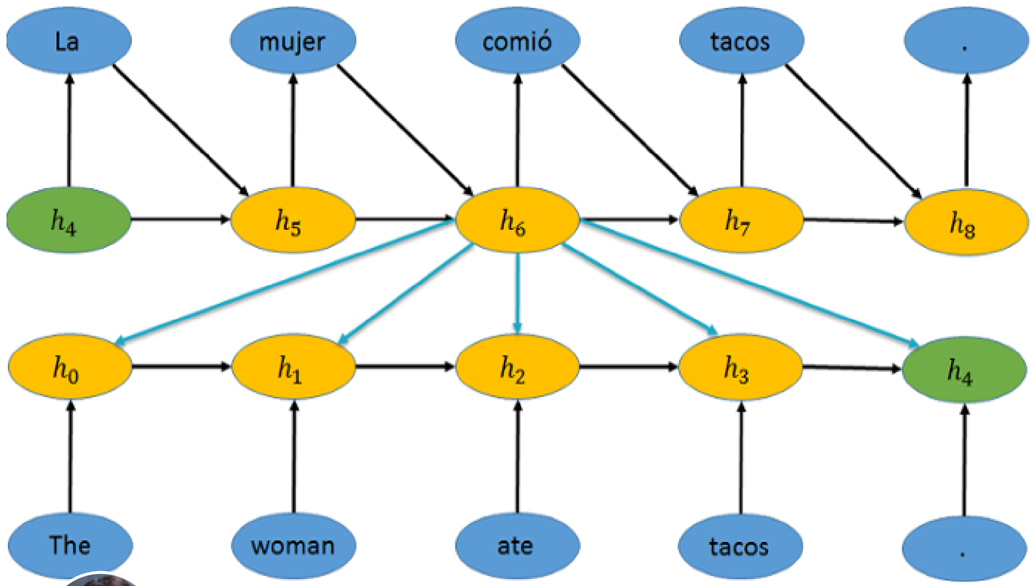



Jonathan Mugan
Computer scientist: machine learning and data science



# Deep Learning for Natural Language Processing



*Content originally posted at http://21ct.com/blog/.*

Before deep learning, the meaning embedded in the words we write was communicated to computers using human-engineered symbols and structures. In this blog post, I first review the symbolic methods WordNet, ConceptNet, and FrameNet as a contrast from which to understand the capabilities of deep learning. I then discuss how deep learning represents meaning using vectors and how vectors allow for more flexible representations. I then discuss how meaning encoded in vectors can be used to translate languages and even generate captions for images and answer questions about text. I conclude with a discussion of what deep learning still needs to truly understand human language.

WordNet [1] from Princeton University is probably the most famous corpus of symbolic meaning. It groups words together when they have the same meaning and represents hierarchical links between groups. For example, it says that a car is the same thing as an automobile and that both are types of vehicles. ConceptNet [2] is a semantic network that comes from MIT. It represents broader relationships than does WordNet. For example, ConceptNet says that bread is typically found near a toaster. However, there are too many interesting relationships to write down. Ideally, we'd like to represent that toasters are things into which one should not insert a fork.

FrameNet [3] is a project from Berkeley University that attempts to catalog meaning using frames. Frames represent concepts and their associated roles. As we saw in my last blog post, a child's birthday party frame has roles for its different parts, such as venue, entertainment, and sugar source. Another frame is that of a "buying" event where there is a seller, a buyer, and a good being sold. Computers can "understand" text by searching for keywords that trigger frames. These frames have to be manually created by humans, and their trigger words need to be manually associated. We can represent a lot of knowledge this way, but it is hard to explicitly write everything down. There is just too much, and writing it down makes it too brittle.

Symbols can also be used to create language models that compute the probability of the next word in a sentence. For example, the probability that the next word is "tacos" given that I have written "I ate" is the number of times a corpus of text has "I ate tacos" divided by the number of times the corpus has the text "I ate." Such models can be useful, but we all know that tacos are pretty similar to chimichangas, at least compared to toasters, but this model doesn't take advantage of that similarity. There are lots of words, to store all of the triples you need storage the size of (*num words* by *num words* by *num words*), and this illustrates the problem with using symbols, there's just too many of them and too many combinations. There needs to be a better way.

