Graduate Student Orientation

Vasant Honavar
Artificial Intelligence Research Laboratory
Department of Computer Science
Bioinformatics and Computational Biology Graduate Program
Center for Computational Intelligence, Learning, & Discovery
Iowa State University
honavar@cs.iastate.edu
www.cs.iastate.edu/~honavar/
www.cild.iastate.edu/
www.cs.iastate.edu
www.bcb.iastate.edu/

Overview

- Computer Science – The big picture
- Computer Science Research – what, why, and how
- Graduate school survival tips
- Responsible conduct - ethics

Computer Science
The science of information processing

- The language of computation is the best language we have so far for describing how information is encoded, stored, manipulated and used by natural as well as synthetic systems
- Algorithmic or information processing models provide for biological, cognitive, and social sciences what calculus provided for classical physics

A little history quiz

- Aristotle
- Panini
- Al-Khowirizmi
- Aristotle
- Hobbs
- Leibnitz
- Boole
- Frege
- Tarski
- Hilbert
- Godel

- Turing
- Church
- Kleene
- Post
- Markov
- Atanasoff
- Wiener
- Shannon
- Von Neumann
- McCullough
- Rashevsky

- Bardeen
- Schockley
- Brattain
- Backus
- Chomsky
- McCarthy
- Minsky
- Newell
- Simon
- Uhr
- Cobham

- Edmunds
- Cook
- Karp
- Codd
- Gates
- Jobs
- Cerf
- Berners-Lee
- Gosling
- Brin
- Page

The Road to Computer Science

- History of computer science is really a history of human attempts to understand nous (the rational mind)
  - intelligence – processes of acquiring, processing, and using information
- Aristotle (384-322 BC) distinguishes matter from form thereby laying the foundations of representation
- Panini (350 BC) develops a formal grammar for Sanskrit
- Al Khowarizmi (825) introduces algorithms in his text on mathematics

The road to Computer Science

- Hobbs (1650) suggests that thinking is a rule-based process analogous to arithmetic
- Leibnitz (1646-1716) seeks a general method for reducing all truths to a kind of calculation
- Boole (1815-1864) proposes logic and probability as the basis of laws of thought
- Frege (1848-1925) further develops first order logic
- Tarski (1902-1983) introduces a theory of reference for relating objects in a logic to objects in the world

The road to Computer Science

- Hilbert (1862-1943) presents the decision problem – Is there an effective procedure for determining whether or not a given theorem logically follows from a given set of axioms?
- Godel (1906-1978) shows the existence of an effective procedure to prove any theorem in Frege's logic and proves the incompleteness theorem
- Turing (1912-1954) invents the Turing Machine to formalize the notion of an effective procedure

The road to Computer Science

- Church, Kleene, Post, Markov (1930-1950) develop other models of computation based on alternative formalizations of effective procedures
- Several special purpose analog and digital computers are built (including the Atanasoff-Berry Computer – 1937-1942)
- Turing and Church put forth the Church-Turing thesis that Turing machines are universal computers (1948)

The road to Computer Science

- Wiener (1947-1948) introduces Cybernetics – the science of communication and control in humans and machines
- Shannon (1948) and Wiener develop information theory – offering a means to quantify information
- Von Neumann (1956) works out a detailed design for a stored program digital computer
- Several digital computers are constructed and universal languages for programming are developed – Lisp, Snobol, Fortran…

The road to Computer Science

- Von Neumann, McCulloch, Rashevsky (1940-1956), investigate the relationship between the brain and the computer
- Chomsky (1956) develops the Chomsky hierarchy of languages
- Von Neumann and Morgenstern develop a formal framework for rational decision making under uncertainty
- Von Neumann (1956) develops a theory of self-reproducing automata
- McCarthy, Minsky, Selfridge, Simon, Newell, Uhr et al (1956) begin to investigate the possibility of artificial intelligence

