The *Communications* Web site, <http://cacm.acm.org>, features more than a dozen bloggers in the BLOG@CACM community. In each issue of *Communications*, we'll publish selected posts or excerpts.

Follow us on Twitter at <http://twitter.com/blogCACM>

DOI:10.1145/3131066

<http://cacm.acm.org/blogs/blog-cacm>

Manipulating Word Representations, and Preparing Students for Coding Jobs?

Robin K. Hill mulls an aspect of natural language processing research, while Mark Guzdial ponders why coding is taught in public schools.

Robin K. Hill
Deep Dictionary

<http://bit.ly/2tRnVZN>

June 20, 2017

Recent research in natural language processing using the program *word2vec* gives manipulations of word representations that look a lot like semantics produced by vector math. For vector calculations to produce semantics would be remarkable, indeed. The word vectors are drawn from context, big, huge context. And, at least roughly, the meaning of a word is its use (in context). Is it possible some question is begged here?

We represent words by vectors (one-dimensional arrays) of a large number of elements, the numeric values of which are determined by reading and processing a vast number of examples of context in which those words appear, and which are functions of the distances between occurrences of words in sequences. A neural network

refines the values of each element of each word vector in the training phase. The network, through iterated adjustment of the elements of the vector based on errors detected on comparison with the text corpora, produces the values in continuous space that best reflect the contextual data given. The end result is the word vector, a lengthy list of real numbers that do not seem to have any particular interpretation that would reflect properties of the thing itself, nor properties of words.

In the example provided by Mikolov et al.,¹ they "... assume relationships are present as vector offsets, so that in the embedding space, all pairs of words sharing a particular relation are related by the same constant offset." Let's use 'd' to stand for "distributed representation by vector" of a word.

d("king")	=	v1
d("male")	=	v2
d("female")	=	v2
d("queen")	=	v4

Mikolov and his colleagues find that $v1 - v2 + v3 \approx v4$ (in vector offset mathematics in the n-dimensional space). This is certainly an intriguing result, in accord with our understanding of the meanings of the four words, in which taking "king," removing the "male" aspect, and replacing it with a "female" aspect, gives "queen."

The word vectors can also capture plurality and other shades of meaning the researchers regard as syntactic. For example, the offset between singular and plural is a (learned) constant:

$d(\text{"apple"}) - d(\text{"apples"}) \approx d(\text{"families"}) - d(\text{"family"}) \approx d(\text{"cars"}) - d(\text{"car"})$

More details of *word2vec* can be found in the explanation by Xin Rong.³ It looks like, not by direct coding but by some fortuitous discovery, the system has figured out some mathematical analog for semantics. (There is no claim that individual elements of the vector are capturing features such as gender, status, or grammatical number.)

We already have a compendium of data on relationships of words to other words through contexts of use: the dictionary. The use of a word is largely given by its context. Its context can be inferred also from its dictionary definition. Most dictionaries will offer a direct or indirect connection through "king" to "ruler" or "sovereign" and "male" and through "queen" to "ruler" or "sovereign" and "female," as :

queen

The female ruler of an independent state, especially one who inherits the position by right of birth

king

The male ruler of an independent state, especially one who inherits the position by right of birth

ruler

A person exercising government or dominion

These definitions² show gender can be “factored out,” and in common usage the gender aspect of sovereigns is notable. We would expect those phenomena to show up in vast text corpora. In fact, we would expect that to show up in text corpora *because* of the dictionary entries. Since we base word use on definitions captured by the dictionary, it is natural for any graph-theoretic distance metric based on node placement to (somehow) reflect that cross-semantic structure.

Suppose that, employing the English slang terms “gal” and “guy” for male and female, the word for queen were “ruler-gal,” and for king “rulguy,” (perhaps the word for mother were “parentgal,” and for father, “parentguy”). Then the word vector offsets calculated would not appear as remarkable, the relationships exposed in the words themselves.

The system *word2vec* constructs and operates through the implicit framework of a dictionary, which gave rise to the input data to *word2vec*. How could it be otherwise? As we understand the high degree of contextual dependency of word meanings in a language, any representation of word meaning to a significant degree will reflect context, where context is its interassociation with other words.

