Experimental Mathematics and Data Mining:

Extracting Identities from the Online Encyclopedia of Integer Sequences

Hieu D. Nguyen

Rowan University

Mathfest - Lexington, KY
August 4, 2011

Online Encyclopedia of Integer Sequences (OEIS)

- **Searchable online database - http://oeis.org/**
- **Contains almost 200,000 integer sequences**
- **Created by Neil Sloane (AT & T Bell Labs)**
- **Maintained by OEIS Foundation**
- **Example: F_n = 0, 1, 1, 2, 3, 5, 8, 13, 21, ...**

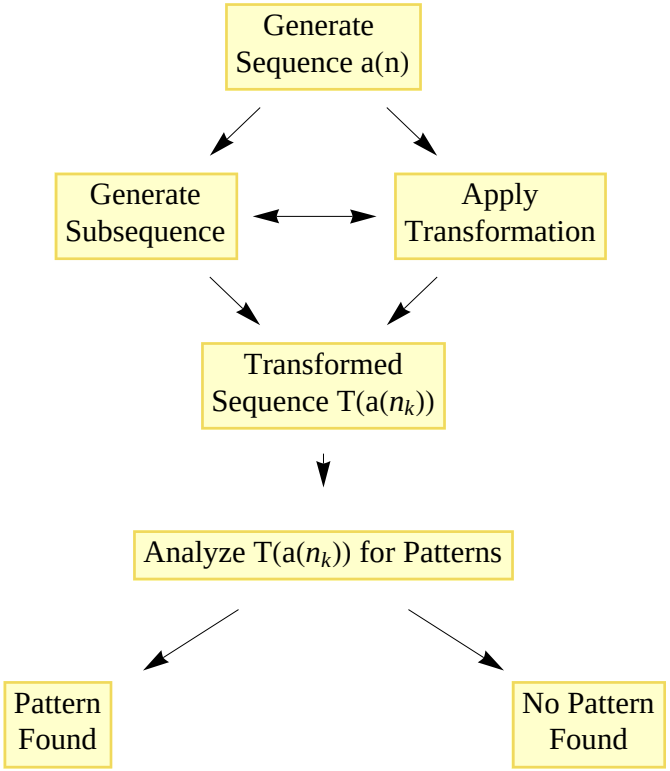
Data Mining

- **Large Scale Pattern Recognition**

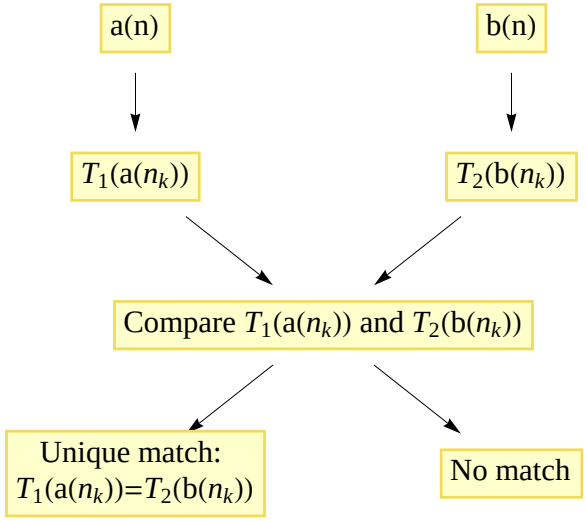
  Data mining is the process of extracting patterns from large datasets using computer science, mathematics, and statistics.

- **Mining the OEIS**
- **Number Patterns of Integer Sequences**
- **Integer Sequence Identities (Clusters)**

Number Pattern Search Algorithm for Integer Sequences



Identity Search Algorithm for Integer Sequences



Experimental Mathematics

Example 1

A000045 : Fibonacci sequence
F_n = 0, 1, 1, 2, 3, 5, 8, 13, 21, …, 39088169 (39 terms)

A000045S1T3: Sums of Squares Transformation
$\sum_{k=0}^{n} F_k^2$ = 0, 1, 2, 6, 15, 40, 104,…, 2472169789339634

A000045S1T8: Product of Consecutive Terms Transformation
$F_n F_{n+1}$ = 0, 1, 2, 6, 15, 40, 104,…, 2472169789339634

EXPERIMENTAL CONJECTURE: $\boxed{\sum_{k=0}^{n} F_k^2 = F_n F_{n+1}}$

Example 2

- **A131524: Number of possible palindromic rows in an n X n crossword puzzle**
  a_n = 0, 0, 1, 1, 2, 2, 4, 4, 7, 7, 12, ..., 121392; $n \geq 1$ (50 terms)

  A131524S2T4: Binomial Transform of a_{2n} (pad $a_0 = 0$):

  $\sum_{k=0}^{n} (-1)^k \binom{n}{k} a_{2k}$ = 0, 0, 1, 1, 2, 3, 5, 8, 13,… 4181 $(n \geq 0)$

- **A018910S1T4: Pisot sequence L(4,5)**
  b_n = 4, 5, 7, 10, 15, 23, 36, 57, ..., 165580143 $n \geq 0$ (39 terms)

  A018910S1T4: Binomial Transform of b_n:

  $\sum_{k=0}^{n} (-1)^k \binom{n}{k} b_k$ = 4, -1, 1, 0, 1, 1, 2, 3, 5, 8, 13, …, 4181, ..., 9227465 $(n \geq 0)$

  EXPERIMENTAL CONJECTURE: $\boxed{\sum_{k=0}^{n} (-1)^n \binom{n}{k} a_{2k} = \sum_{k=0}^{n+2} (-1)^{n+2} \binom{n+2}{k} b_k}$ $(n \geq 1)$

  PROBLEM: Determine an appropriate distance function (or similarity measure) to match two sequences that are similar, but not exactly the same.

EUREKA Project

GOAL: Mine the OEIS for new identities

- **Implementation: *Mathematica* and MySQL**
- **Use *Mathematica* to apply transformations to each integer sequence in OEIS and store in MySQL database**
- **Use *Mathematica* to search MySQL database for *identity clusters* (transformations which match); each cluster corresponds to a set of equivalent identities**
- **Programming Challenges**
- **Sequence matching: variations of the same sequence, offsets, false positives**
- **Extremely large integers**
- **Large data sets**
- **Parallel/distributed computing**

■ Example: Fibonacci Identity Cluster

Current Progress

- **Scope**
- **Approximately 8 million transformed sequences have been calculated (A000001-A170000)**
- **MySQL database of transformed sequences contains over 142 million rows (each row stores a term of a sequence) - 6 GB file**
- **Results so far**
- **Completed search for identity clusters involving A000001-A050000**
- **Over 400,000 matches between transformed sequences found by EUREKA**
- **Preliminary analysis shows:**
  - Most matches are trivial or already mentioned in OEIS (> 99%)
  - Small fraction of false positives (> 0.9%)

Next Steps

- **Scale up processing power and memory**
- **Perform search on a cluster of computers**
- **Implement parallel/distributed computing**
- **Improve sequence matching algorithms**
- **Reduce search-times**
- **Reduce false positives**
- **Expand Scope of Search**
- **Enlarge collection of sequence transformations**
- **Extend algorithms to 2-D sequences, rational sequences (e.g. Bernoulli numbers)**

- **Disseminate Work**
- **Create database website**
- **Publish new interesting (non-trivial) EUREKA's experimental conjectures**
- **Seek Help**
- **Need collaborators to analyze EUREKA's experimental conjectures: filter out trivial matches and false positives**
- **Need good programmers (recruit students!)**

The End

Thank you