What is Research?

Mohammed J. Zaki

zaki@cs.rpi.edu

Research

- You need to be aware of what previous work has been done, to have a good feel for the level of difficulty of a proposal and how long it is likely to take.

- Some features are common to all research:
  - review of existing literature
  - original work
  - critical view of your own and other work

What is a valid area of Research?

- Is the work original?
- Is it non-trivial? e.g. a straight-forward implementation of a program or system is not sufficient, but may be if significant research is associated with it.
- Are there existing papers/theses in this area
- Will the examiners think so!

Choosing a Topic

- Good ideas often come from reading, discussing, explaining (and best of all, teaching) what someone else is doing

- Group discussions can be fertile breeding grounds for new ideas

- Read current research papers in areas that interest you, force yourself to present and explain them to others, and ideas will strike you

Finding Research Problems

- Suppose you think idea **X** is very good
- Can you extend **X** by…
  - Making it more accurate (*statistically significantly* more accurate)
  - Making it faster (usually an order of magnitude, or no one cares)
  - Making it an anytime algorithm
  - Making it an online (streaming) algorithm
  - Making it work for a different data type (including uncertain data)
  - Making it work on low powered devices
  - Explaining *why* it works so well
  - Making it work for distributed systems
  - Applying it in a novel setting (industrial/government track)
  - Removing a parameter/assumption
  - Making it disk-aware (if it is currently a main memory algorithm)
  - Making it simpler
- Caveat: Can lead to incremental, boring, low-risk papers

Choosing a Topic: Caveats

- Good to look at suggested open questions in related works, but…

- The authors' suggestions for future research may not be the ones that spawn the best questions

- Those suggestions are probably ones the authors themselves haven't been able (or bothered) to pursue successfully

- Capitalize on your more detached position to escape from the author's mindset and think more laterally about what (s)he's working on, rather than joining him in the tunnel of his vision and identifying open issues through her/his eyes

What Makes a Good Research Problem?

- **It is important:** If you can solve it, you can make money, or save lives, or help children learn a new language

- **You can get real data**: Doing DNA analysis of the Loch Ness Monster would be interesting, but…

- **You can make incremental progress**: Some problems are all-or-nothing. Such problems may be too risky for young scientists

- **There is a clear metric for success**: Some problems fulfill the criteria above, but it is hard to know when you are making progress on them

What do I do now I have a topic?

- Review the literature in the field to ensure that the problem proposed on the basis of the supervisor's knowledge is indeed the most appropriate one to tackle
- Prepare a written statement clearly defining the problem to be studied, carefully stating the aims of the project in such operational terms that it can be reasonably known when the aims have been achieved
- Produce a detailed plan of work for at least the first year of the study including
  - methods to be used, together with an assessment of the suitability of available equipment
  - a list of the materials and resources required and an assessment of their availability
  - an assessment of the time required to undertake the various operations projected to ensure the feasibility of the program
- Produce a general account of likely development beyond the first year

Who is responsible?

- Supervisor responsible for
  - Ensuring the topic is appropriate. However they may not be able to guarantee success, but in this case they should warn of the danger
  - Be available to discuss, suggest, read, comment,…
- Student responsible for
  - Approaching the supervisor at regular intervals to discuss progress
  - Doing the work and writing the thesis

Framing Research Problems

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

  "I hate it when a paper under review does not give a concise definition of the problem"

- Your research statement should be **falsifiable**

  Falsifiability (or **refutability**) is the logical possibility that an claim can be shown false by **Falsifiability** (or **refutability**) is the logical possibility that an claim can be shown false by an observation or a physical experiment. That something is 'falsifiable' does not mean it is an observation or a physical experiment. That something is 'falsifiable' does not mean it is false; rather, that *if* it is false, then this can be shown by observation or experiment

- Avoid the contradiction of claiming that the problem is very important, but there is no real data

Avoid Complex Solutions

- …are less likely to generalize to datasets.
- …are much easer to overfit with.
- …are harder to explain well.
- …are difficult to reproduce by others.
- …are less likely to be cited.
- Simplicity is a strength, not a weakness, acknowledge it and claim it as an advantage
- Always start by eliminating simple answers
- Your paper is implicitly claiming "*this is the simplest way to get results this good*"

Reproducibility

- Reproducibility is one of the main principles of the scientific method, and refers to the ability of a test or experiment to be accurately reproduced, or replicated, by someone else working independently.

