Software Engineering: Research Overview, Papers, Proposals and Presentations

Yu Huang
Vanderbilt University
yu.huang@vanderbilt.edu

Where can you find the information?

- Course website: **https://huang.isis.vanderbilt.edu/cs8395**

  You can find (almost) everything on the course website!

- Demo

What is Software Engineering Research?

- “My favorite operational definition of engineering is '**design under constraint**.' Engineering is creating, designing what can be, but it is constrained by nature, by cost, by concerns of safety, reliability, environmental impact, manufacturability, maintainability, and many other such 'ilities.'”

[Bill Wulf, NAE President, The Urgency of Engineering Education Reform, 2008]

Everything that is involved in the entire process!

- “[Software Engineering is] The Establishment and use of sound **engineering principles** in order to obtain **economically** software that is **reliable** and works **efficiently** on **real** machines.”

[Fritz Bauer 1975, S. 524]

What is Software Engineering Research?

- Where are the good work published?
  - https://csrankings.org/ ---> Is any ranking fair?
  - **ICSE**
    - International Conference on software Engineering
  - **FSE** (ESEC/*FSE*):
    - The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering
  - **ASE**:
    - International Conference on Automated Software Engineering
  - ISSTA, ESEM, ICPC, ISSRE, MSR, etc.

What is Software Engineering Research?

- Where are the good work published?
  - **ICSE**
    - International Conference on software Engineering

Technical Track

About | ICSE 2023 open science policy | Q/A

Call for Papers

Research of Interest

ICSE welcomes submissions addressing topics across the full spectrum of Software Engineering, being inclusive of quantitative, qualitative, and mixed-methods research.

Topics of interest include:

- API design and evolution
- Apps and app store analysis
- Autonomic systems and self adaptation
- Configuration management
- Crowd-based software engineering
- Debugging and fault localization
- Design for quality, incl. privacy and security by design
- Distributed and collaborative software engineering
- Diversity, inclusion, fairness of software
- Embedded and cyber-physical systems
- Ethics in software engineering
- Evolution and maintenance
- Feedback, user, and requirements management
- Formal methods
- Green and sustainable technologies
- Human aspects of software engineering
- Human-computer interaction
- Legal aspects of software engineering
- Machine learning with and for SE
- Mining software repositories
- Model checking
- Modeling and model-driven engineering
- Parallel and distributed systems
- Performance analysis and testing
- Privacy and security
- Program analysis
- Program comprehension
- Program repair
- Program synthesis
- Programming languages
- Recommender systems
- Refactoring
- Release engineering and DevOps
- Reliability and safety
- Requirements engineering
- Reverse engineering
- SE for machine learning
- Search-based software engineering
- Software architecture and product design
- Software economics

Core Course Topics

- Program analysis
  - Is a program correct? Does a program have certain property?
- Program testing and repair
  - How to effectively conduct this boring but extremely important task?
- Software security
  - Vulnerability in programs
- Human factors in SE
  - SE is a human activity, what can go wrong?
- AI for SE
  - Leverage AI tools to automate SE tasks
- Open source software
  - Global collaboration, diverse applications, unique challenges

Program Analysis

Operate on the programs

- The systematic examination of a program to determine its properties
  - Is my program correct?
  - Where is the bug?
  - What does a program do (without running it)?
  - How to prove theorems about the behavior of a program?
  - ...
- Why should I care?
  - Automatic testing and bug finding
  - Language design and implementations (compilers, VMs)
  - Program transformation (optimization, repair)
  - Program synthesis

Program Analysis

Operate on the programs

- What issues can you find using program analysis?
  - Defects that result from inconsistently following simple design rules
    - Security: Buffer overruns, improperly validated input
    - Memory safety: Null Pointer Dereference, uninitialized data
    - Resource leaks: Memory, OS resources
    - API protocols: Device drivers, GUI frameworks
    - Exceptions: Arithmetic/library/user-defined
    - Encapsulation: Accessing internal data, calling private functions
    - Data races: Two threads access the same data without synchronization

Check compliance to simple, mechanical design rules

Testing and Repair

- Program testing: ***This is dumb, but let's just find some stuff!***
  - Run your programs again and again, wait, the output doesn't look right! Wait, my program crashed!...
- Program Repair: ***Now I found some stuff, then what?!***
  - Challenge #1: what is wrong?? Where should I check?? -> **Fault Localization**
  - Challenge #2: This line is super suspicious!! How should change it? -> **Patch Generation**

What about distributed systems? Concurrent programs? Multi-threaded programs?

Software Security

- Technically, this is not a bug (i.e., functionality bug). The developer doesn’t intend the system to do something. The attacker who finds a vulnerability is using something that wasn't modeled by the developer in the first place.
  - Buffer overrun
  - DoS: Denial-of-Service
  - ROP: Return-oriented Programming
- "Vulnerability"
- How can we find these vulnerabilities before attackers find them?
  - Program analysis, testing

Human Factors

- Human is an important component in SE (if not the most important)!
- They conduct almost all the activities
  - Requirement
  - Design
  - Writing code
  - Reading code
  - Code review
  - Testing
  - Fixing bugs
  - Educating next generation programmers
  - …
- Understand how they conduct them, what is important, and improve them
  - Productivity
  - Effectiveness
  - Diversity
  - Sustainability
- Use methods in CS, Psych, CogSci, NeuroSci, Social Science, etc., to understand it; use CS/engieering/education/AffectiveComputing/... to improve it (interventions)

AI and SE

- AI4SE
  - How to leverage AI tools to improve productivity in SE?
    - Automation
    - Complicated tasks
    - Treating programs as text?
  - Domain knowledge
    - Program is in between math/logic and natural languages
  - Human-guided AI for SE
- SE4AI
  - Treating AI system/model design/development as a special software development process
    - Does Agile work? How to test it? DevOps for AI system?

