Research Methods in computer science

Spring 2019

Lecture 10

Omprakash Gnawali

February 18, 2019

Agenda

HW4 Live Grading

Paper Review

Conference Organization

Hypothesis and metrics

HW5

Typical Template

Summary

Strengths

Weaknesses

Detailed Comments

Justification for these sections?

We looked at a few examples of paper reviews.

We also looked at paper review software.

I almost never print out papers for review; I prefer to work with the electronic version. I always read the paper sequentially, from start to finish, making comments on the PDF as I go along. I look for specific indicators of research quality, asking myself questions such as: **Are the background literature and study rationale clearly articulated? Do the hypotheses follow logically from previous work? Are the methods robust and well controlled? Are the reported analyses appropriate? (I usually pay close attention to the use—and misuse—of frequentist statistics.)** Is the presentation of results clear and accessible? **To what extent does the Discussion place the findings in a wider context and achieve a balance between interpretation and useful speculation versus tedious waffling?**

sciencemag.org

I subconsciously follow a checklist. First, is it well written? That usually becomes apparent by the Methods section. (Then, throughout, if what I am reading is only partly comprehensible, I do not spend a lot of energy trying to make sense of it, but in my review I will relay the ambiguities to the author.) I should also have a good idea of the hypothesis and context within the first few pages, and it matters whether the hypothesis makes sense or is interesting. Then I read the Methods section very carefully. I do not focus so much on the statistics—a quality journal should have professional statistics review for any accepted manuscript—**but I consider all the other logistics of study design where it's easy to hide a fatal flaw**. Mostly I am concerned with credibility: Could this methodology have answered their question? Then I look at how convincing the results are and how careful the description is. Sloppiness anywhere makes me worry. **The parts of the Discussion I focus on most are context and whether the authors make claims that overreach the data. This is done all the time, to varying degrees. I want statements of fact, not opinion or speculation, backed up by data.**

- ***Michael Callaham***, *emergency care physician and researcher at the University of California, San Francisco*

sciencemag.org

Most journals don't have special instructions, so I just read the paper, usually starting with the Abstract, looking at the figures, and then reading the paper in a linear fashion. I read the digital version with an open word processing file, keeping a list of “major items” and “minor items” and making notes as I go. There are a few aspects that I make sure to address, though I cover a lot more ground as well. First, I consider how the question being addressed fits into the current status of our knowledge. Second, **I ponder how well the work that was conducted actually addresses the central question posed in the paper. (In my field, authors are under pressure to broadly sell their work, and it's my job as a reviewer to address the validity of such claims.)** Third, I make sure that the design of the methods and analyses are appropriate.
- *McGlynn*

sciencemag.org

First, I read a printed version to get an overall impression. What is the paper about? How is it structured? I also pay attention to the schemes and figures; if they are well designed and organized, then in most cases the entire paper has also been carefully thought out.

When diving in deeper, first I try to assess whether all the important papers are cited in the references, as that also often correlates with the quality of the manuscript itself. Then, right in the Introduction, you can often recognize whether the authors considered the full context of their topic. After that, I check whether all the experiments and data make sense, paying particular attention to whether the authors carefully designed and performed the experiments and whether they analyzed and interpreted the results in a comprehensible way**. It is also very important that the authors guide you through the whole article and explain every table, every figure, and every scheme.**