The road to Computer Science

- McCarthy (1957-61) introduces the first time sharing operating system (CTSS)
- Dantzig and Edmunds (1960-62) introduce reduction – a general transformation from one class of problems to another
- Cobham and Edmunds (1964-65) introduce polynomial and exponential complexity
- Codd introduces the relational database (1970)
- Cook and Karp (1971-72) develop the theory of NP-completeness which helps recognize problems that are intractable

The road to Computer Science

- Vinton Cerf invents the Internet (1973-1974)
- Internet rolled out (1984)
- Tim Berners-Lee invents the World-Wide Web (1989-1991)
- James Gosling and colleagues invent Java – a platform independent programming language and environment (1991-1994)
- Sergei Brin and Larry Page launch Google (1996-1998)
- The rest is recent history ☺

Computer science

- Computer science is the science of information processing
- The language of computation is the best language we have so far for describing how information is encoded, stored, manipulated and used by natural as well as synthetic systems

Algorithms as Theories

- Computation: Cognition : : Calculus : Physics
  (Artificial Intelligence, Cognitive Science)
- Computation: Life : : Calculus : Physics
  (Computational Biology)
- Computation: Society : : Calculus : Physics
  (Computational Economics, Computational Organization Theory)

Conceptual Impact of Computer Science
Algorithms as theories

We will have a theory of

- Learning when we have precise information processing models of learning (computer programs that learn from experience)
- Protein folding when we have an algorithm that accepts a linear sequence of amino acids as input and produces a description of the 3-dimensional structure of a protein as output
- Economic behavior when ..

Conceptual impact of Computer Science

Pre-Turing

- Focus on physical basis of the universe with the objective of explaining all natural phenomena in terms of physical processes

Post-Turing

- Focus on information processing basis of the universe with the objective of explaining natural phenomena in algorithmic terms – in terms of processes that acquire, store, process, manipulate, and use information

Conceptual impact of Computer Science

We understand a phenomenon when we can write a computer program that models it at the desired level of detail

- When theories and explanations in science take the form of algorithms, all sciences morph into computer science!
- Implication – new horizons to explore for computer scientists!

All Science is Computer Science!

Computer Science has given birth to:

- Bioinformatics and computational biology
- Cognitive science
- Computational chemistry and chemo-informatics
- Computational economics
- Computational neuroscience
- Computational organization theory
- Ecological informatics
- Engineering informatics
- Medical informatics
- Social informatics
- ….

All Science is Computer Science!

Computer scientists are being hired by departments of

- Biological Sciences
- Chemistry
- Economics
- Engineering
- Philosophy
- Physics
- Psychology
- Sociology ……….

Some Grand Challenges for Computer Scientists

- Organizing and managing the world's knowledge
- Facilitating collaboration and interaction
- Improving Healthcare
- Educating the masses
- Extending human cognitive abilities
- Understanding the informational basis of life
- Understanding thought, learning, and behavior
- Understanding complex systems – organizations, civilizations, economies, ecologies..
- Building smart artifacts – smart homes, smart highways.. artifacts that can improve our lives

Organizing and managing the world's knowledge

- Information explosion: text, audio, video, life experiences, experimental data
- We need tools to automatically organize, retrieve, summarize, transmit, secure this information
  - Billions of web pages
  - Verizon calling graphs
  - Wal-Mart transactions
  - Millions of satellite images
  - Millions of genomes
  - Thousands of macromolecules
  - Recordings from millions of neurons

Understanding the informational basis of life

- Instructions for building organisms are digital codes (A, G, C, T)
- Nature's molecular computers decode and execute these instructions
- Living systems are information processing systems
  - Computer scientists can help cure cancer
  - Computer scientists can help design draught resistant crops

Implications for research and education

- What should a computer scientist know?
  - Suppose you want to build software for analysis of genome sequences – you will need to know some biology
  - Suppose you want to build a domain-specific programming language for programming software agents that monitor and trade stocks on the internet – you better know something about economics and how markets work
- What should every literate person know?
  - Suppose you are an epidemiologist interested in modeling the spread of infectious diseases – you better know how to model and simulate hundreds of agents

Computer Science Research Strengths @ ISU

Department of Computer Science

- Algorithms
- Artificial Intelligence and Robotics
- Bioinformatics and Computational Biology
- Databases and Information Systems
- Distributed Computing, Networks, and Systems
- Programming Languages
- Software Engineering
- Theory of Computation and Complexity

Remember – these areas are not necessarily disjoint!