Open Source Software

- + Human Factors
  - Global collaborations
  - Sustainability issue
  - Challenge of open science
- + program analysis, testing, repair, security, etc.
- + AI
- + Software Security
- How do we serve non-engineering/CS applications?

How to read a scientific paper?

The Myth of Double-Blind Review Revisited: ACL vs. EMNLP

Cornelia Caragea[1], Ana Sabina Uban[2], and Liviu P. Dinu[2]
[1]Computer Science, University of Illinois at Chicago
Chicago, Illinois
[2]Faculty of Mathematics and Computer Science, University of Bucharest
Bucharest, Romania
cornelia@uic.edu, ana.uban@gmail.com, ldinu@fmi.unibuc.ro

Abstract

The review and selection process for scientific paper publication is essential for the quality of scholarly publications in a scientific field. The double-blind review system, which enforces author anonymity during the review period, is widely used by prestigious conferences and journals to ensure the integrity of this process. Although the notion of anonymity in the double-blind review has been questioned before, the availability of full text paper collec-

conferences in order to guarantee *fairness* of the paper selection, and thus, plays an essential role in how scientific quality is eventually measured (Meadows, 1998). It is designed to reduce the risk of bias in paper reviews, ensuring that all papers are judged solely based on their content and intrinsic quality and that any author has a fair chance of having a paper accepted, regardless of their prestige or previous work. The double-blind review process implies that the submitted papers have to be anonymized, i.e., the authors' names are not ex-

What is a paper?

- A **documented** presentation of **new knowledge** that is placed **in context**
  - A **convincing logical argument** supported by **evidence**
- A **conversation** with a community of **peers**
- **Currency** in academic circles

Why papers? How to read them?

- Theory of knowledge (epistemology)
  - Investigating what distinguishes “justified belief” from “opinion”.
- Philosophy of Science
- Peer review process
- Structure of a paper

Knowing requires specification

- Specification is **hard**

- **Written** communication (complete with citations for context) help build a **trace of knowledge** established over history
  - You can trace recent NLP work back to computational linguistics
  - You can trace addition back to *Principia Mathematica*
  - *You can trace the newest program repair work to the original GenProg*

What is knowing?

- **Knowledge = Justified true beliefs**
- Belief
  - Trust in some concept or entity
  - *“I believe the earth is flat.”*
  - *“I believe the earth is round.”*
- Truth
  - Agreeing with reality
  - *The earth behaves as though it were round*
- Justification
  - Truth with account
  - *Gravity, motion of celestial bodies, horizon effect, satellite photography, etc.* **justify** *one’s belief that the earth is round;* **falsify** *one’s belief that the earth is flat.*

Reason

- Photos of the earth from orbit make it appear round, therefore I believe the earth is round

- Reducing calorie intake has led others to become thinner, therefore I can consume fewer to become thinner

Science: Gaining knowledge and reason

- Science is a **method** for acquiring new **knowledge by applying reason!**
  - Identify a question
  - Think real hard…
  - Develop a hypothesis
  - Test the hypothesis to justify a conclusion
  - Have **peers** test the hypothesis and conclusion
  - Apply **reason**: If *Hypothesis* is true, then *prediction of data is* true
- **Methodological naturalism**
  - Scientists assume that all causes are empirical and naturalistic, which means they can be measured, quantified and studied methodically.
  - Anyone can redo the experiment -> replicable