- *"The vast body of results being generated by current computational science practice suffer a large and growing credibility gap: it is impossible to believe most of the computational results shown in conferences and papers" – David Donoho et. al. 2008 (15 Years of Reproducible Research in Computational Harmonic Analysis)*

Idealized Algorithm for Writing a Paper

- Find problem/data
- Start writing (*yes*, start writing *before* and *during* research
- Do research/solve problem
- Finish 95% draft
- Send preview to mock reviewers
- Send preview to the rival authors (virtually or literally)
- Revise
- Submit

Writing the Paper

- Make a working title
- Introduce the topic and define (informally at this stage) terminology
- Motivation: Emphasize why is the topic important
- Relate to current knowledge: what's been done
- Indicate the gap: what need's to be done?
- Formally pose research questions
- Explain any necessary background material.
- Introduce formal definitions.
- Introduce your novel algorithm/representation/data structure etc.
- Describe experimental set-up, explain what the experiments will show
- Describe the datasets
- Summarize results with figures/tables
- Discuss results
- Explain conflicting results, unexpected findings and discrepancies with other research
- State limitations of the study
- State importance of findings
- Announce directions for further research
- Acknowledgements
- References

Make Reviewer's Life Easy

- *I have often said reviewers make an initial impression on the first page and don't change 80% of the time* -- Mike Pazzani
- The introduction acts as an anchor. By the end of the introduction the reviewer *must* know.
  - What is the problem?
  - Why is it interesting and important?
  - Why is it hard? why do naive approaches fail?
  - Why hasn't it been solved before? (Or, what's wrong with previous proposed solutions?)
  - What are the key components of my approach and
  - If possible, an interesting figure on the first page helps results? Also include any specific limitations.
  - A final paragraph or subsection: "Summary of Contributions". It should list the major contributions in bullet form, mentioning in which sections they can be found. This material doubles as an outline of the rest of the paper, saving space and eliminating redundancy.
- Avoid "Junk" paragraphs: In section 2, we do blah, in Sec 3, we do more blah, and we conclude in section 6 with blah blah!

Reproducibility

- Mention all parameter setting, and make all data and code available
- Set up a web-site as supplementary material
- Treat this as an obligation to the community
- But it will also lead to more citations – the actual currency of the research community
- It will help document your work for later extension

Avoid Plagiarism Like the Plague!

- Never copy and paste
- Acknowledge all sources of text, figures, data, software, etc.
- Try to write in your own words
- Avoid temptation of repeating other papers' description in your related work section

Other Issues

- Motivation: It is very important to convince the reviewers that your work is *original*
  - Do a detailed literature search
  - Use mock reviewers
  - Explain why your work is different
- Avoid "Laundry List" Citations
- Always write your paper imagining the most cynical reviewer looking over your shoulder. This reviewer does not particularly like you, does not have a lot of time to spend on your paper, and does not think you are working in an interesting area. But he *will* listen to reason.

Heilmeier Questions
(George Heilmeier, Director of ARPA in 70s)

1. What are you trying to do? Articulate your objectives using absolutely no jargon. What is the problem? Why is it hard?

2. How is it done today, and what are the limits of current practice?

3. What's new in your approach and why do you think it will be successful?

4. Who cares?

5. If you're successful, what difference will it make? What impact will success have? How will it be measured?

6. What are the risks and the payoffs?

7. How much will it cost?

8. How long will it take?

9. What are the midterm and final "exams" to check for success? How will progress be measured?

Quora: 2015 Hot Topics in CS Research

- **Abundant-data applications, algorithms, and architectures** are a meta-topic that includes research avenues such as data mining (quickly finding relatively simple patterns in massive amounts of loosely structured data, evaluating and labeling data, etc), machine learning (building mathematical models that represent structure and statistical trends in data, with good predictive properties), hardware architectures to process more data than is possible today.

Quora: 2015 Hot Topics in CS Research

- **Artificial intelligence and robotics** - broadly, figuring out how to formalize human capabilities, which currently appear beyond the reach of computers and robots, then make computers and robots more efficient at it. Self-driving cars and swarms of search-and-rescue robots are a good illustration. In the past, once good model were found for something (such as computer-aided design of electronic circuits), this research moves into a different field – the design of efficient algorithms, statistical models, computing hardware, etc.

Quora: 2015 Hot Topics in CS Research

- **Bio-informatics and other uses of CS in biology, biomedical engineering, and medicine,** including *systems biology* (modeling interactions of multiple systems in a living organism, including immune systems and cancer development), *computational biophysics* (modeling and understanding mechanical, electrical, and molecular-level interactions inside an organism), *computational neurobiology* (understanding how organisms process incoming information and react to it, control their bodies, store information, and think). There is a very large gap between what is known about brain structure and the functional capabilities of a living brain – closing this gap is one of the grand challenged in modern science and engineering. DNA analysis and genetics have also become computer-based in the last 20 years. Biomedical engineering is another major area of growth, where microprocessor-based systems can monitor vital signs, and even administer life-saving medications without waiting for a doctor. Computer-aided design of prosthetics is very promising.