## Representing meaning using vectors

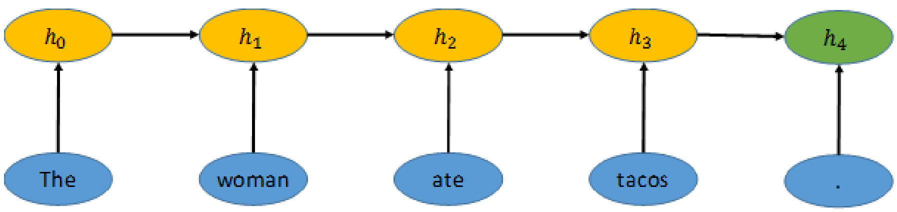
Deep learning represents meaning using vectors, so that instead of representing a concept using a monolithic symbol, a concept is represented as a large vector of feature values. Each index of the vector represents some learned feature of the neural network, and vectors are usually around length 300. This is a more efficient way of representing concepts because the concept now is composed of features [Bengio and LeCun, 2007]. Whereas two symbols can only be the same or not the same, two vectors have a degree of similarity. The vector for "taco" will be similar to the vector for "burrito", and both vectors will be very different from the vector for "car." Similar vectors can be grouped together like in WordNet.

Vectors even have internal structure so that if you take the vector for Italy and subtract the vector for Rome, you get a vector that is close to the vector you get when you subtract the vector for Paris from the vector for France [Mikolov et al., 2013]. We can write this as an equation as Italy – Rome = France – Paris. Another example is King – Queen = Man – Woman.

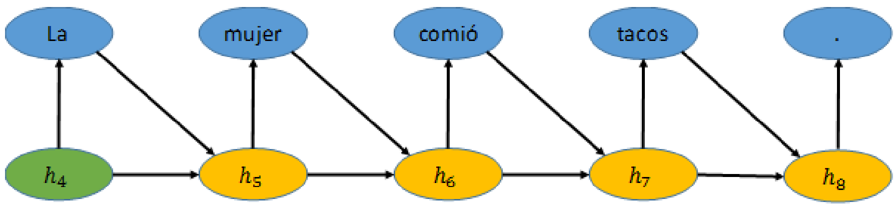
We learn vectors with these properties by training neural networks to predict the words that surround each word [Mikolov et al., 2013]. You can download already learned vectors from Google [4] or Stanford [5], or you can learn your own using the Gensim software library [6]. It's surprising that this works, and that the word vectors have such intuitive similarity and relationships, but it does work.

## Composing meaning from word vectors

We have vectors that represent the meanings of individual words, how can we compose meaning from words and even write sentences? We use something called a recurrent neural network (RNN), shown below. The RNN encodes the sentence "The woman ate tacos." into a vector represented by *h4*. The word vector for the word "The" is taken to be *h0*, and then the RNN combines *h0* and the word vector for the word "woman" to get *h1*. The vector *h1* then feeds along with the word vector for "ate" into *h2*, and so on, all the way to *h4*. The vector *h4* then represents the entire sentence.




Once information is encoded into a vector, we can decode that information into something else [Cho et al., 2014], as shown below. For example, the RNN then translates (decodes) the sentence encoded into vector *h4* into Spanish. It does this by taking the encoded vector *h4* and generating the most likely word given that vector. That generated word "La" and *h4* are then used to generate the new state vector *h5*. Given vector *h5*, the RNN then generates the most likely next word, "mujer." This process continues until a period is generated and the network stops.