Scientific Method

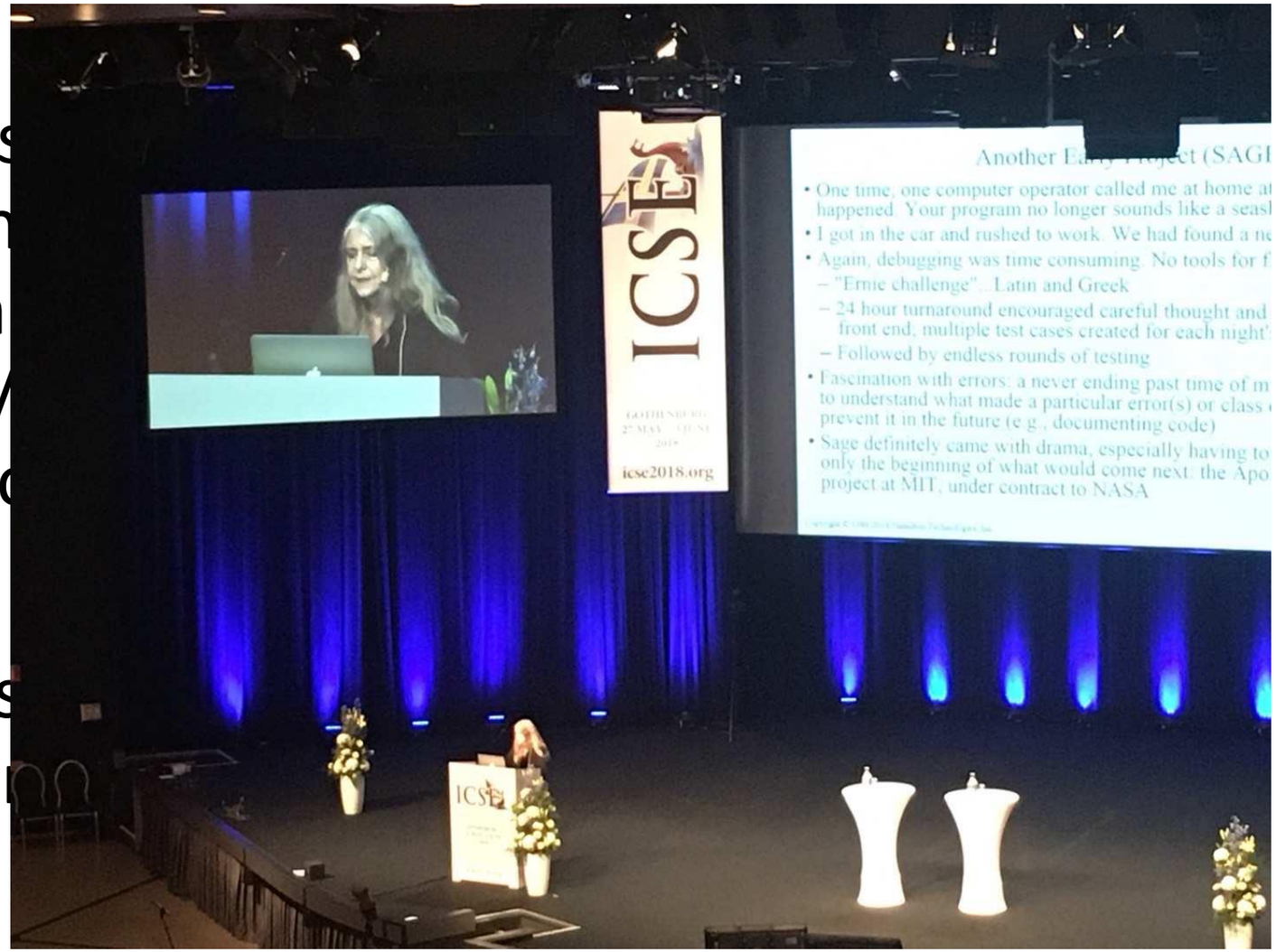
- **Knowledge** acquired by **Science** enables **prediction** of future data
  - We *predict* QuickSort will work better than BubbleSort, controlled experiment demonstrates better runtime performance
- We **design** controlled **experiments** whose **data is attributed** to our new **knowledge**
  - e.g., citations from an anonymous paper can be used to predict authorship information
- *Appropriate* Metrics

"You can't improve what you don't measure" – Bill Wulf, NAE

Trivia:

- This computer scientist, system engineer, and business owner, was director of the Software Engineering Division of the MIT Instrumentation Lab, which developed on-board flight software for NASA's Apollo program. This computer scientist is one of the people credited with coning the term "Software Engineering".
- In Apollo 11 Mission, this computer scientist's on-board flight software averted an abort of the landing on the moon.

Trivia: **Margaret Hamilton**

- 
- 

Peer Review and Paper Reading

- A **paper** represents a chunk of new knowledge
  - It is documentation that someone already solved a problem or answered a question (i.e., you don't have to again—just read the paper!)

- **Skepticism** must be applied to **ensure** humanity has acquired **new knowledge**

- **Peer review** is when others apply **skepticism** to another's claim to knowledge (e.g., a paper)

Reading a Paper: A giant logical argument (story)

The review and selection process for scientific paper publication is essential for the quality of scholarly publications in a scientific field.

The double-blind review system, which enforces author anonymity during the review period, is widely used by prestigious conferences and journals to ensure the integrity of this process. Although the notion of anonymity in the double-blind review has been questioned be-fore, the availability of full text paper collections brings new opportunities for exploring the question: Is the double-blind review process really double-blind? We study this question on the ACL and EMNLP paper collections and present an analysis on how well deep learning techniques can infer the authors of a paper. Specifically, we explore Convolutional Neural Networks trained on various aspects of a paper, e.g., content, style features, and references, to understand the extent to which we can infer the authors of a paper and what aspects contribute the most. Our results show that the authors of a paper can be inferred with accuracy as high as 87% on ACL and 78% on EMNLP for the top 100 most prolific authors.

Reading a Paper: A giant logical argument (story)

The review and selection process for scientific paper publication is essential for the quality of scholarly publications in a scientific field.

The double-blind review system, which enforces author anonymity during the review period, is widely used by prestigious conferences and journals to ensure the integrity of this process.

Although the notion of anonymity in the double-blind review has been questioned be-fore, the availability of full text paper collections brings new opportunities for exploring the question: Is the double-blind review process really double-blind? We study this question on the ACL and EMNLP paper collections and present an analysis on how well deep learning techniques can infer the authors of a paper. Specifically, we explore Convolutional Neural Networks trained on various aspects of a paper, e.g., content, style features, and references, to understand the extent to which we can infer the authors of a paper and what aspects contribute the most. Our results show that the authors of a paper can be inferred with accuracy as high as 87% on ACL and 78% on EMNLP for the top 100 most prolific authors.

