here there is structure within an abstract
Revisiting Text Rank

Capitalize on the conventional structure in documents

Identify the logical structure of a scientific document

Find the best sentences within the sections of the document

Jun Ping Ng, Praveen Bysani, Ziheng Lin, Min-Yen Kan and Chew Lim Tan (2011) SWING: Exploiting Category-Specific Information for Guided Summarization. In Proceedings of the Text Analysis Conference 2011 (TAC 2011), Gaithersburg, Maryland, USA.
What about **scholarly** documents?

They have references and citations **“citation sentences”**

Often describe a paper from the community’s point of view

A representation of a key point of a work

Citation sentences and in- article sentences have complementary purposes:

Results and evaluations usually not mentioned in citation sentences.

Sentences in a paper that describe its method are usually too detailed for a summary.
<table>
<thead>
<tr>
<th>TITLE</th>
<th>CITED BY</th>
<th>YEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic theory: A formal introduction</td>
<td>1502</td>
<td>1999</td>
</tr>
<tr>
<td>IA Sag, T Wasow, EM Bender, IA Sag</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center for the Study of Language and Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The grammar matrix: An open-source starter-kit for the rapid</td>
<td>275</td>
<td>2002</td>
</tr>
<tr>
<td>development of cross-linguistically consistent broad-coverage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>precision grammars</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM Bender, D Flickinger, S Oepen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proceedings of the 2002 workshop on Grammar engineering and evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficient deep processing of Japanese</td>
<td>142</td>
<td>2002</td>
</tr>
<tr>
<td>M Siegel, EM Bender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proceedings of the 3rd workshop on Asian language resources and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Efficient deep processing of Japanese</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntactic variation and linguistic competence: The case of AAVE</td>
<td>116</td>
<td>2000</td>
</tr>
<tr>
<td>copula absence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM Bender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stanford university</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The syntax of Mandarin Bà: Reconsidering the verbal analysis</td>
<td>90</td>
<td>2000</td>
</tr>
<tr>
<td>E Bender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journal of East Asian Linguistics 9 (2), 105-145</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grammar customization</td>
<td>75</td>
<td>2010</td>
</tr>
<tr>
<td>EM Bender, S Orellishak, A Fokkens, L Poulson, S Saleem</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research on Language and Computation 8 (1), 23-72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapid prototyping of scalable grammars: Towards modularity in</td>
<td>74</td>
<td>2005</td>
</tr>
<tr>
<td>extensions to a language-independent core</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM Bender, D Flickinger</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Companion Volume to the Proceedings of Conference including Posters/Demos</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arboretum: Using a precision grammar for grammar checking in CALL</td>
<td>68</td>
<td>2004</td>
</tr>
<tr>
<td>EM Bender, D Flickinger, S Oepen, A Walsh, T Baldwin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instil/cola symposium 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road-testing the English Resource Grammar Over the British National</td>
<td>67</td>
<td>2004</td>
</tr>
<tr>
<td>Corpus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T Baldwin, EM Bender, D Flickinger, A Kim, S Oepen</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COLING (Santa Fe)</td>
<td>22/38</td>
<td></td>
</tr>
</tbody>
</table>
The context of the citation influences its importance: via its function and referent

- Its location (which section)
- Its syntactic structure

In addition, we proposed a set of new features that used verb class information induced from the frame files of the Chinese PropBank, as well as features that were designed to exploit the grammatical constructions that Semantic Role Labeling of Chinese Predicates are unique to Chinese, specifically the BA (Bender 2000) and BEI (Huang 1999) constructions.

For discussion of the modern Mandarin disposal construction the reader is referred to Li and Thompson (1981), Cheng (1988), Sybesma (1999), Bender (2000), among many others.
Interactions with others play an important role in populating the slow hunch.
Tell me something I don’t know

Collaborative Filtering: “People like you also read…”

Q: Whom do we ask for help?
We attend conferences (like this one) in part to help learn from each other.

A key artifact is the slide presentation, which often summarizes the work in an accessible manner.