Quora: 2015 Hot Topics in CS Research

- **Computer-assisted education**, especially at the high-school level. Even for CS, few high schools offer competent curriculum, even in developed countries. Cheat-proof automated support for exams and testing, essay grading, generation of multiple-choice questions. Support for learning specific skills, such as programming (immediate feedback on simple mistakes and suggestions on how to fix them, peer grading, style analysis).

Quora: 2015 Hot Topics in CS Research

- **Databases, data centers, information retrieval, and natural-language processing**: collecting and storing massive collections of data and making them easily available (indexing, search), helping computers understand (structure in) human-generated documents and artifacts of all kinds (speech, video, text, motion, biometrics) and helping people search for the information they need when they need it. There are many interactions with abundant-data applications here, as well as with human-computer interaction, as well as with networking.

Quora: 2015 Hot Topics in CS Research

- **Emerging technologies for computing hardware, communication, and sensing**: new models of computation (such as optical and quantum computing) and figuring out what they are [not] good for. Best uses for three-dimensional integrated circuits and a variety of new memory chips. Modeling and using new types of electronic switches (memristors, devices using carbon nano-tubes, etc), quantum communication and cryptography, and a lot more.

Quora: 2015 Hot Topics in CS Research

- **Human-computer interaction** covers human-computer interface design and focused techniques that allow computers to understand people (detect emotions, intent, level of skill), as well as the design of human-facing software (social networks) and hardware (talking smart-phones and self-driving cars).

Quora: 2015 Hot Topics in CS Research

- **Large-scale networking**: high-performance hardware for data centers, mobile networking, support for more efficient multicast, multimedia, and high-level user-facing services (social networks), networking services for developing countries (without permanent high-bandwidth connections), various policy issues (who should run the Internet and whether the governments should control it). Outer-space communication networks. Network security (which I also listed under Security) is also a big deal.

Quora: 2015 Hot Topics in CS Research

- **Limits of computation and communication** at the level of problem types (some problems cannot be solved in principle!), algorithms (sometimes an efficient algorithm is unlikely to exist) and physical resources, especially space, time, energy and materials. This topic covers Complexity Theory from Theoretical CS, but also the practical obstacles faced by the designers of modern electronic systems, hinting at limits that have not yet been formalized.

Quora: 2015 Hot Topics in CS Research

- **Multimedia:** graphics, audio (speech, music, ambient sound), video – analysis, compression, generation, playback, multi-channel communication etc. Both hardware and software are involved. Specific questions include scene analysis (describing what's on the picture), comprehending movement, synthesizing realistic multimedia, etc.

Quora: 2015 Hot Topics in CS Research

- **Programming languages and environments**: automated analysis of programs in terms of correctness and resource requirements, comparisons between languages, software support for languages (i.e., compilation), program optimization, support for parallel programming, domain-specific languages, interactions between languages, systems that assist programmers by inferring their intent.

Quora: 2015 Hot Topics in CS Research

- **Security of computer systems and support for digital democracy**, including network-level security (intrusion detection and defense), OS-level security (anti-virus SW) and physical security (biometrics, tamper-proof packaging, trusted computing on untrusted platforms), support for personal privacy (efficient and user-friendly encryption), free speech (file sharing, circumventing sensors and network restrictions by oppressive regimes), as well as issues related to electronic polls and voting. Security is also a major issue in the use of embedded systems and the Internet of Things (IoT).

Quora: 2015 Hot Topics in CS Research

- **Verification, proofs, and automated debugging** of hardware designs, software, networking protocols, mathematical theorems, etc. This includes formal reasoning (proof systems and new types of logical arguments), finding bugs efficiently and diagnosing them, finding bug fixes, and confirming the absence of bugs (usually by means of automated theorem-proving).

Slide Sources

- http://www.cs.ucr.edu/~eamonn/Keogh_SIGKDD09_tutorial.pdf
- http://www.design.caltech.edu/erik/Misc/Heilmeier_Questions.html
- http://www.cs.waikato.ac.nz/GradConf/talks/bruce/ChoosingTopic.pdf
- http://www.forbes.com/sites/quora/2015/04/22/13-of-2015s-hottest-topics-in-computer-science-research/