Computer Science Research Opportunities @ ISU

Interdepartmental research-based training programs

- Bioinformatics and Computational Biology
  - One of the strongest such program in the US
  - Ph.D. program
  - Ph.D. co-major
  - No MS program, but courses open to MS students
- Human-Computer Interaction (HCI)
  - Ph.D. program
  - MS Program
- Information Assurance
  - MS Program

Computer Science Research Opportunities @ ISU

Research Centers and Institutes

- Center for Computational Intelligence, Learning, and Discovery
- Center for Nondestructive Evaluation
- Center for Integrated Animal Genomics
- Information Assurance Center
- Information Infrastructure Institute
- Cyberinnovation Institute
- Scalable Computing Laboratory (Ames Lab)
- Statistical Computing Laboratory
- Laurence H. Baker Center for Bioinformatics and Biological Statistics
- Virtual Reality Applications Center

Internship Opportunities

- Google
- Intel Research Labs
- IBM Research Labs
- Mayo Clinic
- Microsoft Research Labs
- NASA Ames Research Center
- Philips Research
- Siemens Research
- Xerox PARC
- Yahoo!

and many more.. Keep an eye out for announcements

Overview

- Computer Science – The big picture
- Computer Science Research – what, why, and how
- Graduate school survival tips
- Responsible Conduct - ethics

Research – what is research?

Everything has been thought of before, but the problem is to think of it again – Goethe

- Goal of research: Advancing the state of knowledge
  - High school – your own knowledge
  - Industry – your company's knowledge
  - Research university – the world's knowledge
- How do you know the state of the world's knowledge?
  - Peer reviewed publications
  - Research monographs
  - Conferences and Workshops

Research – Why and How

Why do research?

- For the pleasure of finding things out
- To improve the quality of our lives

How to do research?

- That is what you are here to learn
- You learn to do research by doing research – under the guidance or one or more mentors
- It helps to have a good mentor
- Drive, curiosity, hard work, creativity, persistence, thoroughness, intelligence do pay off

How to evaluate research

- If you don't know how to evaluate what you have done, you won't know if you have succeeded – and you won't be able to convince others either
- Defining evaluation criteria can be as important as the research activity itself
  - In early phases, it is part of the problem definition
  - In later phases, better criteria are often identified
  - Early evaluation can guide or redirect research

How to evaluate quality of research?

General criteria

- Correctness
- Novelty, originality, creativity, surprise value
- Impact (economic, social, scientific, expository)
- Difficulty, relevance to other known hard problems
- Foundational value (usefulness in further developments)
- Cumulative value (building on existing work)
- Impact on the practice of computing

Computing Research Paradigms

- Theoretical – Define abstraction, prove results (theorems) e.g., computability theory, complexity theory, learning theory …
- Experimental – Build and experiment with, measure, evaluate systems
- Creative – Invent new artifacts e.g., Computer, Internet, World-wide web, search engine
- Synthetic – Unify a body of research results – generalize, specialize results
- Cross-disciplinary – Apply computing to solve problems in other disciplines (e.g., bioinformatics), construct computational models (e.g., cognitive science, computational biology)

Theoretical paradigm

- Define abstraction, prove results (theorems) e.g., automata theory, formal languages, computability theory, complexity theory, learning theory …
- Quality
  - Mathematical soundness of results
  - Abstraction should discard irrelevant details while maintaining essential aspects of real problems
  - Contribution to understanding of the domain
  - Ideally opens up new areas of research, offers new proof or analysis techniques, or applications

