Mining the OEIS: Ten Experimental Conjectures

Hieu D. Nguyen and Douglas Taggart

Department of Mathematics
Department of Computer Science
Rowan University

Joint Math Meetings - San Diego, CA

1/9/2013

Online Encyclopedia of Integer Sequences (OEIS)

- Searchable online database containing information on over 200,000 integer sequences: http://oeis.org

Online Encyclopedia of Integer Sequences (OEIS)

- Searchable online database containing information on over 200,000 integer sequences: http://oeis.org
- Created by Neil Sloane originally in book form the 1970's

Online Encyclopedia of Integer Sequences (OEIS)

- Searchable online database containing information on over 200,000 integer sequences: http://oeis.org
- Created by Neil Sloane originally in book form the 1970's
- Sample entry - A000045: Fibonacci sequence - $\{0, 1, 1, 2, 3, 5, ..., 39088169\}$

Mining the OEIS

- GOAL: Discover new mathematical identities involving integer sequences.

Mining the OEIS

- GOAL: Discover new mathematical identities involving integer sequences.
- Classical (manual or by hand) approach:

Mining the OEIS

- GOAL: Discover new mathematical identities involving integer sequences.
- Classical (manual or by hand) approach:
  - Great bookkeepers: Wallis, Newton, Euler, Gauss, etc.

Mining the OEIS

- GOAL: Discover new mathematical identities involving integer sequences.
- Classical (manual or by hand) approach:
  - Great bookkeepers: Wallis, Newton, Euler, Gauss, etc.
  - Use human intuition to infer number patterns

Mining the OEIS

- GOAL: Discover new mathematical identities involving integer sequences.
- Classical (manual or by hand) approach:
  - Great bookkeepers: Wallis, Newton, Euler, Gauss, etc.
  - Use human intuition to infer number patterns
- Modern (automated or by computer) approach:

Mining the OEIS

- GOAL: Discover new mathematical identities involving integer sequences.
- Classical (manual or by hand) approach:
  - Great bookkeepers: Wallis, Newton, Euler, Gauss, etc.
  - Use human intuition to infer number patterns
- Modern (automated or by computer) approach:
  - Small-scale: Use OEIS to investigate a single sequence or family of sequences

Mining the OEIS

- GOAL: Discover new mathematical identities involving integer sequences.
- Classical (manual or by hand) approach:
  - Great bookkeepers: Wallis, Newton, Euler, Gauss, etc.
  - Use human intuition to infer number patterns
- Modern (automated or by computer) approach:
  - Small-scale: Use OEIS to investigate a single sequence or family of sequences
  - Large-scale (data mining): Mine the entire OEIS database as a whole

Project Eureka

- Experimental mathematics research project aimed at mining the OEIS for new identities.

Project Eureka

- Experimental mathematics research project aimed at mining the OEIS for new identities.
- Our approach is to store integer sequences and their transformations in a database and apply an appropriate similarity measure to match sequences numerically.

Project Eureka

- Experimental mathematics research project aimed at mining the OEIS for new identities.
- Our approach is to store integer sequences and their transformations in a database and apply an appropriate similarity measure to match sequences numerically.
- Filter matches (experimental conjectures) to obtain interesting new identities.

Project Eureka

- Experimental mathematics research project aimed at mining the OEIS for new identities.
- Our approach is to store integer sequences and their transformations in a database and apply an appropriate similarity measure to match sequences numerically.
- Filter matches (experimental conjectures) to obtain interesting new identities.
- Example: The two transformations below match:
  A000045S1T1: $F_n = \{0, 1, 1, 2, 3, ..., 39088169\}$
  A000045S1T2: $\sum_{k=0}^{n} F_k = \{0, 1, 2, 4, 7, ..., 102334154\}$
  Corresponds to identity