As I go along, I use a highlighter and other pens, so the manuscript is usually colorful after I read it. Besides that, I make notes on an extra sheet.
- ***Melanie Kim Müller***, *doctoral candidate in organic chemistry at the Technical University of Kaiserslautern in Germany*

sciencemag.org

I first familiarize myself with the manuscript and read relevant snippets of the literature to make sure that the manuscript is coherent with the larger scientific domain. Then I scrutinize it section by section, noting if there are any missing links in the story and if certain points are **under- or overrepresented**. **I also scout for inconsistencies in the portrayal of facts and observations**, assess whether the exact technical specifications of the study materials and equipment are described, consider the adequacy of the sample size and the quality of the figures, and assess whether the findings in the main manuscript are aptly supplemented by the supplementary section and whether the authors have followed the journal's submission guidelines.
- ***Chaitanya Giri***, *postdoctoral research fellow at the Earth-Life Science Institute in Tokyo*

sciencemag.org

I spend a fair amount of time looking at the figures. In addition to considering their overall quality, sometimes figures raise questions about the methods used to collect or analyze the data, or they fail to support a finding reported in the paper and warrant further clarification. I also want to know whether the authors' conclusions are adequately supported by the results. Conclusions that are overstated or out of sync with the findings will adversely impact my review and recommendations.

- ***Dana Boatman-Reich****, professor of neurology and otolaryngology at Johns Hopkins University School of Medicine in Baltimore, Maryland*

sciencemag.org

I generally read on the computer and start with the Abstract to get an initial impression. Then I read the paper as a whole, thoroughly and from beginning to end, taking notes as I read. For me, the first question is this: Is the research sound? And secondly, how can it be improved? Basically, I am looking to see if the research question is well motivated; if the data are sound; if the analyses are technically correct; and, **most importantly, if the findings support the claims made in the paper.**

\- *Walsh*

sciencemag.org

I generally read on the computer and start with the Abstract to get an initial impression. Then I read the paper as a whole, thoroughly and from beginning to end, taking notes as I read. For me, the first question is this: Is the research sound? And secondly, how can it be improved? Basically, I am looking to see if the research question is well motivated; if the data are sound; if the analyses are technically correct; and, **most importantly, if the findings support the claims made in the paper.**

- *Walsh*

sciencemag.org

The main aspects I consider are the novelty of the article and its impact on the field. I always ask myself what makes this paper relevant and what new advance or contribution the paper represents. Then I follow a routine that will help me evaluate this. First, I check the authors' publication records in PubMed to get a feel for their expertise in the field. I also consider whether the article contains a good Introduction and description of the state of the art, as that indirectly shows whether the authors have a good knowledge of the field. Second, I pay attention to the results and whether they have been compared with other similar published studies. Third, **I consider whether the results or the proposed methodology have some potential broader applicability or relevance, because in my opinion this is important. Finally, I evaluate whether the methodology used is appropriate. If the authors have presented a new tool or software, I will test it in detail.**

*- **Fátima Al-Shahrour**, head of the Translational Bioinformatics Unit in the clinical research program at the Spanish National Cancer Research Centre in Madrid*

sciencemag.org

Some Questionable Ideas…

- A paper addresses a problem that will become important?
- A paper will become important to a community?
- A paper has lots of graphs and data
- The authors did a lot of experiments

Some Easier Ideas…

- Technical Claims
  - What are the claims?
  - Have the claims been validated?
    - Experimental
    - Data
    - Theory
- Technical Correctness

Activities in a Conference

Keynote

Paper presentations

Panels

Poster and demo

Competitions

Open mic sessions

Conference Organization

- Different roles
  - General Chair
  - Finance Chair
  - Arrangement Chair
  - Technology Chair
  - Program Chair
  - Publication Chair
  - Technical Program Committee
  - Many other roles
- Schedule for activities

Technical Program Committee

Review papers

Types of discussions and meetings

Future of CS Publications

- Virtual Conferences
- Blogs?
- Tweets?
- Open Access
- Hybrid Conference/Journal

We formed the organization and technical program committee for the conference. We also decided tentative schedule for the conference.

Hypothesis

“A hypothesis (plural hypotheses) is a proposed explanation for a phenomenon. For a hypothesis to be a scientific hypothesis, the scientific method requires that one can test it. Scientists generally base scientific hypotheses on previous observations that cannot satisfactorily be explained with the available scientific theories. Even though the words "hypothesis" and "theory" are often used synonymously, a scientific hypothesis is not the same as a scientific theory. A working hypothesis is a provisionally accepted hypothesis proposed for further research, in a process beginning with an educated guess or thought.”