Reading a Paper: A giant logical argument (story)

The review and selection process for scientific paper publication is essential for the quality of scholarly publications in a scientific field. The double-blind review system, which enforces author anonymity during the review period, is widely used by prestigious conferences and journals to ensure the integrity of this process.

Although the notion of anonymity in the double-blind review has been questioned before, the availability of full text paper collections brings new opportunities for exploring the question: Is the double-blind review process really double-blind?

We study this question on the ACL and EMNLP paper collections and present an analysis on how well deep learning techniques can infer the authors of a paper. Specifically, we explore Convolutional Neural Networks trained on various aspects of a paper, e.g., content, style features, and references, to understand the extent to which we can infer the authors of a paper and what aspects contribute the most. Our results show that the authors of a paper can be inferred with accuracy as high as 87% on ACL and 78% on EMNLP for the top 100 most prolific authors.

Reading a Paper: A giant logical argument (story)

The review and selection process for scientific paper publication is essential for the quality of scholarly publications in a scientific field. The double-blind review system, which enforces author anonymity during the review period, is widely used by prestigious conferences and journals to ensure the integrity of this process. Although the notion of anonymity in the double-blind review has been questioned be-fore, the availability of full text paper collections brings new opportunities for exploring the question: Is the double-blind review process really double-blind?

We study this question on the ACL and EMNLP paper collections and present an analysis on how well deep learning techniques can infer the authors of a paper. Specifically, we explore Convolutional Neural Networks trained on various aspects of a paper, e.g., content, style features, and references, to understand the extent to which we can infer the authors of a paper and what aspects contribute the most.

Our results show that the authors of a paper can be inferred with accuracy as high as 87% on ACL and 78% on EMNLP for the top 100 most prolific authors.

Reading a Paper: A giant logical argument (story)

The **review and selection process** for scientific paper publication is **essential** for the **quality of scholarly publications** in a scientific field. **The double-blind review system**, which enforces author anonymity during the review period, **is widely used** by prestigious conferences and journals **to ensure the integrity of this process**. Although the notion of anonymity in the double-blind review has been questioned be-fore, the availability of full text paper collections brings new opportunities for exploring the question: **Is the double-blind review process really double-blind?** We study this question on the ACL and EMNLP paper collections and present an analysis on how well deep learning techniques can infer the authors of a paper. Specifically, we explore Convolutional Neural Networks trained on various aspects of a paper, e.g., content, style features, and references, to understand the extent to which we can infer the authors of a paper and what aspects contribute the most.

Our results show that the authors of a paper can be inferred with accuracy as high as 87% on ACL and 78% on EMNLP for the top 100 most prolific authors.

Reading a Paper: A giant logical argument (story)

The **review and selection process** for scientific paper publication is **essential** for the **quality of scholarly publications** in a scientific field. **The double-blind review system**, which enforces author anonymity during the review period, **is widely used** by prestigious conferences and journals **to ensure the integrity of this process**. Although the notion of anonymity in the double-blind review has been questioned be-fore, the availability of full text paper collections brings new opportunities for exploring the question: **Is the double-blind review process really double-blind?** We study this question on the ACL and EMNLP paper collections and present an analysis

Implicit Argument: If authorship can be inferred with high accuracy, the answer is "**no**."

contribute the most.

Our results show that the authors of a paper can be inferred with accuracy as high as 87% on ACL and 78% on EMNLP for the top 100 most prolific authors.

Reading a Paper: Structure

- **Academics** are **old** and obsessed with **ritual**
- Papers in a field follow a structure and best practices
- In computing fields: (the order I read them)
  - Abstract (1)
  - Introduction (2)
  - Background/Related Work (2.5, if needed)
  - Method/Study Design (4)
  - Experiments/Results (3)
  - (sometimes) Threats to Validity (5)
  - (or not at all) Conclusion (2)

Heilmeier Catechism

- George Heilmeier, former DARPA director

- (1) What are you doing?
- (2) How is it done today, and what are the limits?
- (3) What’s the novelty?
- (4) What are the risks and benefits?
- (5) What are metrics for success?

Reading a Paper: Questions to Answer (as a reviewer)

- (0): Does it make sense?
  - Your credibility as a reviewer is at stake…. It's safer to reject if you don't get it

- Example: "The Myth of Double-Blind…"
  - The abstract tells us they are claiming authorship is predictable from paper text and citations (i.e., without explicit author names)
  - This "makes sense" in that, if they can do that, it impacts the soundness of double-blind review (e.g., there is a logically valid argument)

  - (btw it's also readable English – in fact, if it is bad English, I will directly reject it)

Reading a Paper: Questions to Answer

- (1) What are the authors doing?
- Can you understand (and believe) the problem?
  - A problem "should" be well-motivated
  - Is it *actually* **worth** effort to **know** about the paper's **solution** or answer?
    - **Your time** is valuable… you can't read *every* paper out there

- Example: "The Myth of Double-Blind…"
  - Is it really "urgent" that we care about double-blind?
    - **Yes**—double blind is expensive to implement and maintain during review
      - We shouldn't do it if it doesn't actually obscure author identity…
    - The integrity of science and peer review is at stake!

