There ain't no cure for math

April 19, 2013 - 6:42pm **BY JASON BROWN**

I was sitting at the Leonard Cohen concert last Saturday, up in the nosebleed section, but even at that height, I could appreciate a master at his craft.

As both a songwriter and a mathematician, I can't help but analyze the tunes as they float up to the rafters of the Metro Centre. And while the music is very good, I continue to be bowled over by Leonard's lyrics. He certainly is one of a kind.

But what makes him that? What makes Cohen so "cohen-y," so inherently different from his contemporaries like Paul Simon and Bob Dylan? Could one distinguish one songwriter's words from another? Now that sounds like the beginning of a mathematical problem to me!

Indeed there has been interest in mathematically identifying characteristics that distinguish one writer from another. One common technique was to look at the richness of the vocabulary used by each author, coupled with the difference in use of common words.

This brings me back to my work at Dalhousie. One of the many parts of my job at the university is as a reader on thesis defence committees. Graduate students, at the doctoral level, have to write a thesis on some research they have carried out over x number of years. As a member of the defence committee, I was to be one of three examiners of the thesis.

Sometimes it is a bit of an onerous task (a thesis can be a couple of hundred pages or more), but oftentimes surprises are to be had.

A couple of months ago, I was asked to be a thesis reader for a PhD student in computer science. The work was about modelling the essence of music, finding a way to statistically isolate what makes one composer's songs different from another's. The work is based on very interesting research that a colleague of mine at Dalhousie, Vlado Keselj, and some colleagues did on a new way to distinguish the writings of authors.

So I located a copy of their original article and read it. And the ideas therein are really interesting!

So while you might think that to recognize patterns in the work of a writer, you might have to consider what words they use, the idea put forward is that you can forget about the words — remove all the spaces, push the letters together, and then collect statistics about what a called n-grams, sequences of n consecutive letters. So, for example, in the sentence "The Leonard Cohen concert was great," the 2-grar start off as th,he,el,le … and the 3-grams start off as the, hel,ele,leo,eon …

Certainly among these incomprehensible streams, there is no author information, right? Wrong! It turns out that the frequencies of these grams can be used very often to distinguish between authors. Moreover (and this is the part I love!), you can use the process for Asian languages where there is often no separation between words.

I haven't had time yet to see how Leonard's n-grams might differ from Paul's or Bob's, but I would expect the statistics to be different among them. And anyways, the whole approach is a fascinating insight into how writers might be distinguishable on a content-less but mathematical basis.

On a more sombre note, I read recently that Google has produced a software tool to keep a digital imprint of people after death. And Leonard Cohen did warn all of us at the concert that we are all on a farewell tour. But perhaps all we need for posterity is to keep a jar of our n-grams on the mantelpiece!

Jason I. Brown is a professor of mathematics at Dalhousie University in Halifax. He is the author of Our Days Are Numbered: How Mathematics Orders Our Lives.

About the Author »

By JASON BROWN

0 comments • 2 reactions

Comments for this thread are now closed.

Oldest Community Share

No one has commented yet.

Comment feed Subscribe via email