-- wikipedia

Hypothesis in Engineering

The hypothesis-model is good for research where you want to understand how something works, but I think it is ill-suited for capturing the full scope of engineering research. After all, in engineering, you're primary goal is not to learn something about how the world works, but rather to **change how the world works**! So, instead of a hypothesis on how something works, I'd put up existing gaps in the ability to do something as a working basis. That will then put a focus on your research result as an extension of technical capabilities. In order to evaluate your research results, you would then have to show that your results actually close the existing gap.

Of course, also engineering research needs to understand something about how existing things work in order to be able to create something new. Hypotheses **are suitable** in engineering to clarify these preliminary things. In your case, you state that *"the existing 'role-based access control' of MS-Windows does not solve some problems"* - that sounds like a perfect hypothesis to test for. But verifying this hypothesis is certainly not the key step in your research, and maybe it has already been done previously. That's why I'd recommend not to focus on a hypothesis as the basis for engineering research (though one might use them to clarify preliminaries), but focus on **identified gaps in current technical abilities**.

-- silvado, Aug 26, 2013 on stackexchange

https://academia.stackexchange.com/questions/12156/hypothesis-for-an-engineering-oriented-research-thesis

Hypothesis in Engineering

Effectively, what you are doing is *development* of existing research, rather than designing something *de novo*. The notion of a research hypothesis is therefore somewhat inappropriate to such work, and you wouldn't write a paper describing this work specifying a definitive "hypothesis."

Instead, you'd write the paper emphasizing that your model does something "better," "faster," "more securely," or specifying whatever other accomplishments advance your work from the previous state of affairs. Your thesis should then show how that is accomplished, and give some evidence thereof.

-- aeismail, Aug 25, 2013

https://academia.stackexchange.com/questions/12156/hypothesis-for-an-engineering-oriented-research-thesis

Hypothesis and Engineering Thesis

Because engineers invent rather than discover, does an engineering thesis need a hypothesis?... because invention is a more tightly directed activity than discovery; and the two are not mutually exclusive anyway…uppose your project involves using Artificial Neural Networks (ANNs), in conjunction with appropriate hardware, to sort good apples from bad. The hypothesis for this project may be, 'It is possible to sort good apples from bad using ANNs and suitable hardware'…. Suppose that on completing your project, you discovered that the system you had devised works well with green apples, but not with red ones. You would have discovered new knowledge and would be able to suggest a revised hypothesis as the starting point for further investigation. Your own project would have demonstrated the correctness of a hypothesis like 'It is possible to sort good green apples from bad green apples, with an accuracy of better than 90%, using ANNs and suitable hardware'.

http://thesishub.org/does-an-engineering-thesis-need-a-hypothesis/

Sample Hypotheses

Only an extraordinarily skilled attacker can break into our firewall. [?]

The firewall accepts all well-formed packets and sessions, and handles malformed packets and sessions as documented in the firewall’s manual.

From Sean Peisert and Matt Bishop

Types of Experiments

Model / Analysis

Simulations

Testbed (Real word ^{lite})

"Real world"

Which one to use when?

Metric

Why do we want to measure?