Reading a Paper: Questions to Answer

- (2) What existing techniques are there, and what are their limitations?
  - The paper claims new knowledge—why hasn't it been known before?
    - Implicitly: "If this problem is so important, shouldn't it have been solved already?"

- Example: "The Myth of Double-Blind…"
  - Authorship prediction studies have been **done before**, but **not** with large amounts of **textual data** (*only* on citation patterns and references)
  - Existing techniques get pretty low accuracy (60% on top 10% authors)

- Side note: "60% on top 10% authors" is **not directly comparable** to abstract: "87% and 78% on top 100 authors" (i.e., how much is 10% vs. 100?)

Reading a Paper: Questions to Answer

- (3) What's the novelty?
  - If there's a limitation in current approaches, what *insight* is brought to bear?
  - What's the magic sauce that makes the problem solvable now?

- Example: "The Myth of Double-Blind…"
  - Large archives of **paper text** are readily **available** for analysis
    - (previously, only the citations were easily tabulated)
  - Use CNN model ("more advanced" according to authors)

- Side note: The new insight is about new data becoming available
  - Implicitly, the authors **hypothesize** that **new data** will **change** the outcome of the **question** (i.e., the new data allows better prediction of authorship)

Reading a Paper: Questions to Answer

- (4) What are the risks and benefits?
  - Tradeoffs are inevitable, there's no free lunch, etc.
  - Does the approach solve the problem better?

- Example: "The Myth of Double-Blind…"
  - Implicitly, authors argue their approach is higher accuracy at identifying authors (benefit)
  - Risk: more expensive, data intensive, etc.

- Side note: the benefit seems fishy to me:
  - Overly-qualified… top 100 authors => 87 or 78%
  - Old work was 60% on top 10%
    - I bet top 10% > 100, so the problem is harder

pects contribute the most. Our results show that the authors of a paper can be inferred with accuracy as high as 87% on ACL and 78% on EMNLP for the top 100 most prolific authors.

Reading a Paper: Questions to Answer

- (5) What are the metrics for success?
  - How do you know the authors won?
  - Do you believe them?

- Example: “The Myth of Double-Blind…”
  - Accuracy of identifying authorship
  - Features that are most relevant

- Side note: There are others:
  - Runtime performance (it is useless if it takes months to run)
  - “Accuracy” is underdefined (what if there are multiple authors? What about ordering? How about institutional affiliation?)

Reading a Paper: Heilmeier Summary

- Heilmeier's Questions help give a general picture.
  - A good paper has an introduction, abstract, and conclusion that answer them all
- Next, we need more **scrutiny**. Do the knowledge claims hold water?
- Example: "The Myth of Double-Blind…"
  - The authorship problem has been examined many times over… double-blind prevails as an important technique
  - I am **skeptical** of the 87/78% numbers for top 100 authors. This number seems really qualified… How is it computed?
    - Moreover, is it actually surprising that we can find the most prolific authors' papers if anonymized?

Reading a Paper: More depth

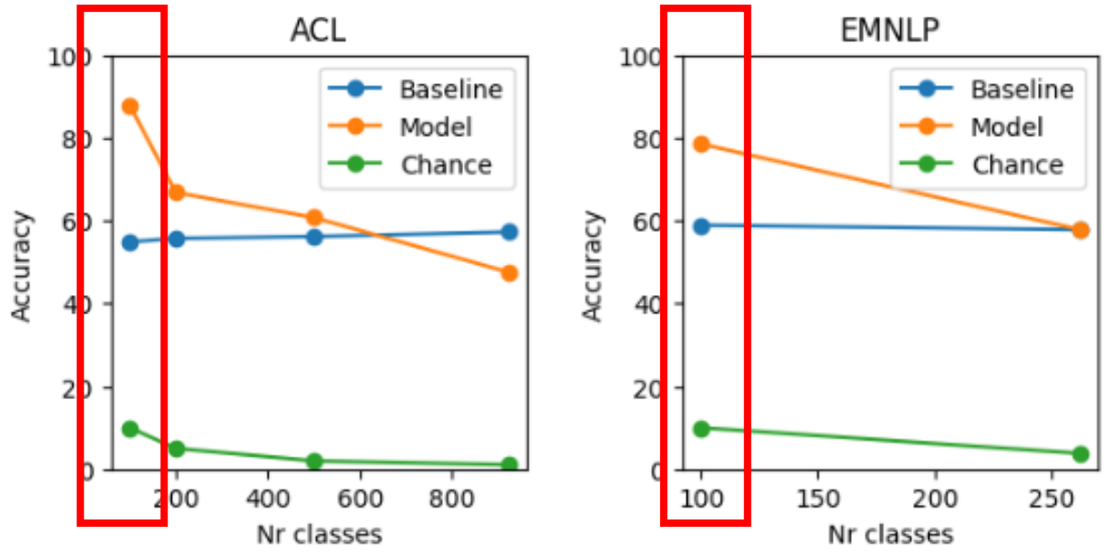
- Do the claims hold up? Do the experiments support conclusions (constructive validity)?

- Example: “The Myth of Double-Blind…”

- Accuracy@k - computed as the number of articles for which at least one true author was in the top k predicted authors. We use $k = 10$ as this number was shown to perform well in other search and retrieval tasks (Spink and Jansen, 2004).

