A Structured Wikipedia for Mathematics

JANUARY 6, 2010
ICS POSTER TALK
BY HENRY LIN

OR "MATHEMATICS IN A WEB 2.0 WORLD"

Overview

- Background and motivation
- A new system for organizing mathematics
- An interesting new phenomenon: how the Internet is changing the way research is conducted
- Open questions

Background

- Problem: Related work is difficult to find
  - Large amount of research is produced each year
  - Many overlapping fields of research
  - Naming/keyword classification can be inadequate

- In theoretical CS, often need to search for related:
  - Mathematical theorems
  - Algorithmic problems
  - Complexity classes
  - Open questions
  - Etc.

Existing Resources on the Internet

Wikipedia/PlanetMath:

- Stores general knowledge and knowledge on mathematics
  - www.planetmath.org by Nathan Egge, Aaron Krowne, and others
  - www.wikipedia.org by Jimmy Wales, Larry Sanger, and others

Complexity Zoo/Complexity Garden:

- Stores complexity classes and computational problems
  - qwiki.stanford.edu/wiki/complexity_zoo by Scott Aaronson, and others
  - qwiki.stanford.edu/wiki/complexity_garden by Hunter Monroe, and others

- Open Problem Garden: stores open problems
  - garden.irmacs.sfu.ca by Matt DeVos and Robert Šámal

- The Scheduling Zoo: stores results on scheduling problems
  - www.lix.polytechnique.fr/~durr/query/ by Peter Brucker and Sigrid Knust

A Structured Wikipedia for Mathematics

- Except for the scheduling zoo, the previous systems are difficult to use to search for related work
- Can a better system be designed to help organize existing results and open questions?
- The system must be:
  - Simple to understand
  - Easy to use
  - Organize and link related results, for better searching

A Structured Wikipedia for Mathematics

One idea to help organize related work:

- A collaborative website like Wikipedia
- Use indentation to structure results
- Structure allows system to organize and automatically link related results

A special case of Chernoff's Bound

Initial Conditions:

- Let $x_1, x_2, \ldots, x_n$ be
  - Independent random variables
  - Binary random variables
    - With probability ½ of being 0 or 1
- Let $X = x_1 + x_2 + \ldots + x_n$
- Let $\mu = E[X]$
- Let $0 < \delta < 1$

Conclusion:

- $Pr[\ X \geq (1-\delta)\cdot\mu\ ] \leq \exp(-\ \mu\cdot\delta^2)$

A special case of Chernoff's Bound

Initial Conditions:

- Let $x_1, x_2, \ldots, x_n$ be
  - Independent random variables
  - Binary random variables
    - With probability ½ of being 0 or 1
- Let $X = x_1 + x_2 + \ldots + x_n$
- Let $\mu = E[X]$
- Let $\delta > 0$

Conclusion:

- $Pr[\, X \geq (1+\delta)\cdot\mu \,] \leq \exp(-\mu\cdot\delta^2)$

A general case of Chernoff's Bound

Initial Conditions:

- Let $x_1, x_2, \ldots, x_n$ be
  - Independent random variables
  - Binary random variables
    - With probability p_i of being 1 and $(1-p_i)$ of being 0
- Let $X = x_1 + x_2 + \ldots + x_n$
- Let $\mu = E[X]$
- Let $\delta > 0$

Conclusion:

- $Pr[\, X \geq (1+\delta)\cdot\mu \,] \leq (e^{\delta} / (1 + \delta)^{(1+\delta)})^{\mu}$

Hoeffding's Inequality

Initial Conditions:

- Let $x_1, x_2, \ldots, x_n$ be
  - Independent random variables
  - Such that $x_i \in [a_i, b_i]$ almost surely

- Let $X = x_1 + x_2 + \ldots + x_n$
- Let $\mu = E[X]$
- Let $\delta > 0$

Conclusion:

- $Pr[X \geq \mu + \delta] \leq \exp(- 2 \cdot \delta^2 / \sum (a_i - b_i)^2)$

Azuma–Hoeffding Inequality

Initial Conditions:

- Let $x_1, x_2, \ldots, x_n$ be
  - Such that $Y_i = x_1 + \ldots + x_i$ forms a martingale
  - Such that $|x_i| < c_i$ almost surely

- Let $X = x_1 + x_2 + \ldots + x_n$
- Let $\mu = E[X]$
- Let $\delta > 0$

Conclusion:

- $Pr[\, X \geq \delta \,] \leq \exp(-\delta^2 / 2 \cdot \sum c_i^2)$

Linking related complexity classes

NP: Nondeterministic Polynomial-Time

- The class of decision problems solvable by an NP machine such that:
  - If the answer is 'yes,' at least one of the computation paths accept.
  - If the answer is 'no,' all of the computation paths reject.

RP: Randomized Polynomial-Time

- The class of decision problems solvable by an NP machine such that:
  - If the answer is 'yes,' at least 1/2 of the computation paths accept.
  - If the answer is 'no,' all of the computation paths reject.

BPP: Bounded-Error Probabilistic Polynomial-Time

- The class of decision problems solvable by an NP machine such that:
  - If the answer is 'yes,' at least 2/3 of the computation paths accept.
  - If the answer is 'no,' at least 2/3 of the computation paths reject.

Open Questions

- Is the above system adequate for organizing and linking related results?
- Should more rules be imposed on the system?
- What other functionality should be implemented into the system?

Other interesting projects

Tricki:

- A collaborative website to store problem-solving techniques in math
  - www.tricki.org by Tim Gowers and others

Polymath:

- Collaborative efforts to solve open math problems
  - www.polymathprojects.org by Tim Gowers and others

Vdash:

- A collaborative website to store proofs in a formal, checkable manner
  - www.vdash.org by Cameron Freer

The Polymath Project

- Collaborative work done over blogs, discussion forums, and Wiki pages

- Derived a new simpler proof for a variation on the density Hales-Jewett theorem.
  - Initiated Feb 1, 2009
  - New result being written
  - Article about result published in Nature

Open Questions

- Can a similar structured system be developed to link and organize proofs, or ideas for proofs of theorems?
- Is there some structure that proofs can be written in to make them more understandable and checkable?
- Are there other ways to make collaborative research more feasible?

Future Work

- Future work to be found at:

http://itcs.tsinghua.edu.cn/~henry/

- Or Google “Henry Lin ITCS”

- Questions?