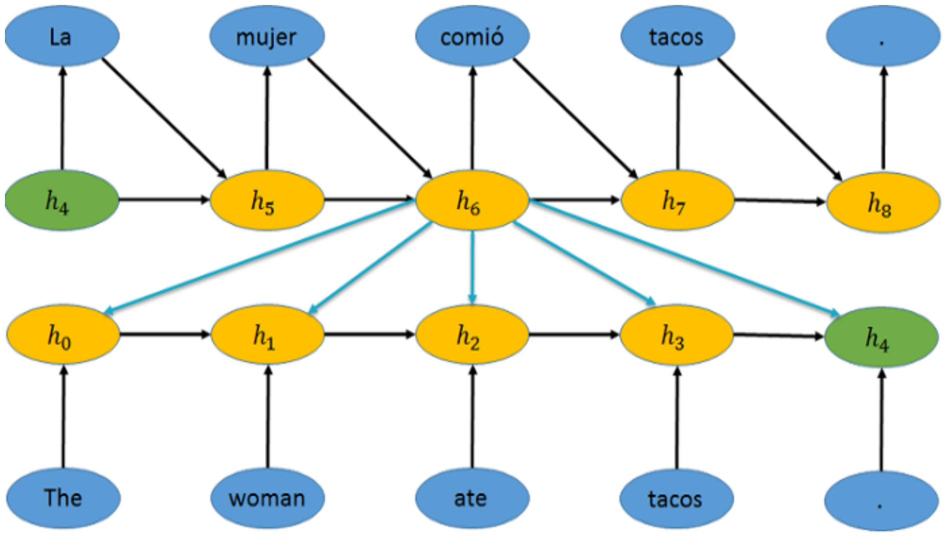

To do language translation using this encoder-decoder model, RNNs are trained on large corpuses of data where there are sentences translated between a source and a target language. These RNNs often have nodes that are quite complex [Hochreiter and Schmidhuber, 1997], and the whole model can have millions of parameters that need to be learned.

We can make the output of the decoding any sequence, such as a parse tree [Vinyals et al., 2014], or even captions for images if you have enough examples of images with captions. To generate captions for images, you train a neural network on images to identify the objects in images. Then you take the weights from the top part of that neural network as a vector representation of the image and feed that into the decoder to generate the description [Karpathy and Fei-Fei, 2014; Vinyals et al., 2014]. (For examples, see [7].)

## From compositional meaning to attention, memory, and question answering

It's tempting to think of the encoder-decoder method as a parlor trick, but we are slowly making our way toward real intelligence. We can think about decoding as answering the question, "What is the translation of this phrase?" Or, given the sentence to translate, and the words already written, "What should be the next word?"

To answer questions, the algorithm needs to remember facts. In the examples we've seen previously, it only remembers the current vector state *h* and the last word written. But what if we wanted to give it access to everything it had learned and seen? In the machine translation example, this would mean being able to go back and look at vectors *h0*, *h1*, *h2*, and *h3* while it was deciding which word to write next. Bahdanau et al. [2014] created an architecture that does just that, depicted in the image below. The network learns how to determine which previous memory is most relevant at each decision point. We can think of this as a focus of attention over memories.




What this means is that since we can encode concepts and sentences as vectors, and we can use large numbers of vectors as memories and search through to find the right answer to questions, deep learning can answer questions from text. One example is a method [Weston et al., 2014] that, at its simplest, multiplies embedded question vectors by embedded memory vectors and takes the best fit to be the best answer to the question. Another example is a method [Peng et al., 2015] that encodes the question and the facts through multiple layers, and at the final layer the question is put through a function that results in an answer. Both of these methods learn by training on simulated stories with questions and answers and then answer questions like the ones shown below from Weston et al. [2014].

Question Answering

1. Joe went to the kitchen.
2. Fred went to the kitchen.
3. Joe picked up the milk.
4. Joe travelled to the office.
5. Joe left the milk.
6. Joe went to the bathroom.

Q: Where is the milk now? A: office
Q: Where is Joe? A: bathroom
Q: Where was Joe before the office? A: kitchen

## The next frontier is grounded understanding

The methods just discussed learned how to answer questions by reading stories, but some important aspects of stories are so obvious that we don't bother writing them down. Imagine a book on a table. How could a computer learn that if you move the table you also move the book? Likewise, how could a computer learn that it only rains outside? Or, as Marvin Minsky asks, how could a computer learn that you can pull a box with a string but not push it? Since these are the kinds of facts that we don't write down, stories will be limited in the kind of knowledge they can convey to our algorithms. To acquire this knowledge, our robots may need to learn through physical experience or simulated physical experience.