  So you can guess 1/10 authors right and get a point?
  Is the k the same in related work? (“baseline”)
  Doesn’t this inflate numbers at high Nr?

They reported these in the abstract!



Reading a Paper: More depth

- Are findings statistically rigorous?

the total number of articles authored by them. An interesting finding is that the correlation between the rank of the author (in order of their number of written articles) and the misclassification rate is 0.35, showing that more prolific authors tend to be more accurately classified. One of the most mis-

is this significant? Where's the p-value?

Reading a Paper: More depth

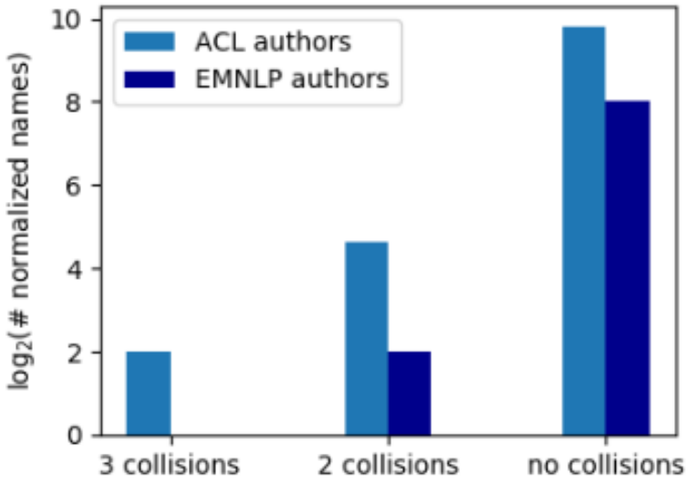
- Are there missing analyses?

authors. Other 45 articles not authored by Christopher Manning were predicted as being written by this author, possibly due to a large number of his citations in the articles' references list and/or similar keywords with those of Christopher Manning.

Does this mean we can get around the prediction by introducing lots of Chris Manning citations?

Reading a Paper: Input Assumptions

- What assumptions do authors make?
  - Does it break any result or analysis?



ACL. From each dataset, we normalized the author names to consist of the initial of the first name and the full last name and removed the authors with less than three articles (to ensure enough data for training and evaluation), leaving us with 922 authors for the ACL dataset and 262 authors for the EMNLP dataset (which represent our classes).

(this is probably fine)

(also, “fewer than” #grammar)

Reading a Paper: More depth

- Do you **know the area?**
  - If not, pick up the related work and try to find issues with novelty claims
- Example: “The Myth of Double-Blind…”
  - Related work is placed in the context of models… authors argue that CNNs over complete paper text are new wrt other work (is it true?)

As far as we are aware, no other study has dealt with analyzing the authorship of articles published at ACL or EMNLP (or a comparably prestigious conference) without restricting the scenario to only a subtask (for example, focusing only on a subset of the data), or limiting the analysis to one aspect of the text (for example, focusing on the stylistic level). While previous studies support the hypothesis that authors of a scientific article are possible to predict from an anonymized paper, we attempt to provide a fuller picture regarding what exactly it is about an anonymous article that can give away its authors.

Reading a Paper: More Depth

- Do the conclusions follow from the experimental data?

- Example: “The Myth of Double-Blind…”
  - Issue: The model is good for top 100 authors. *However*, in practice, you don’t know how prolific the author is for a random input paper.
    - If the output of the model is a non-prolific author, you haven’t gained as much information

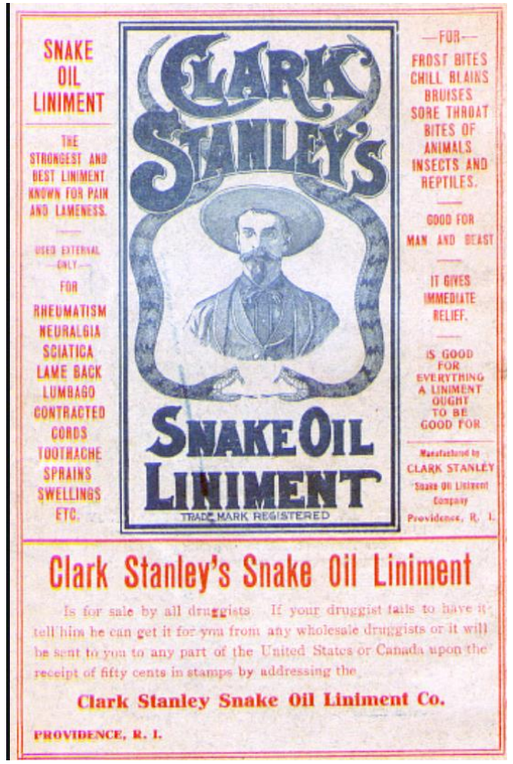
(so it’s not a “Myth”?)

AI. Although we found that the most prolific authors can be inferred with accuracy as high as 87.88% on ACL and 78.49% on EMNLP, the authors with less papers are more and more difficult to infer, which enforces the benefits of the double-blind review in offering any author a fair chance of having their papers accepted in top venues.