$$\sum_{k=0}^{n} F_k = F_{n+2} - 1 \quad (1)$$

List of Transformations

Symbol (Txx)	Transformation Name	Formula
T1	Identity	a_n
T2	Partial Sums	$\sum_{k=0}^{n} a_k$
T3	Partial Sums of Squares	$\sum_{k=0}^{n} a_k^2$
T4	Inverse Binomial Transform	$\sum_{k=0}^{n} (-1)^n \binom{n}{k} a_k$
T5	Self-Convolution	$\sum_{k=0}^{n} a_k a_{n-k}$
T6	Linear Weighted Partial Sums	$\sum_{k=0}^{n} k a_k$
T7	Binomial	$\sum_{k=0}^{n} \binom{n}{k} a_k$
T8	Product of Two Consecutive Elements	$a_k a_{n-k}$
T9	Cassini	$a_{n-1} a_{n_1} - a_n^2$
T10	First Stirling	$\sum_{k=0}^{n} s(n,k) a_k$
T11	Second Stirling	$\sum_{k=0}^{n} S(n,k) a_k$
T12	Boustrophedon	$\sum_{k=0}^{n} \binom{n}{k} E_{n-k} a_k$
T13	First Differences	$a_n - a_{n-1}$
T14	Catalan	$\sum_{k=0}^{n} \frac{k}{n} \binom{2n-k-1}{n-k} a_k$
T15	Hankel	$\det(a_{i+j})_{i,j=0}^{n}$
T16	Sum of Divisors	$\sum_{d\mid n} a_d$
T17	Moebius	$\sum_{d\mid n} \mu(n/d) a_d$

MySQL

- Apply T1-T17 to A000001-A170000

MySQL

- Apply T1-T17 to A000001-A170000
- Over 3 million sequence transformations

MySQL

- Apply T1-T17 to A000001-A170000
- Over 3 million sequence transformations
- Terms are stored in a MySQL table using a window format

MySQL

- Apply T1-T17 to A000001-A170000
- Over 3 million sequence transformations
- Terms are stored in a MySQL table using a window format
- Table contains over 100 millions rows

Table: Sequence Transformations - Sample Entries

ID	Label	Position	EntryOne	EntryTwo	EntryThree
1	A000045S1T1	0	0	1	1
2	A000045S1T1	1	1	1	2
3	A000045S1T1	2	1	2	3
4	A000045S1T1	3	2	3	5
...	...	...	...	...	...
38	A000045S1T1	37	24157817	39088169	Null
39	A000045S1T1	38	39088169	Null	Null

Mathematica

- Challenges with matching sequences:

Mathematica

- Challenges with matching sequences:
    - Sequences stored in OEIS vary in length from 4 to 100 terms

Mathematica

- Challenges with matching sequences:
  - Sequences stored in OEIS vary in length from 4 to 100 terms
  - Many sequences have the same initial terms 0 and 1.

Mathematica

- Challenges with matching sequences:
  - Sequences stored in OEIS vary in length from 4 to 100 terms
  - Many sequences have the same initial terms 0 and 1.
  - Sequences may be shifts, translations or scalar multiples (or all three) of one another as illustrated by previous Fibonacci identity.

Mathematica

- Challenges with matching sequences:
  - Sequences stored in OEIS vary in length from 4 to 100 terms
  - Many sequences have the same initial terms 0 and 1.
  - Sequences may be shifts, translations or scalar multiples (or all three) of one another as illustrated by previous Fibonacci identity.
- Match sequences using a similarity measure based on *head-bites-tail* (HBT) overlap $L_{\max}$ and relative HBT distance d_r.

Mathematica

- Challenges with matching sequences:
  - Sequences stored in OEIS vary in length from 4 to 100 terms
  - Many sequences have the same initial terms 0 and 1.
  - Sequences may be shifts, translations or scalar multiples (or all three) of one another as illustrated by previous Fibonacci identity.
- Match sequences using a similarity measure based on *head-bites-tail* (HBT) overlap $L_{\max}$ and relative HBT distance d_r.
- Match parameters:
  - $L_{\max} \geq 4$
  - $d_r \leq 1/2$

Linear Matches

Definition

Two sequences $\{a_n\}$ and $\{b_n\}$ are said to be *linear* if there exists constants s, t, and C such that

$$sa_n + tb_n = C \tag{2}$$

Linear Matches

Definition

Two sequences $\{a_n\}$ and $\{b_n\}$ are said to be *linear* if there exists constants s, t, and C such that

$$sa_n + tb_n = C \tag{2}$$

Lemma

Let a_n and b_n be two non-trivial finite sequences with first differences $\Delta a_n = a_{n+1} - a_n$ and $\Delta b_n = b_{n+1} - b_n$, respectively. Moreover, let $A = GCD\{\Delta a_n\}$ and $B = GCD\{\Delta b_n\}$. Then

$$\frac{\Delta a_n}{A} = \frac{\Delta b_n}{B} \tag{3}$$

if and only if a_n and b_n are linear.