Experimental paradigm

- Build and experiment with systems
  - Performance measurements, Lesion studies ..
- Quality
  - Sound experimental design, controls, statistics
  - Relevance of hypothesis, experiments relative to theory
  - New information, not already available from theory or other experimental data
  - Advance theory, supplement theory, or uncover limitations of existing theory

Creative paradigm

- Invent new artifacts e.g., Computer, Internet, World-wide web
- Quality:
  - Utility
  - Originality
  - Impact
  - Application of sound theory
  - Demonstrable improvements over the state of the art

Synthetic paradigm

- Unify a body of research results
- Quality:
  - Contribution to understanding of diverse results within a common framework
  - Contribution to clarity, new insights, new directions
  - Generalization or specialization of existing results

Cross-disciplinary research

- Apply computing to solve problems in other disciplines
  - Bioinformatics, Cheminformatics, Social informatics
- Develop and analyze computational models
  - Computational biology, Computational organizational theory, Computational economics, Computational ecology
- Quality:
  - Originality relative to the application area
  - Soundness and currency of the CS applied
  - Utility – advancing science or practice in the application area
  - Contribution to opening up new CS problems

Summary

Research is

- A lot of fun
- Extremely rewarding
- Hard work
- Requires learning and practice
- The most important component of graduate study

Learn to be a competent -- if not outstanding -- researcher, and have fun doing it!

Research Tips

- Work on relevant problems
- Read with a purpose
- Know the literature your area
- Maintain a research notebook
- Avoid tunnel vision
- Be engaged with the scientific community
- Write well
- Give compelling presentations

Overview

- Computer Science – The big picture
- Computer Science Research – what, why, and how
- Graduate school survival tips
- Responsible conduct - ethics

Masters versus Ph.D.

- Masters – recommended for those interested in working in non research and development jobs in industry
- Ph.D. – recommended for those interested in
  - Research positions in industry, government, or academia, or
  - Faculty positions in academia
  - Leadership positions in industry

Masters versus Ph.D. – Placement of Graduates

- Masters
  - Software development positions at Oracle, Cisco, Microsoft, IBM, HP, Nokia …
  - Ph.D. programs at other schools
- Ph.D.
  - Faculty positions at Kansas State, Utah State, North Texas..
  - Postdoctoral positions at RPI, IBM Research..
  - Research and development positions at Lucent, IBM, Yahoo!, Fair Isaac, Hughes Research Labs, Mayo Clinic
  - Software development positions in industry

Masters versus PhD

- Masters degree
  - With project
    - Terminal degree
    - Aimed at industry-bound
  - With thesis
    - Not a terminal degree
    - Often in preparation for Ph.D.
    - Can help you find out whether you like research

Masters Thesis

- Topic is usually more clearly defined at the outset – less open-ended than a Ph.D. topic
- If judiciously chosen, can be good preparation before beginning work for a Ph.D.
- Typically you have less control over the overall direction of an M.S. thesis as compared to a Ph.D. thesis
- Ideally should result in a publishable paper

Masters project (non thesis option)

- Has more of a software design and implementation focus
- Ideally should result in a piece of well-designed software that solves some specific problem
- Often involves working in teams
- May contribute to an open source project
- May result in a publishable paper

Ph.D. Thesis

- Should demonstrate ability to define, pursue, and complete independent and original research
- Should result in 3 or 4 journal publications (and several conference publications)
- Should prepare the recipient for an independent research career in academia or industry

Advise to Ph.D. students

- Forget it unless you are *strongly* motivated
- Find an area that really interests you – don't blindly follow the fashion
- Study the subject until you become an expert
- Get involved in a research group as soon as possible
- Become part of the research community – attend conferences
- Publish early and often but without sacrificing quality

Finding a research problem

- Pick a problem that is
  - Hard enough to be interesting
  - Not already solved
  - Has a good chance of being solvable by you
  - Not being solved concurrently by lots of other people
- Take advantage of your advisor’s expertise
- Don’t be overly afraid of reinventing – it is not uncommon when you are at the threshold of new work