“The Myth of Double-Blind…”

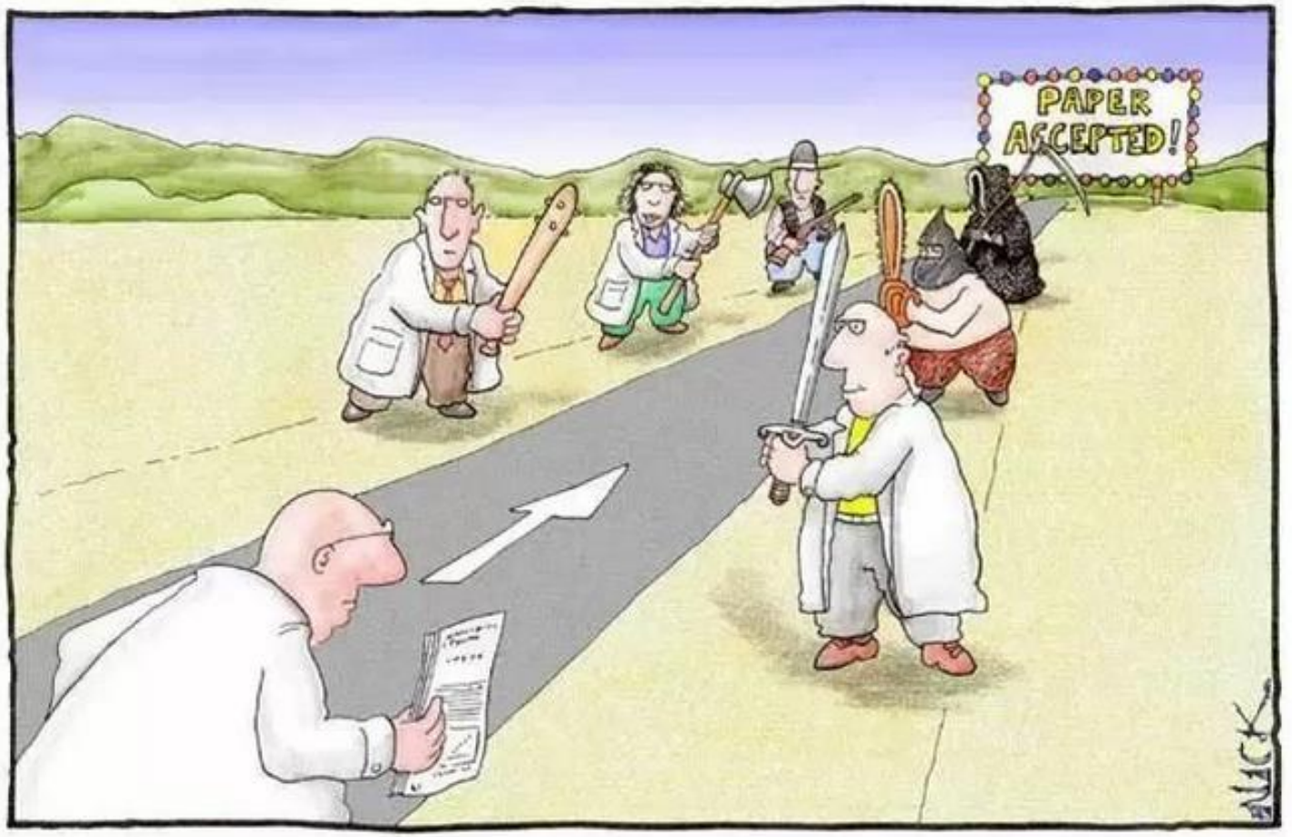
- While skeptical of some reported numbers, they ultimately make true claims:
  - They built a CNN to analyze paper text to predict lists of authors
  - They attained high accuracy for EMNLP/ACL
- The impact is not as clear to me (the result is not surprising…)
- The main contribution is the novel architecture (CNNs + paper text)
  - It’s still a ways off before getting to human-performance at breaking double blind

Peer review



- "Peer review: the worst way to judge research… except for all the others." – Aaron Carroll

- No scrutiny leads to false claims
  - Intent? (this work is too important to fail)
  - Laziness? (do I really need all 100 samples?)
  - Career? (publish/perish)
  - Morality?
- **Peers** have a **responsibility** to assess with **skepticism**



Most scientists regarded the new streamlined peer-review process as 'quite an improvement.'

Scientific Revolution: Peer Review

- **Peer review** enables the development of **consensus** among members of a community
- This is **not** the same as **simple majority**
- A peer-reviewed paper “should” be convincing to all:
  - A **sound** and **valid** logical argument presents a **conclusion** that follows ***necessarily*** from its **premises**

Peer Review: Reality

- A **conference** or **journal** is associated with an **organization or board** (e.g., IEEE or ACM or special interest group)
- That **board** selects (an) **organizer**(s) from academia or industry
  - The “General Chair” or “Editor-in-Chief”
- That **organizer** solicits participants for a **program committee**
- The **program committee** reviews **manuscripts**
  - Often “double-blind” (the authors don’t know the reviews and vice versa)

- A paper’s fate depends on a “discussion” (usually a vote)

Scientific Revolution: Peer review in practice

	OveMer
Review #143A	4
Review #143B	4
Review #143C	4
Review #143D	

Congratulations again for your accepted ICSE paper. Decisions have been made for all conditionally accepted papers, and we have finally accepted 129 papers out of 617 submissions, for an acceptance rate of 20.9%.

Congratulations! Your paper (title below) has been selected as one of the ICSE 2019 ACM SIGSOFT Distinguished Paper awardees. Distinguished Papers represent the very best contributions to the ICSE Technical Track, and are awarded to up to 10% of the papers.

