Equation Hypergraph

Benedict W. J. Irwin[1,2]

[1]Theory of Condensed Matter, Cavendish Laboratories, University of Cambridge, Cambridge, United Kingdom
[2]Optibrium, F5-6 Blenheim House, Cambridge Innovation Park, Denny End Road, Cambridge, CB25 9PB, United Kingdom
ben.irwin@optibrium.com

July 30, 2020

Abstract

We find a graph representation for equations.

Combinatoric objects have many identities. Furthermore, there are deep and interesting patterns in mathematical equations. We propose a hypergraph to link all of this information. For example the binomial equation

$$(x + a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}$$

links the node $(x + a)^k$ to the product hypernode $\{\binom{n}{k}, x^k, a^{n-k}\}$, with the labelled hyperedge $\sum_{k=0}^{n}$.

We see these cycles in equations for transformations between 'bases' in terms of Stirling numbers and Lah numbers. This also relates to Bell numbers, factorials etc.

The key thing here, would be a query to the hypergraph, which says "Express \$node in terms of \$othernode". Additions may be "but not using \$restrictednode", or additional boolean operations.

Clearly 'Objects' here are either functions, which have input(s) and output(s) or functionals which have both arguments as well as additional inputs, i.e. sum the summand from a to b.

Functional nodes/edges are related because "sum the summand from a to b" is a parent to "sum the summand from 0 to ∞" with the edge label "$a \to 0$ and $b \to \infty$".

The question is what data is required to form such a graph. Can machine learning tools be used to assist in node creation? I.e. by image recognition and scanning of Wikipedia/Papers etc. and how do we check for errors? What is the complexity of maintaining such a large collection of identities and traversing the hypergraph?

Can we define new special functions and polynomials? Can the algorithm figure out which edge addition would give a swift and useful result? How do we cancel out terms in the hypernodes? I.e. to simplify expressions? To what extent does Mathematica already do this? How do we represent symmetry? For example $(a + b) = (b + c)$ as $SumSet\{a, b\}$.

Even having a lookup that gives equations that link objects together would be very useful.

For unknown equations, for example if a mathematician just discovers a formula, a routine that tries to express that formula in terms of human readable equations, or human defined terms would be ideal!

Notation Example

```
N = define node {TeX: "(x+a)^{k}"}
S = define node {TeX: "\sum_{a}^b"}
E = define edge {N,"=",S,pset(A,B,C)}
```

Really, we just want the language to be something like mathematica...

```
Power[x_,n_]:=x^n
Add[a_,b_]:=a+b
Define Add[a_,b_]=Add[b,a]
Define Power[Add[x,a],n&Integer]=Sum[Binomial[n,k]Power[x,k]Power[a,n-k],{k,0,n}]
```

They key thing, is that we will trawl the OEIS to get identities. Another source may be Math Stack exchange... (this is likely to be somewhat corrupted with incorrect information, or equations that have strict bounds which are not mentioned).

Obviously the hypergeometric definitions of functions will immediately link many generating functions together.

One interesting idea is that in maths we often have functions of more than one input i.e. $\binom{n}{k}$ but we rarely use simple objects of the form $(f, g)(x)$ as opposed to $f(x, y)$. Consider the solution to a differential equation which has two solutions... perhaps this is actually a fundamental object with two outputs, therefore. Roots are also examples of these.

We might want to limit hypergraph walks from accessing the identity. Otherwise random parts of mathematics could be connected. This would be a query option.

Other nodes would surely be functions for example $e^x \to e^{-x}$ with the edge $x \to -x$. Then a common edge might be an integral transform, for example $e^{-x} \to \Gamma(s)$, where the edge is the Mellin transform.

Thus, asking the question, how does e^x link to $\Gamma(s)$, one possible route through the graph is $x \to -x, \mathcal{M}_{x \to s}$. Likewise, $e^x \to \frac{1}{k!}$ with the edge being "ordinary generating function of". We then also have $e^x \to 1$ as "exponential generating function of".

$$\text{GammaDistribution}(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, x > 0, \alpha, \beta > 0$$

so we have an edge $\Gamma(x) \to \frac{1}{\Gamma(x)}$. But we also have a node representing $\Gamma(x), x > 0 \to \Gamma(x)$ where the edge is of type "drop conditions".

What we want to avoid is having an enormous number of nodes, i.e. $1, 2, 3, 4, \cdots$ an so on just to explain integers and rational numbers etc. We need a way of finding $3x$ is the same as cx for a real constant. It should be fine to allow nx for integer constant to be separate and ncx for integer n and real c etc. as well as complex numbers. Thus p/q is important for integer p and q, and that should really lead to only a few categories of positive and negative and all rationals etc.

Even with this being said, a single equation with 3 parameters, for example, the gamma distribution, may take on many permutations of each parameter belonging to each set of objects.

$$
\begin{array}{ccccccccc}
7 & \leftarrow & 14 & \leftarrow & 21 & \leftarrow & 28 & \leftarrow & 35 \\
\downarrow & & & & & & & & \\
5 & \leftarrow & 10 & \leftarrow & 15 & \leftarrow & 20 & \leftarrow & 25 \\
\downarrow & & & & & & & & \\
3 & \leftarrow & 6 & \leftarrow & 9 & \leftarrow & 12 & \leftarrow & 18 \\
\downarrow & & & & & & & & \\
2 & \leftarrow & 4 & \leftarrow & 8 & \leftarrow & 16 & \leftarrow & 32 \\
\downarrow & & & & & & & & \\
1 & & & & & & & &
\end{array}
$$