But they:
- Miss important technical details

Idea: Use both together
Presentations and their Relationship to Documents

6. EXPERIMENTAL STUDY

In this section, we present an extensive experimental study of TREE S KETCH on real-life and synthetic data sets. Our results verify the effectiveness, in terms of accuracy and construction time, of the TREE S KETCH synopses as structural summaries for large XML data sets. These benefits become even more apparent in a comparison to previously proposed techniques, where TREE S KETCH es perform consistently better in all aspects. Overall, this empirical study indicates that TREE S KETCH es are a viable and effective solution for the structural summarization of large XML data sets in real-world applications.

6.1 Testbed and Methodology

Techniques. We have experimented with two techniques. TREE S KETCH es. We have implemented a fully functional prototype of the TREE S KETCH framework that we describe in this paper. Throughout our experiments, the construction algorithm uses an upper limit of d lex AV operations and rebuilds the heap when its size is reduced below d IMY operations. Twig-KS KETCH es.
Aligning Documents to their Presentations

Better to juxtapose both media together in a fine-grained manner.

Output: an alignment map
Slide Demographics

Prevalence of Slide Types

Thank You!

Agenda
- Problem
- Solution: Conceptual Partitioning Monitoring (CPM)
- Extensions of the Solution
- Performance Analysis
- Conclusion

Sample Tuning

Synopsis Model - Example

XML Data Graph

Synopsis graph

\[ \text{Count}(A) = \text{Extent}(A) \]

\[ \#(A, B, C) \]
Error Analysis: It’s a multimodal problem

<table>
<thead>
<tr>
<th>Slide Type</th>
<th>Common reason</th>
<th>% Incorrectly Aligned by Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nil</td>
<td>Doesn’t know where to align (\rightarrow) align to best fit</td>
<td>64%</td>
</tr>
<tr>
<td>Outline</td>
<td>Name of some sections in it (\rightarrow) align to longest one</td>
<td>36%</td>
</tr>
<tr>
<td>Image</td>
<td>Very little text available</td>
<td>81%</td>
</tr>
<tr>
<td>Drawing</td>
<td>Noisy data: lots of shapes and text boxes</td>
<td>53%</td>
</tr>
<tr>
<td>Table</td>
<td>Little text, noisy data</td>
<td>50%</td>
</tr>
<tr>
<td>Text</td>
<td></td>
<td>24%</td>
</tr>
</tbody>
</table>
Research Slow

Ignore the gradient:
Open up the **Adjacent Possible**

Develops the **Slow Hunch** through persistent exposure to ancillary evidence

*The Swedish Fika*: collision allows scholars to benefit from each other
Not the Eureka Moment
Where Good Ideas Come From

The Adjacent Possible
Liquid Networks
The Slow Hunch
Serendipity
Error
Exaptation
Platforms

Steven Johnson
Fuben-Eki
不便益

FUrther BENEfit of a Kind of Inconvenience

Make key operations deliberate
Fosters ownership, self-affirmation

Key idea: synergize with human computational abilities

Fig. 1. Relations among inconvenience and benefits
Acceleration has it out for us

Others are public advocates, but what about ourselves?

We need to ramp up our research slow

Use the time gifted by research fast

Adapted from Astro Teller's graph
Weapons of Math Destruction

A WMD is

• Massive
• Opaque
• No feedback loop

The Class Break

Cathy O’Neil
Reprise: Research Slow — The Challenge of Story

Be an advocate
Write the story, wonder “what if?”
Blog your opinion or write columns

Stories give context
We are human
Aids association, catalyzes the slow hunch

Ensembles are better than any single classifier
Conclusion: We combine theory and practice

... v4: It works and we think we know why and we’ll advocate for it

Let’s fix that: Research Fast AND Slow
Fast begets fast, but we need to consciously support slow
NLP to do intelligence augmentation (IA), put scholars on the slow hunch
Close the loop: Let data speak and inform our models

Thanks to:

WING members: Bamdad Bahrani Muthu Chandrasekaran Tao Chen
Emily Bender Leon Derczynski Pierre Isabelle Marti Hearst Gertjan van Noord Chris Manning

Xiangnan He Aminesh Prasad Kazunari Sugiyama Aobo Wang Yohei Seki

all those students and faculty at NUS whom I subjected to a dry run
And research slow technology of audiobooks and digital commonplace books