	OveMer
Review #920A	2
Review #920B	2
Review #920C	3

Thank you for submitting your paper to IEEE Security and Privacy ("Oakland") 2020.

The Oakland 2020 program committee is sorry to inform you that your paper #355 has not been accepted and will not appear at the symposium.

Scientific Revolution: Peer review in practice

- Noisy reviews (on the same paper)

```
Overall merit: 1. Reject
      Novelty: 2. Incremental improvement
Writing quality: 3. Adequate

Overall merit: 2. Weak reject
      Novelty: 2. Incremental improvement
Writing quality: 3. Adequate

Overall merit: 3. Can't decide (to be used sparingly)
      Novelty: 2. Incremental improvement
Writing quality: 4. Well-written

Overall merit: 4. Weak accept
      Novelty: 3. New contribution
Writing quality: 4. Well-written

Overall merit: 5. Accept
      Novelty: 3. New contribution
Writing quality: 4. Well-written
```

Actual photo of me on an emotional roller coaster

Scientific Revolution: Peer review in practice

- Overt negativity (especially in computing)

I am growing weary of papers abusing existing decades-old legacy mechanisms for niche applications. System Management Mode needs to die and be removed from processors, to pay down the technical debt carried along by x86's history (e.g., so that x86 systems' firmware complexity / TCB can be reduced). Using SMM to aid debugging / dynamic analysis is a Bad Thing long term, and is definitely not justified given that the gains presented by MALT are just one move in the cat-and-mouse game. I would much prefer that the authors invest their energy in more significant architectural changes that impact dynamic analysis / debuggability (or lack thereof).

(it later got into Oakland at a 13.5% accept rate #rekt)

Scientific Revolution: Peer review in practice

- Can vary a lot with different reviewers/venues

It first got rejected by ICSE

Dear Yu Huang,

Thank you for your submission to ICSE 2023. We regret to inform you that your submission

has not been selected to appear in the conference.

It later got into FSE and also got the ACM Distinguished Paper Award (didn't change anything!)! #rekt

Dear Yu,

Congratulations! Your manuscript will be receiving a SIGSOFT Distinguished Paper Award at ESEC/FSE 2023. We hope you can join us at the award session on Wednesday morning at 9am in Golden Gate A next week.

Twelve of the 127 accepted manuscripts will receive this award.

Reading a Paper, as a peer reviewer

- A **scientific paper** makes an **argument** to a community of **peers**:
  - If some hypothesis is true, then some predicted data should be obtained
  - "If we use textual paper data, then we can predict authorship to break double blind."
    - "predict authorship" -> how accurately? (87%) what source of data? (EMNLP/ACL)
    - "use textual data" -> what's the method? (CNNs) Other things? (references/citations)
- **Peers** read **papers** to assess the **soundness** and **validity** of that **argument**
  - Do the accuracy numbers contribute to the argument?
  - Do I believe CNNs can operate on this data?
  - Is the approach new/better with respect to the baseline?
  - Can I actually break double blind in general? Does it work for other disciplines?
- As a **peer**, you have a **responsibility** to scrutinize **work**!
  - Truth must out!

How to write a paper/research proposal? (more in the future lecture)

- How can I get a research **idea**?
  - From your own experience
    - What do you want to change?
    - What is important to you?

How to write a paper/research proposal? (more in the future lecture)

- How can I get a research **idea**?
  - From your own experience
    - What do you want to change?
    - What is important to you?
  - ***No, I just cannot come up with any idea from my own experience.***

How to write a paper/research proposal?

- How can I get a research **idea**?
  - From your own experience
    - What do you want to change?
    - What is important to you?
  - ***No, I just cannot come up with any idea from my own experience.***
    - **Final trick: read a paper, lift the assumption(s) in it**

TSE, 2012 GenProg: A Generic Method for Automatic Software Repair

Claire Le Goues, ThanhVu Nguyen, Stephanie Forrest, *Senior Member, IEEE*, and Westley Weimer

Abstract—This paper describes GenProg, an automated method for repairing defects in off-the-shelf, legacy programs without formal specifications, program annotations, or special coding practices. GenProg uses an extended form of genetic programming to evolve a program variant that retains required functionality but is not susceptible to a given defect, using existing test suites to encode both the defect and required functionality. Structural differencing algorithms and delta debugging reduce the difference between this variant and the original program to a minimal repair. We describe the algorithm and report experimental results of its success on 16 programs totaling 1.25 M lines of C code and 120K lines of module code, spanning eight classes of defects, in 357 seconds, on average. We analyze the generated repairs qualitatively and quantitatively to demonstrate that the process efficiently produces evolved programs that repair the defect, are not fragile input memorizations, and do not lead to serious degradation in functionality.

Index Terms—Automatic programming, corrections, testing and debugging.

ICSE, 2019

Harnessing Evolution for Multi-Hunk Program Repair

Seemanta Saha†
University of California Santa Barbara
Email: seemantasaha@cs.ucsb.edu

Ripon K. Saha, Mukul R. Prasad
Fujitsu Laboratories of America, Inc.
{rsaha, mukul}@us.fujitsu.com

Schedule

- Plan 1
  - Lectures -> paper presentations -> Lectures -> paper presentations…
- Plan 2
  - All lectures -> all presentations