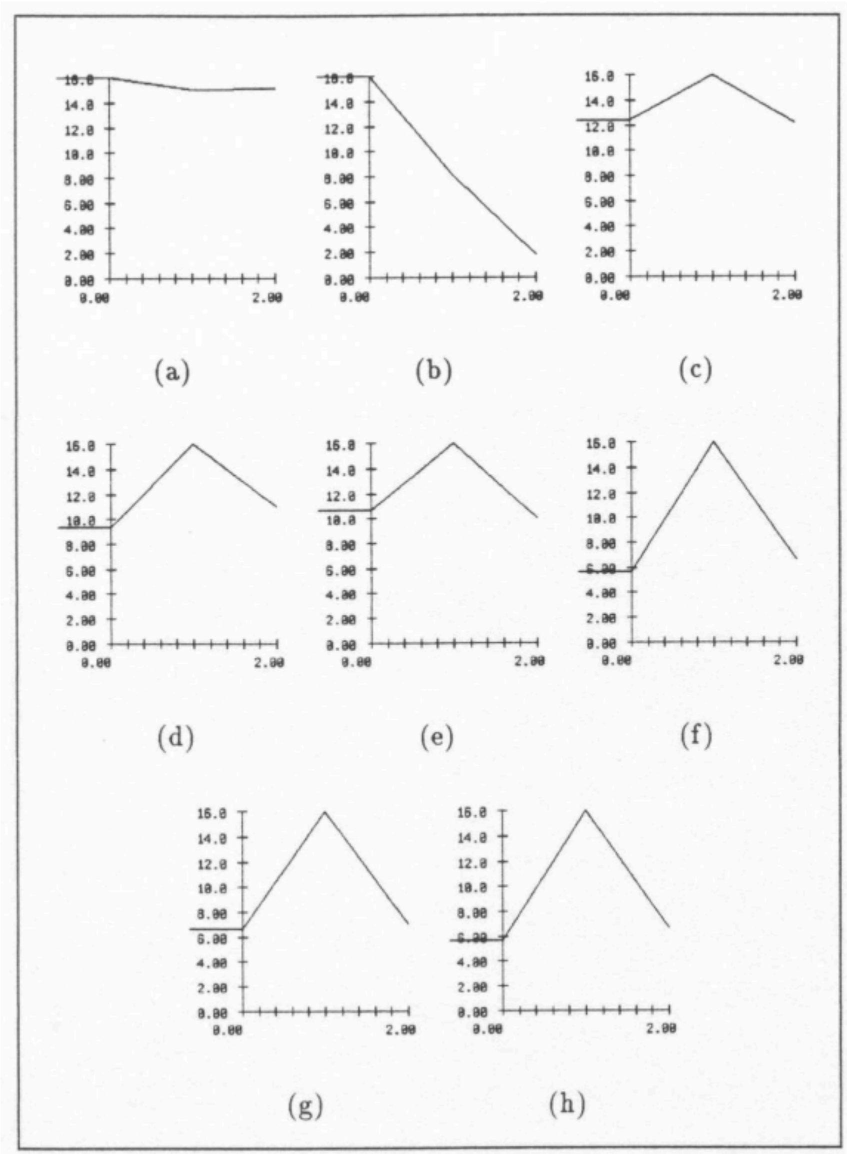
What to measure?

Most of the time we measure improvements

Eigenfaces for Recognition

[Turk '91]

"We have developed a near-real-time computer system that can locate and track a subject's head, and then recognize the person by comparing the characteristics of the face to those of known individuals."



Scenarios and metrics from [Turk '91]

Figure 9. Results of experiments measuring recognition performance using eigenfaces. Each graph shows averaged performance as the lighting conditions, head size, and head orientation vary—the *y*-axis depicts number of correct classifications (out of 16). The peak (16/16 correct) in each graph results from recognizing the particular training set perfectly. The other two graph points reveal the decline in performance as the following parameters are varied: **(a)** lighting, **(b)** head size (scale), **(c)** orientation, **(d)** orientation and lighting, **(e)** orientation and size (#1), **(f)** orientation and size (#2), **(g)** size and lighting, **(h)** size and lighting (#2).

The Anatomy of a Large-Scale Hypertextual Web Search Engine

[Brin and Page ’98]

What hypothesis, scenarios, and metrics should we expect to see in this paper?