Robots must take this physical experience and encode it in deep neural networks so that general meaning can build upon it. If a robot frequently sees a block fall off a table, it should create neural circuitry that is activated for this event. This circuitry should be associated with the word "fell" when mommy says, "Oh, the block fell." Then, as an adult robot, when it reads the sentence, "The stock fell by 10 points." it should understand what happened using this neural circuitry.

Robots also must tie generalized physical experience to abstract reasoning. Consider trying to understand the implications of the sentence, "He went to the junkyard." WordNet only provides a grouping of words for "went." ConceptNet ties "went" to the word "go" but never actually says what "go" means. FrameNet has a frame for self-motion, which is pretty close, but is still insufficient. Deep learning can encode the sentence as a vector and can answer a question such as "Where is he?" with "junkyard." However, none of these methods conveys a sense of a person being in a different place, which means that he is not here or anywhere else. We need either an interface between natural language and logic, or we need to encode some kind of abstract logic in neural networks.

## Back to the practical: resources for getting started with deep learning

There are a lot of ways to get started. There is a Stanford class on deep learning for NLP [8]. One can also take Professor Hinton's Coursera Course and get it right from the horse's mouth [9]. Also, there is a clear and understandable online textbook in preparation on deep learning from Professor Bengio and friends [10]. To get started programming, if you are a Python person, you can use Theano [11], and if you are a Java person, you can use Deeplearning4j [12].

## Conclusion

The revolution in deep learning is enabled by the power of our computers and the increased digitization of our lives. Deep learning models are successful because they are so big, with models often having millions of parameters. Training these models requires a lot of training data, which comes from our digital lives, and a lot of computation, which comes from Moore's law. To progress to creating real intelligence, we still have to go deeper. Deep learning algorithms must learn from physical experience, generalize this experience, and tie this generalized experience to abstract reasoning.

## Links and References

[1] http://wordnet.princeton.edu/

[2] http://conceptnet5.media.mit.edu

[3] https://framenet.icsi.berkeley.edu/fndrupal/home

[4] https://code.google.com/p/word2vec/

[5] http://nlp.stanford.edu/projects/glove/

[6] https://radimrehurek.com/gensim/models/word2vec.html

[7] For some great examples, see http://cs.stanford.edu/people/karpathy/deepimagesent/ and http://googleresearch.blogspot.com/2014/11/a-picture-is-worth-thousand-coherent.html.

[8] http://cs224d.stanford.edu/syllabus.html

[9] https://www.coursera.org/course/neuralnets

[10] http://www.iro.umontreal.ca/~bengioy/dlbook/

[11] http://deeplearning.net/software/theano/ and http://www.deeplearning.net/tutorial/

[12] http://deeplearning4j.org/

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards AI. *Large-scale kernel machines*, 34(5), 2007.

Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.



Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

Andrej Karpathy and Li Fei-Fei. Deep visual-semantic alignments for generating image descriptions. *arXiv preprint arXiv:1412.2306*, 2014.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Baolin Peng, Zhengdong Lu, Hang Li, and Kam-Fai Wong. Towards neural network-based reasoning. *arXiv preprint arXiv:1508.05508*, 2015.

Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. Grammar as a foreign language. *arXiv preprint arXiv:1412.7449*, 2014.

Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *arXiv preprint arXiv:1411.4555*, 2014.

Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *arXiv preprint arXiv:1410.3916*, 2014.

Deep Learning for Natural Language Processing
Jonathan Mugan

The Power of Purpose at Work
Reid Hoffman

Daily Pulse: Jobs Report Shatters Estimates, Strikes in Germany, Uber Goes Analog
Katie Carroll

Bihar Exit Poll Summary: BJP moves further ahead
Subhash Chandra

Daily Pulse: Obama Nixes Keystone XL, Square Marks Itself Down, Toyota's Billion-…
John C Abell

Genomics and big data analytics in healthcare
Dr Malpani

No Place for Stress in Teams
Anu Acharya

Before Separating Your Company, Ask Yourself These Three Questions
Meg Whitman

Natural Language Processing, Neural Networks, Machine Learning

Written by
Jonathan Mugan
Follow

Like Comment

28 likes 5 comments

Add your comment

Newest

Balaji K
Cloudadic.com

Nice share. R u interested in this one. Let us chat.