Search Run Times

Table: Search Run Times Based on Window Size

Window Size (Number of Terms)	Run Time (Days)
1	38.96
2	3.5
3	2.67

Table: Search Run Times Based on Computer Model

Computer (Model/Year)	Configuration (Processor/RAM)	Run Time (Days)
Apple iMac (mid-2011)	2.7 GHz Intel Core i5 quad-core 4 GB RAM	2.67
Apple Mac Pro (mid-2010)	3.2 GHz Intel Xeon quad-core 32 GB RAM	0.62

Current Results

- Over 300,000 linear matches found.

Current Results

- Over 300,000 linear matches found.
- Large fraction of matches are either known, redundant or trivial, e.g. A000045S1T1 (F_n) $\sim$ A000071S1T1 ($F_n - 1$).

Current Results

- Over 300,000 linear matches found.
- Large fraction of matches are either known, redundant or trivial, e.g. A000045S1T1 (F_n) $\sim$ A000071S1T1 ($F_n - 1$).
- Matches are stored in a MySQL table, publicly available at Eureka database website: http://elvis.rowan.edu/datamining/eureka

Table: Sample linear match: A000045S1T1 $\sim$ A000045S1T2

ID	Label1	Label2	Overlap	Distance	Scaling	Translation	Shift
2087	A000045S1T1	A000045S1T2	34	0.02857	1	1	-2

$$\sum_{k=0}^{n} F_k = F_{n+2} - 1$$

Ten Experimental Conjectures

- Present a sample of ten experimental conjectures (linear matches) that we believe to be new, interesting, and not mentioned on OEIS website.

Ten Experimental Conjectures

- Present a sample of ten experimental conjectures (linear matches) that we believe to be new, interesting, and not mentioned on OEIS website.
- Many conjectures are suitable for advanced undergraduate math students to investigate and hopefully develop into research projects.

Ten Experimental Conjectures

- Present a sample of ten experimental conjectures (linear matches) that we believe to be new, interesting, and not mentioned on OEIS website.
- Many conjectures are suitable for advanced undergraduate math students to investigate and hopefully develop into research projects.
- All conjectures can be accessed on the Eureka database website using its search engine.

Conjecture 1

A002212S1T15 $\sim$ **A032908S1T1** ($L_{\max} = 10$, $d_r = 0.43$)

$$\det[(a_{i+j})_{i,j=0}^{n}] = b_{n+1} - 1 \qquad (4)$$

where

- a_n = A002212 - Number of restricted hexagonal polyominoes with n cells.
- b_n = A032908 - One of 4 3rd-order recurring sequences for which the first derived sequence and the Galois transformed sequence coincide.

Conjecture 2

A004441S1T12 $\sim$ **A065619S1T7** ($L_{\max} = 21$, $d_r = 0.45$)

$$\sum_{k=0}^{n} \binom{n}{k} E_{n-k} a_k = \sum_{k=0}^{n} \binom{n}{k} b_k \tag{5}$$

where

- a_n = A004441 - Numbers that are not the sum of 4 distinct nonzero squares.
- b_n = A065619 - E.g.f. $x(\tan(x) + \sec(x))$.

Conjecture 3

A008410S1T17 $\sim$ **A022523S1T2** ($L_{\max} = 16,\ d_r = 0.16$)

$$\sum_{d|n} \mu(n/d) a_d = 480 \sum_{k=0}^{n-1} b_k \tag{6}$$

where

- a_n = A008410 - a(0) = 1, $a(n) = 480\sigma_7(n)$, where $\sigma_7(n)$ is the sum of divisors function.
- b_n = A022523 - Nexus numbers $(n+1)^7 - n^7$.

Conjecture 4

A026375S1T5 $\sim$ **A144180S1T10** ($L_{\max} = 17$, $d_r = 0.11$)

$$\sum_{k=0}^{n} a_k a_{n-k} = \frac{5}{4} \sum_{k=0}^{n} s(n,k) b_k - \frac{1}{4} \tag{7}$$

where

- a_n =A026375 - $a(n) = \sum_{k=0}^{n} \binom{n}{k} \binom{2k}{k}$.
- b_n = A144180 - Number of ways of placing n labeled balls into n unlabeled (but 5-colored) boxes.