The result is still intriguing. We have to ask how co-occurrence of words can reliably lay out semantic relationships. We might explore the aspects of semantics missing from context analysis, if any. We might (and should) ask what sort of processing of a dictionary would deliver the same sort of representations, if any.

The word vectors produced by the method of training on a huge natural text dataset, in which words are given distributed vector representations refined through associations present in the input context, reflect the cross-referential semantic compositionality of a dictionary.

That close reflection is derived from the fact that words in natural text will be arranged in accordance with dictionary definitions. The *word2vec* result is revelation of an embedded regularity.

References

1. Mikolov, T., Chen, K., Corrado, C., and Dean, J. Efficient Estimation of Word Representations in Vector Space, <https://arxiv.org/abs/1301.3781>.
2. Oxford Living Dictionary, 2017, Oxford University Press, <https://en.oxforddictionaries.com/>.
3. Rong, X. *word2vec* Parameter Learning Explained, <https://arxiv.org/abs/1411.2738>.

Mark Guzdial
Coding in Schools as
New Vocationalism:
Larry Cuban on
What Schools Are For
<http://bit.ly/2tpSgip>
July 18, 2017

Larry Cuban is an educational historian who has written before on why requiring coding in schools is a bad idea (<http://bit.ly/2uocMLQ>). Jane Margolis and Yasmine Kafai wrote an excellent response about the importance of coding in schools (<http://bit.ly/2vmi52U>). Cuban penned a three-part series about “Coding: The New Vocationalism,” likely inspired by a recent *New York Times* article about the role tech firms are having on school policy (<http://nyti.ms/2uodaKi>).

► In Part 1 (<http://bit.ly/2v2ULoo>), he describes the ‘dance’ schools have had with industry over more than 100 years, between preparing future *citizens* and preparing future *workers*.

Preparation for the workplace is not the only goal for public schooling, yet that has been the primary purpose for most reformers over the past three decades. A century ago, reformers also elevated workplace preparation as the overarching purpose for tax-supported public schools.

In the new vocationalism, Cuban sees schools have been tied to economic growth and the needs of information-age society. He sees coding advocates blending the roles of school in preparing citizens and school as preparing workers by arguing that computing is necessary for modern society.

► In Part 2 (<http://bit.ly/2vmjCG8>), he points out any education reform faces the reality of what teachers *know* and what they will actually *do* in the classroom. He draws on efforts in the 1950s and 1960s, and uses the story of Logo to explain how reformers get it wrong.

The lessons to be learned from earlier school reformers are straightforward.

► *Build teacher capabilities in content and skills since both determine to what degree, if any, a policy gets past the classroom door.*

► *With or without enhanced capabilities and expertise, teachers will adapt policies aimed at altering how and what they teach to the contours of the classrooms in which they teach. If policymakers hate teacher fingerprints over innovations, if they seek fidelity in putting desired reforms into practice, they wish for the impossible.*

► *Ignoring both of the above lessons ends up with incomplete implementation of desired policies and sorely disappointed school reformers.*

► In Part 3 (<http://bit.ly/2wpo9o8>), he returns to the question of what school is for. He describes successful reform as a collaboration between top-down designers and policy-makers and bottom-up teachers. He describes a successful model for reform that created “work circles” of researchers and teachers (at Northwestern University) to achieve the goals of the researchers’ curriculum by adapting it with the help of the teachers.

Cuban is not necessarily against teaching computing in schools; he says it doesn’t make sense to impose it as a mandate from industry. More importantly, he offers a path forward: *mutual adaptation* of curricular goals, between designers and teachers.

Mutual adaptation can benefit teachers and students. While this is only one study of four teachers wrestling with teaching a science unit, it is suggestive of what can occur.

Will similar efforts involve teachers and make the process of mutual adaptation work for both teachers and students? I have yet to read of such initiatives as districts and states mandating computer science courses and requiring young children to learn to code. Repeating the errors of the past and letting mutual adaptation roll out thoughtlessly has been the pattern thus far. The “New Vocationalism,” displaying a narrowed purpose for tax-supported public schools, marches on unimpeded.

Robin K. Hill is an adjunct professor in the Department of Philosophy at the University of Wyoming. Mark Guzdial is a professor in the College of Computing at Georgia Institute of Technology.

Copyright held by authors.