Finding a research problem: Where to look

- Other people's papers
  - Stated open problems – may not be a good bet unless you have special ability or background that the author does not
  - Unstated problems – oversights, variations, generalizations, combinations
- Your thesis advisor's research interests and ongoing projects in his or her research lab
  - Stated open problems
  - Unstated problems

Approaching a research problem

- Tackle a hard, sufficiently rich problem
  - Hopefully you will solve it, first
  - If not, you will probably discover something else interesting along the way
- Tackle a not so hard, marginally interesting problem
- Keep in mind that the most obviously interesting, the most obviously high impact problems as well as problems opened up by a new development (the low hanging fruit) are likely to attract the most competition

Exploit your strengths and opportunities

- Personal strengths
  - Hard work?
  - Exceptional intelligence?
  - Exceptional creativity?
  - Cross-disciplinary background?
- Local opportunities
  - Collaborative opportunities
  - Funded research projects
  - Special facilities
  - Advance information through other local research

Publication – why?

- Get feedback and criticism from peers
- Disseminate research results – advance the world's knowledge
- Stake claim to ideas
- Establish evidence of your research accomplishments

Publication – the process

- Technical report
  - Stake claim
  - Get draft out quickly for review and feedback
- Conference paper – short and focused
  - Engage the research community, enhance visibility
  - Get more review and feedback from peers
- Journal paper – archive important ideas
- Book – collect related work, organize for teaching

Publication – the process

- Model your first paper on a paper that you admire
- Explain the motivation behind your research and how it relates to other work
- Overview, detail, summary
- For conferences and workshops, focus clearly on a few exciting ideas and results
- Give credit where credit is due – cite other related work

http://www.cs.iastate.edu/~honavar/grad-advice.html

Publication – software

- Free dissemination (e.g., via Internet)
  - Will people use it for free?
  - Will they take your code and improve on it?
- Commercial
  - Will people pay for it?
  - Will they steal it or imitate it?

Overview

- Computer Science – The big picture
- Computer Science Research – what, why, and how
- Graduate school survival tips
- Responsible conduct - ethics

Getting Started – the first year

- Take courses
- Make the transition to being a graduate student
- Learn about research opportunities
- Attend research seminars (610) in your area of interest
- Attend departmental colloquia and other seminars on campus
- Remedy any deficiencies in background, language skills (e.g., spoken and written English)
- Find an advisor, form POS
- Maintain satisfactory academic progress
- Work hard, but also make friends, have fun
- Begin research with advisor or pursue an internship in the first summer

Resources for graduate students

- Surviving in graduate school
- Ethics – Responsible conduct in research, writing, and authorship
- How to do research
- How to present research results – writing, presentation
- Career resources
- Life after graduate school

http://www.cs.iastate.edu/~honavar/grad-advice.html

Surviving Graduate School

Graduate School Survival Tips (Wanda Pratt, Marie desJardins)

- Getting the most out of the relationship with your research advisor or boss
- Getting the most out of what you read
- Making continual progress on your research
- Finding a thesis topic or formulating a research plan
- Characteristics to look for in a good advisor, mentor, boss, or committee member
- Avoiding the research blues

http://projects.ischool.washington.edu/wpratt/survive.htm
http://www.acm.org/crossroads/xrds1-2/advice1.html
http://www.acm.org/crossroads/xrds1-3/advice2.html

Ethics – Responsible conduct

- Course on Responsible Conduct in Research

  http://ccnmtl.columbia.edu/projects/rcr/index.html
  - Responsible Authorship and Peer Review
  - Collaborative Science
  - Data Acquisition and Management
  - Mentorship
  - Research Misconduct
  - Conflicts of Interest
- Plagiarism and how to avoid it
  http://www.northwestern.edu/uacc/plagiar.html
- Academic misconduct
  - refer to ISU student handbook
- Ethics in Computing

  http://ethics.csc.ncsu.edu/

Useful Resources

http://www.cs.iastate.edu/~honavar/grad-advice.html