Like Reply 17 hours ago



Reid Hoffman
Entrepreneur. Product Strategist. Investor.


# The Power of Purpose at Work



In 2002 at LinkedIn, there were very few of the perks that Silicon Valley is famous for. Our company headquarters was my apartment's living room. Free lunches? Yes, if there was an extra yogurt in my refrigerator, or if you consider a can of Coke "lunch." Nap pods, on-site yoga classes, wellness centers, concierge services, or haircut days? No.

And yet every day my co-founders came to work early and stayed late. We were on a quest to augment search, data analytics, and network connectivity with real identity, reputation, and trust. The goal: Building a platform that creates economic opportunity for every member of the global workforce. It was an ambitious vision and it gave us a great sense of purpose.

In a start-up environment, when teams are small and most everyone who's involved is, by nature, a risk-taker with a desire to create something that will potentially have outsized impact, it's relatively easy to find purpose-driven individuals.

But to become the company we wanted to become, we knew that we'd eventually need more employees, with different skillsets and different temperaments. And inevitably our culture would change. As I and my fellow instructors explain in our Blitzscaling class at Stanford, a company with 100 employees cannot function effectively using the tactics that work for a company with 10 employees: You need an updated playbook.

Still, my co-founders and I were determined to preserve our shared sense of purpose as a core value, even as we grew. In job interviews, and then again, in new hire orientations, we always emphasized our guiding value: Individual LinkedIn members always come first. Any addition or change we make to the platform must improve it in ways that help individual members increase their economic opportunities.

In emphasizing our philosophy so persistently, we inevitably dissuaded some talented potential hires for whom it did not resonate. But we also attracted individuals who did connect with it, and thus ensured our ongoing cohesiveness even as we started to expand beyond our core team.

Today, LinkedIn has more than 9200 employees. Needless to say, we've moved out of my living room, into offices in more than 30 cities around the world. Our selection of complimentary beverages has scaled up rapidly.

But while the Wild Alaskan Salmon with Avocado-Corn Relish in our café and our 401K matching program make it nicer to work here, a strong sense of purpose remains the defining attribute of our employees.

In fact, a consultancy called Imperative that publishes a national index measuring how purpose-oriented the U.S. workforce is across industries, job types, and more, recently surveyed 2000 of LinkedIn's global employees. It found that 41 percent of them fit its purpose-oriented profile – they prioritize meaning and fulfillment over money and status. That's nearly twice as high as the U.S. tech industry average of 21 percent.

LinkedIn benefits from this orientation in a number of ways. According to Imperative's research, purpose-oriented employees are:

* 54 percent more likely to stay at a company for 5-plus years
* 30 percent more likely to be high performers
* 69 percent more likely to be Promoters on Bain & Company's eNPS scale, which measures employee engagement and loyalty

In *The Alliance*, my co-authors and I present the "tour of duty" as a mechanism for creating and maintaining engagement in an era when lifetime employment is no longer a given. The key to a successful tour of duty is a high degree of alignment between employer and employee. And the key to a high degree of alignment is a shared sense of purpose.

When that shared sense of purpose exists, employees stay at a company longer. Their high level of engagement leads to higher levels of loyalty and performance.

In the old days of lifetime employment, the presumed payoffs were security and predictability. Now, more and more professionals look for positions at companies where they can create meaningful impact and experience personal growth.

Companies that understand the increasing emphasis of purpose in today's professional landscape improve their ability to attract such employees and also their ability to retain them for longer periods of time.

And of course a virtuous loop inevitably kicks in. At LinkedIn, when I see how thousands of LinkedIn employees are committed to creating economic opportunity for every member of the global workforce, I get even more inspired about where we're headed. Jeff Weiner calls this shared sense of vision LinkedIn's true north. It has guided my own efforts for more than a decade now, and it continues to help us attract and retain exactly the kind of committed professionals who are helping us realize our vision.

Talent Management, Employee Engagement

Featured In Company Culture, Entrepreneurship, Editor's Picks

Written by
Reid Hoffman

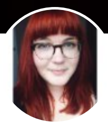
Follow



Patrick Mullen
Sales Rep at Will Foods

I love to see pieces being shared of other people, companies, etc., that understand that while money is important to existence and of course making a much more materialistically fulfilling life for us, for real dreams to come true and goals to be achieved, it cannot be the real, true and ultimate motivator.