5 Results and Performance

The most important measure of a search engine is the quality of its search results. While a complete user evaluation is beyond the scope of this paper, our own experience with Google has shown it to produce better results than the major commercial search engines for most searches. As an example which illustrates the use of PageRank, anchor text, and proximity, Figure 4 shows Google's results for a search on "bill clinton". These results demonstrates some of Google's features. The results are clustered by server. This helps considerably when sifting through result sets. A number of results are from the whitehouse.gov domain which is what one may reasonably expect from such a search. Currently, most major commercial search engines do not return any results from whitehouse.gov, much less the right ones. Notice that there is no title for the first result. This is because it was not crawled. Instead, Google relied on anchor text to determine this was a good answer to the query. Similarly, the fifth result is an email address which, of course, is not crawlable. It is also a result of anchor text.

All of the results are reasonably high quality pages and, at last check, none were broken links. This is largely because they all have high PageRank. The PageRanks are the percentages in red along with bar graphs. Finally, there are no results about a Bill other than Clinton or about a Clinton other than Bill. This is because we place heavy importance on the proximity of word occurrences. Of course a true test of the quality of a search engine would involve an extensive user study or results analysis which we do not have room for here. Instead, we invite the reader to try Google for themselves at http://google.stanford.edu.

Query: bill clinton

http://www.whitehouse.gov/
100.00% (no date) (0K)
http://www.whitehouse.gov/
    Office of the President
    99.67% (Dec 23 1996) (2K)
    http://www.whitehouse.gov/WH/EOP/OP/html/OP_Home.html
    Welcome To The White House
    99.98% (Nov 09 1997) (5K)
    http://www.whitehouse.gov/WH/Welcome.html
    Send Electronic Mail to the President
    99.86% (Jul 14 1997) (5K)
    http://www.whitehouse.gov/WH/Mail/html/Mail_President.html

mailto:president@whitehouse.gov
99.98%
    mailto:President@whitehouse.gov
    99.27%
The "Unofficial" Bill Clinton
94.06% (Nov 11 1997) (14K)
http://zpub.com/un/un-bc.html
    Bill Clinton Meets The Shrinks
    86.27% (Jun 29 1997) (63K)
    http://zpub.com/un/un-bc9.html
President Bill Clinton - The Dark Side
97.27% (Nov 10 1997) (15K)
http://www.realchange.org/clinton.htm
$3 Bill Clinton
94.73% (no date) (4K)
http://www.gatewy.net/~tjohnson/clinton1.html

Figure 4. Sample Results from Google

[Brin and Page '98]

Storage Statistics	
Total Size of Fetched Pages	147.8 GB
Compressed Repository	53.5 GB
Short Inverted Index	4.1 GB
Full Inverted Index	37.2 GB
Lexicon	293 MB
Temporary Anchor Data (not in total)	6.6 GB
Document Index Incl. Variable Width Data	9.7 GB
Links Database	3.9 GB
Total Without Repository	**55.2 GB**
Total With Repository	**108.7 GB**

Web Page Statistics	
Number of Web Pages Fetched	24 million
Number of Urls Seen	76.5 million
Number of Email Addresses	1.7 million
Number of 404's	1.6 million

Table 1. Statistics

[Brin and Page '98]

Why did the authors decide to report these measurements?

Metrics/Experiments?

Accurately Initializing Real Time Clocks to Provide Synchronized Time in Sensor Networks

CTP: An Efficient, Robust, and Reliable Collection Tree Protocol for Wireless Sensor Networks

On the Effectiveness of Energy Metering on Every Node

Surviving Sensor Network Software Faults

Metrics from Classification Research

Classification Accuracy
Logarithmic Loss
Area Under ROC Curve
Confusion Matrix
Classification Report
Precision
Recall
F1-Score

Partly from https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/

Metrics from Regression Research

Mean Absolute Error

Mean Squared Error

R^2

Partly from https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/

Metrics from Systems Research

Reliability

Latency

Coverage

Energy

HW5 – Metrics

List of metrics from the related papers.

Define the metrics.

Observations about common and uncommon metrics.