Conjecture 5

A037164S1T17 $\sim$ **A022527S1T2** ($L_{\max} = 11$, $d_r = 0.19$)

$$\sum_{d|n} \mu(n/d) a_d = \sum_{k=0}^{n-1} b_k \qquad (8)$$

where

- $a_n =$ A037164 - Numerators of coefficients of Eisenstein series $E_12(q)$ (or $E_6(q)$ or $E_24(q)$).
- $b_n =$ A022527 - Nexus numbers $(n+1)^{11} - n^{11}$.

Conjecture 6

A046055S1T3 $\sim$ **A018903S1T9** ($L_{\max} = 16$, $d_r = 0.35$)

$$\sum_{k=0}^{n-1} a_k^2 = \frac{b_{n-1}b_{n+1} - b_n^2 - 13}{3} \tag{9}$$

where

- a_n = A046055 - Orders of finite Abelian groups having the incrementally largest numbers of nonisomorphic forms (A046054).
- b_n = A018903 - Define the sequence $S(a_0, a_1)$ by a_{n+2} is the least integer such that $a_{n+2}/a_{n+1} > a_{n+1}/a_n$ for $n >= 0$. This is $S(1, 5)$.

Conjecture 7

A098411S1T15 $\sim$ **A139685S1T8** ($L_{\max} = 8,\ d_r = 0.16$)

$$\det[(a_{i+j})_{i,j=0}^{n}] = \frac{1}{2} b_n b_{n+1} \tag{10}$$

where

- a_n = A098411 - Expansion of $1/(\sqrt{1-4x} \cdot \sqrt{1-12x})$.
- b_n = A139685 - Number of n x n symmetric binary matrices with no row sum greater than 9.

Conjecture 8

A122162S1T17 $\sim$ **A008384S1T2** ($L_{\max} = 26$, $d_r = 0.05$)

$$\sum_{d|n} \mu(n/d) a_d = \sum_{k=0}^{n-1} b_k \tag{11}$$

where

- a_n = A122162 - Coefficient of q-series for constant term of Tate curve.
- b_n = A008384 - Crystal ball sequence for A_4 lattice.

Conjecture 9

A169344S1T7 $\sim$ **A152262S1T9** ($L_{\max} = 13$, $d_r = 0.07$)

$$\sum_{k=0}^{n} \binom{n}{k} a_k = \frac{43}{252}(b_{n-1}b_{n+1} - b_n^2) - \frac{1}{42} \tag{12}$$

where

- a_n =A169344 - Number of reduced words of length n in Coxeter group on 43 generators S_i with relations $(S_i)^2 = (S_i S_j)^{30} = I$.
- b_n = A152262 - $a(n) = 14 * a(n-1) - 43 * a(n-2)$, $n > 1$; $a(0) = 1$, $a(1) = 7$.

Conjecture 10

$$a_n = \sum_{k=0}^{\infty} a(k)b(n-25k) \tag{13}$$

- a_n = A000009 = {1, 1, 1, 2, 2, 3, 4, 5, 6, 8, 10, ..., 89, 104, 122, 142, 165, 192, ..., 5718} - Expansion of $\prod_{m=1}^{\infty}(1+x^m)$; number of partitions of n into distinct parts; number of partitions of n into odd parts.
- b_n =A034320 = {1, 1, 1, 2, 2, 3, 4, 5, 6, 8, 10, ..., 89, 104, 122, 141, 164, 191, ..., 6082} - McKay-Thompson series of class 50a for the Monster group with a(0) = 1.
- c_n = A058703 = {1, 0, 1, 2, 2, 3, 4, 5, 6, 8, 10, ..., 89, 104, 122} - McKay-Thompson series of class 50a for Monster.

Work in Progress

- Python implementation (open-source)

Work in Progress

- Python implementation (open-source)
- Develop better algorithms to filter interesting matches

Work in Progress

- Python implementation (open-source)
- Develop better algorithms to filter interesting matches
- Mine remaining set of integer sequences in OEIS (A170001-A200000)

Work in Progress

- Python implementation (open-source)
- Develop better algorithms to filter interesting matches
- Mine remaining set of integer sequences in OEIS (A170001-A200000)
- Mine fractional sequences, e.g. Bernoulli numbers

References

Eureka website: elvis.rowan.edu/datamining/eureka.

Mining the Online Encyclopedia of Integer Sequences (preprint), available at: www.rowan.edu/colleges/csm/departments/math/facultystaff/nguyen

The Online Encyclopedia of Integer Sequences website: oeis.org.