Money is one hell of an incentive, but the real motivation, has to be the desire to as Reid states, " Be Purposeful"! An old saying that I've always referred back to is" Necessity is the mother of invention" That's purposeful! Money really has never been the driving motivator in inventions.

Katie Carroll
Senior Editor, Social Media & EMEA at LinkedIn



# Daily Pulse: Jobs Report Shatters Estimates, Strikes in Germany, Uber Goes Analog



**Today's jobs report is a doozy.** "To say the October payrolls report will be closely watched across the world would be an understatement. It'll be monumental," wrote Sho Chandra at Bloomberg before this morning's announcement. That's because this is the penultimate jobs report before the Fed decides whether to raise interest rates in December. And, with 271,000 jobs added last month and unemployment hitting a 7-year low of 5%, a rate hike may very well be on the agenda. Analysts were expecting an additional 182,000 jobs in October, and economists consider 5% unemployment 'full employment'—so both figures suggest a healthy American economy.

***

**Strike Watch:** Fresh from a successful negotiation with Fiat Chrysler, United

Auto Workers is in the midst of negotiations with GM—and things are going well. Union leadership has been pressing plants to approve the deal, saying that rejection could lead to a strike and eventually a contract that's less favorable to employees. Final results should be known by Saturday.

Air travelers, though, should brace for some hiccups in Germany: starting this afternoon (German time), the air crew union is going on strike against Lufthansa. Flight staff will be walking out at Frankfurt (the airline's hub) and Duesseldorf. The walkout comes after both parties failed to reach an agreement on pensions and other cost-cutting measures, which Lufthansa is employing to better compete with budget airlines like Ryanair. These changes have proved a tough pill to swallow after the company reported strong Q3 earnings.

***

**Uber is going old-school**: Although going cashless has been one of Uber's selling points, that hasn't worked so well in certain markets. In Indonesia and the Philippines, for example, only 10% of citizens prefer to pay by credit card. So Uber is experimenting with a cash payment option—all of Uber's 22 cities in India, as well as markets like Nairobi, Riyadh, Hanoi, and Manila, are running the test. It also helps Uber compete with local services, like Ola in India (which boasts 80% of the on-demand taxi market) or GrabTaxi in Indonesia, Vietnam, and the Philippines.

***

**Liar, liar, pants on fire?** New York Attorney General Eric T. Schneiderman is investigating Exxon Mobil to determine whether it was less than truthful about the risks of climate change. The company has included research about climate change in its earnings and other reports to investors, highlighting effects it could have on the oil business. The investigation seeks to illuminate whether these public statements align with the company's internal climate change research. According to the New York Times: "The inquiry would include a period of at least a decade during which Exxon Mobil funded outside groups that sought to undermine climate science, even as its in-house scientists were outlining the potential consequences — and uncertainties — to company executives." Exxon Mobil may not be alone, though; the investigation could spark closer looks at climate change research from other oil companies.

***

*Cover Photo: Remember, remember the fifth of November. Yesterday was Bonfire Night in the UK, which commemorates the failed Gunpowder Plot—an attempt in 1605 to blow up Parliament. It's celebrated across the country,* *and many areas add local flair: Lewes, where the above photo was taken, replaced the traditional Guy Fawkes effigy with one of David Cameron and, erm, certain livestock. The only part of England that doesn't celebrate Guy Fawkes Day? St. Peter's, his former school in York.*

***

**What you may have missed — and *really* should read:**

- With Prop F defeated, Airbnb is unbeatable. Here's why
- "I was bossed into being a boss," explains Gimlet Media's Alex Blumberg
- Disney and Vice Media—this odd couple is a lesson in the future of television

****

- **Missed the last update? No one needs to know: 'Trans Pacific Partnership Revealed, 'Safe Harbor' Assured, Boeing is Lost in Space'**
- **Now you can download LinkedInPulse for Apple Watch! Read more and check it out in the App Store.**



Written by
Katie Carroll